

FAST KRASNOSELSKII–MANN METHOD WITH OVERRELAXATION AND PRECONDITIONERS

RADU I. BOT^{*}, ENIS CHENCHENE^{*}, AND JALAL M. FADILI[†]

ABSTRACT. We study accelerated Krasnoselskii–Mann-type methods with preconditioners in both continuous and discrete time. From a continuous-time model, we derive a generalized fast Krasnoselskii–Mann method, providing a new yet simple proof of convergence that leads to unprecedented flexibility in parameter tuning, allowing up to twice larger relaxation parameters. Our analysis unifies inertial and anchoring acceleration mechanisms and offers a broad range of parameter choices, which prove beneficial in practice. Additionally, we extend our analysis to the case where preconditioners are allowed to be degenerate, unifying the treatment of various splitting methods and establishing the weak convergence of shadow sequences.

Keywords: Inertial Accelerations, Nesterov Acceleration, Halpern Method, Fast-Krasnoselskii–Mann Method, Degenerate Proximal Point Algorithm

1. INTRODUCTION

Let $T: H \rightarrow H$ be a nonexpansive operator on the real Hilbert space H . Our objective is to solve the fixed-point problem:

$$(1.1) \quad \text{Find } x \in H \text{ such that } x = T(x),$$

with fast methods in continuous and discrete time, and apply these results to splitting methods providing convergence of the so-called shadow sequences (to be made precise later).

If $T = I - \nabla f$, where $f: H \rightarrow \mathbb{R}$ is a convex and smooth function whose set of minimizers is nonempty, (1.1) is equivalent to minimizing f . The most popular accelerated method to minimize f in discrete time is the celebrated Nesterov’s algorithm [52]. Su, Boyd, and Candès in [57] later discovered that Nesterov’s algorithm can be understood as a specific finite differences discretization of the following ODE

$$(1.2) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0, \quad \text{for } t \geq t_0 > 0$$

with $\alpha = 3$ and initial data $(x(t_0), \dot{x}(t_0)) = (x_0, v_0) \in H^2$. Although the convergence of $x(t)$ to a minimizer of f remains an open problem for $\alpha = 3$ (except in one dimension where convergence holds [8]), the trajectories of (1.2) satisfy $f(x(t)) - \min f = \mathcal{O}(t^{-2})$ as $t \rightarrow +\infty$, which is an improvement over the $o(t^{-1})$ rate of the standard gradient flow. Later, in [9, 30], it has been shown that replacing the critical parameter $\alpha = 3$ with any $\alpha > 3$ ensures the weak convergence of $x(t)$ to a minimizer of f , and improves the rate of convergence for the function values from $\mathcal{O}(t^{-2})$ to $o(t^{-2})$.

However, these inertial acceleration methods often suffer from significant oscillations. This issue was addressed by Attouch, Peypouquet, and Redont [13], and later by Attouch, Chbani, Fadili, and Riahi [7, 6], who proposed to couple the asymptotically vanishing viscous damping term in (1.2) to an additional *geometric Hessian-driven damping term*, previously introduced in [3, 12], resulting in the system

$$(1.3) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t) \frac{d}{dt} \nabla f(x(t)) + \nabla f(x(t)) = 0, \quad \text{for } t \geq t_0,$$

which, for $\alpha > 3$ and appropriate coefficient $\beta(t) > 0$, has the important benefit of retaining the $o(t^{-2})$ rate on the objective, while ensuring the weak convergence of trajectories, and attenuating oscillations.

The research of RIB and EC has been supported by the Austrian Science Fund (FWF), projects W 1260 and P 34922-N, respectively.

Inspired by (1.3), we here study the much more general case where ∇f is replaced by any monotone and Lipschitz continuous operator $Q: H \rightarrow H$:

$$(1.4) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t) \frac{d}{dt} Q(x(t)) + b(t) Q(x(t)) = 0, \quad \text{for } t \geq t_0.$$

For instance, to solve (1.1), it is natural to consider $Q := I - T$, which is indeed monotone and Lipschitz, and the equilibria of (1.4) are solutions to (1.1). Note that while oscillations attenuation with Hessian-driven damping are quite well-understood and intuitive in optimization contexts (where $Q = \nabla f$), the interpretation of the geometric damping for general monotone operators is not always clear and remains an open problem. In fact, the geometric damping might not always reduce oscillations for general monotone operators as we will see in some case in our numerical results. Nonetheless, we will still stick with this terminology in this paper for consistency reasons.

Although the system in (1.3) generates fast converging trajectories for a range of admissible parameters (cf. [6, Theorem 2.1]), when one turns to the general operator setting (1.4) the situation becomes significantly more delicate. A notable contribution in this direction was made by BoȚ, Csetnek and Nguyen in [20], who studied (1.4) with $b(t) := \frac{1}{2}(\dot{\beta}(t) + \frac{\alpha}{t}\beta(t))$, the so-called fast Optimistic Gradient Descent Ascent (Fast-OGDA) system. In particular, they showed that when β is constant and $\alpha > 2$, this dynamical system generates trajectories weakly converging to a zero of Q (provided that the set of zeros of Q is non-empty) with a convergence rate of $\|Q(x(t))\| = o(t^{-1})$ as $t \rightarrow +\infty$. Therefore, (1.4) can be seen as an accelerated version of the first order system

$$(1.5) \quad \dot{x}(t) + Q(x(t)) = 0, \quad \text{for } t \geq t_0.$$

Indeed, when Q is a cocoercive operator, it is known that each solution trajectory generated by (1.5) weakly converges to a zero of Q with a convergence rate of $\|Q(x(t))\| = o(t^{-\frac{1}{2}})$ as $t \rightarrow +\infty$ [5]. To further emphasize this crucial difference, we refer to rates of the form $\mathcal{O}(t^{-1})$, or $o(t^{-1})$ as *fast rates*.

It turns out that the specific choice $b(t) := \frac{1}{2}(\dot{\beta}(t) + \frac{\alpha}{t}\beta(t))$ in (1.4) considered by [20] is ultimately due to the impossibility of exploiting function values in the Lyapunov analysis. This immediately leads us to the natural question:

Can Fast-OGDA in [20] be extended to accommodate more general parameters b , enabling new fast and flexible numerical algorithms for solving (1.1)?

In this paper, we show that is possible to consider the more general choice¹

$$(1.6) \quad b(t) := (1 - \eta)\dot{\beta}(t) + \theta \frac{\beta(t)}{t}, \quad \text{for } \eta \in (0, 1) \quad \text{and} \quad \theta := (1 - \eta) + \eta(\alpha - 1),$$

which recovers [20] by setting $\eta = \frac{1}{2}$, i.e., $\theta = \frac{\alpha}{2}$. Our convergence guarantees for a general parameter $\eta \in (0, 1)$ cover those in [20]: in particular, we establish that, e.g., if $\alpha > 2$ and β is constant, we have weak convergence of the trajectories to zeros of Q with a convergence rate of $\|Q(x(t))\| = o(t^{-1})$ as $t \rightarrow +\infty$. Additionally, we show that in the two edge cases $\eta = 0$ or $\eta = 1$ the rate $\|Q(x(t))\| = o(t^{-1})$ cannot be achieved in general, unveiling the tightness of our analysis and an intriguing fundamental limitation in the choice of the parameter function b .

The Fast-OGDA system is in turn related to a seemingly completely different mechanism: the Tikhonov regularization, also known as anchoring mechanism. Specifically, for $\beta(t) \equiv 1$ and $\eta = 0$, the dynamical system (1.4), can be easily seen to be equivalent to:

$$(1.7) \quad \dot{x}(t) + Q(x(t)) + \frac{\alpha - 1}{t}(x(t) - v) = 0, \quad \text{for } t \geq t_0,$$

with anchor point $v := x_0 + \frac{t_0}{\alpha - 1}(\dot{x}(t_0) + Q(x_0))$, cf. Section 2.2. The dynamical system (1.7) is nothing but the Tikhonov regularization of the monotone operator flow (1.5), see [28, 55, 34], that in fact not only generates trajectories that converge strongly to the projection of v onto $\text{zer } Q$, but also exhibits the rate $\|Q(x(t))\| = \mathcal{O}(t^{-1})$ as $t \rightarrow +\infty$ for all $\alpha \geq 2$.

¹Observe that $\theta \in (1, \alpha - 1)$ is a convex combination of 1 and $\alpha - 1$ with parameter $\eta \in (0, 1)$. The redundancy in notation will be justified later in the analysis, which turns out more convenient to write with respect to η .

Algorithm 1: (Generalized) Fast Krasnoselskii–Mann Method**Pick** $\sigma > 0$, $\alpha > 2$, $\theta \in (1, \alpha - 1)$ **Initialize** $x^{-1}, x^0 \in H$ **for** $k = 0, 1, 2, \dots$ **do**

$$x^{k+1} = x^k + \frac{\theta}{k + \sigma} (T(x^k) - x^k) + \left(1 - \frac{\alpha}{k + \sigma}\right) (T(x^k) - T(x^{k-1}))$$

The system (1.7) can also be recognized as the continuous-time model of the celebrated Halpern method which is a classical method in fixed-point theory [22, 58]. A simplified version of (1.7) was studied in [58] in finite-dimensional spaces. In the discrete-time setting, Sabach and Shtern in [56] and, independently, Lieder in [47], proved that Halpern’s method achieves the convergence rate of $\mathcal{O}(k^{-1})$ as $k \rightarrow +\infty$ on the norm of the fixed-point residual. Interestingly, this cannot be improved to $o(k^{-1})$ in normed spaces; see [35]. The connection between Halpern’s method and momentum-type algorithms have been observed previously in [35, 62, 61].

The $o(k^{-1})$ convergence rate of the fixed-point residual norm is achieved by Algorithm 1, the main focus of our paper. This algorithm solves the fixed-point problem (1.1) and can be viewed as a discrete-time analogue of the continuous dynamics (1.4), with b given by (1.6), $\beta \equiv 1$, and $Q = I - T$. It yields an accelerated variant of the classical Krasnoselskii–Mann method [45, 49], which for $x^0 \in H$ and *relaxation parameter* $\theta \in (0, 1)$, iterates

$$(1.8) \quad x^{k+1} = x^k + \theta(T(x^k) - x^k), \quad \text{for } k \geq 0.$$

The scheme (1.8) is known to only exhibit the rate $o(k^{-\frac{1}{2}})$ as $k \rightarrow +\infty$ on the fixed-point residual; see [46, 37]. We can see that Algorithm 1 extends (1.8) by incorporating a *momentum* term, governed by the standard coefficient $\alpha_k := 1 - \frac{\alpha}{k + \sigma}$, and features *vanishing* relaxation parameters $\theta_k := \frac{\theta}{k + \sigma}$. Leveraging the enhanced flexibility of b in (1.6), it extends the Fast-Krasnoselskii–Mann method (Fast-KM) introduced by Boţ and Nguyen in [21], allowing for more general parameters $\theta \in (1, \alpha - 1)$ and $\sigma > 0$. It results, in particular, in (almost) *twice larger* relaxation parameters—from $\theta = \frac{\alpha}{2}$ to θ arbitrarily close to $\alpha - 1$, as well as a general parameter $\sigma > 0$. As we will show in Section 5, this enhanced flexibility leads to notable practical improvements, consistently outperforming the baseline method in several benchmark problems. We strongly rely on the important interplay between continuous and discrete-time dynamics to derive a convergence proof for Algorithm 1 that matches almost exactly the continuous-time approach, making it easily accessible using the latter as a *guideline*. Additionally, this allows us to cover, in a unifying fashion, the important special case of $\alpha = \sigma = 2$, for which the method reduces to the Optimal Halpern Method [47], as well as the edge case $\theta = 1$, for which Algorithm 1 reduces to a Halpern-type method, much like (1.7) in continuous time.

Since Algorithm 1 allows us to find fixed points of nonexpansive operators, it emerges as a unified algorithmic framework to design accelerated variants of *operator splitting methods*, such as primal-dual [31], forward–backward [16], Douglas–Rachford [38, 48], Davis–Yin [36, 54], and several other extensions. Following the recent insights in [43, 24, 25], these methods can be studied as fixed-point iterations with respect to nonexpansive mappings in a renormed Hilbert space, see Section 4. In this context, the metric might actually be *degenerate*, i.e., only inducing a Hilbert seminorm, and thus the analysis requires a careful adaptation. We do so by delineating a *degenerate* Opial Lemma, a fundamental ingredient that allows us to adapt the convergence analysis of Algorithm 1 almost effortlessly to this degenerate setting. We apply this result to accelerate the convergence of a recent extension of the Douglas–Rachford splitting method [48, 25], ensuring the weak convergence of the so-called *shadow* sequences in an infinite-dimensional Hilbert space setting.

1.1. Contributions. Our main contributions are summarized as follows:

- We introduce a generalization of the Fast-OGDA system introduced by BoȚ, Nguyen and Csetnek in [20] with b as in (1.6), and show that if β is constant, our analysis is tight in the sense that the edge cases in the parameter bounds (1.6) do not lead to fast little- o convergence rates.
- We introduce Algorithm 1 as a discretization of the proposed continuous-time model and propose a new convergence proof that matches almost exactly the continuous-time approach. The method generalizes [21] integrating two degrees of freedom: it considers a general parameter $\sigma > 0$ and a general parameter $\theta \in (1, \alpha - 1)$, while maintaining the fast little- o rate on the fixed-point residual. Additionally, we cover the convergence analysis also in the edge case $\alpha = \sigma = 2$, which corresponds to the Optimal Halpern Method [62], thereby providing a unified analysis of *inertial* and *anchoring* (Tikhonov) acceleration mechanisms.
- We establish an Opial-type lemma with degenerate preconditioners, and adapt our convergence proof to this setting. We further discuss how this allows us to systematically obtain convergence results for several accelerated splitting methods, establishing the weak convergence of the so-called *shadow sequences* in infinite dimensional Hilbert spaces, thus closing a gap in [21].
- We present numerical experiments that demonstrate that the Fast-KM method with large relaxation parameters seemingly outperforms the baseline choice in [21].

1.2. Notation and Preliminaries. In this section, we fix the notation and introduce some preliminary results on nonexpansive operators in Hilbert spaces and their connection to monotone operators; see [14] for a comprehensive account. An operator $T: H \rightarrow H$ is said to be *nonexpansive* if

$$(1.9) \quad \|T(x) - T(x')\| \leq \|x - x'\|, \quad \text{for all } x, x' \in H.$$

The set of *fixed points* of T is defined by $\text{Fix } T := \{x \in H \mid x = T(x)\}$. In this paper, we always assume that $\text{Fix } T \neq \emptyset$. If T is nonexpansive, the operator $T_s := sT + (1-s)I$ for $s \in (0, 1)$ is said to be *s-averaged*. $T_{1/2}$ is called *firmly nonexpansive*. It follows from [14, Remark 4.34(iv)] that $I - T_s$ is $\frac{1}{2s}$ -cocoercive, i.e., it satisfies the following inequality:

$$(1.10) \quad \langle x - x', (I - T_s)(x) - (I - T_s)(x') \rangle \geq \frac{1}{2s} \|(I - T_s)(x) - (I - T_s)(x')\|^2, \quad \text{for } x, x' \in H.$$

In particular, $I - T_s$ is a *monotone* operator. A set-valued mapping $A: H \rightarrow 2^H$ is *monotone* if it obeys

$$(1.11) \quad \langle a - a', u - u' \rangle \geq 0, \quad \text{for all } (u, a), (u', a') \in \text{gra } A,$$

where the set $\text{gra } A \subseteq H \times H$ is the *graph* of A and is defined by $\text{gra } A := \{(u, a) \mid a \in A(u)\}$. A monotone operator is called *maximal monotone* if its graph is maximal (with respect to the inclusions of sets) among the class of monotone operators. The inverse of a set-valued mapping A is denoted by A^{-1} and defined through its graph by $\text{gra } A^{-1} := \{(a, u) \mid (u, a) \in \text{gra } A\}$. The *resolvent* of A is the mapping $J_A := (I + A)^{-1}$, where I denotes the identity operator on H . From [14, Proposition 23.8], A is maximal monotone if and only if J_A is a single-valued mapping from H to H , which is 1-cocoercive, or simply *cocoercive* (or *firmly nonexpansive*). Moreover, it is clear that the operator $T := 2J_A - I$ is nonexpansive. Conversely, from [14, Corollary 23.9], if T is nonexpansive, there exists a maximal monotone operator A such that $T = 2J_A - I$. Furthermore, fixed points of T correspond to *zeros* of A , i.e., elements of $\text{Fix } T = \text{zer } A := \{u \in H \mid 0 \in A(u)\}$. We also denote $\rightharpoonup$ as weak convergence and $\rightarrow$ as strong convergence. Eventually, we denote $\llbracket 1, n \rrbracket := \{1, \dots, n\}$ for all $n \in \mathbb{N}$.

2. CONTINUOUS-TIME SYSTEM AND ANALYSIS

In this section, we introduce a general dynamical system that extends Fast-OGDA from [20]. Let $Q: H \rightarrow H$ be a monotone and Lipschitz continuous operator² and consider the dynamical system:

$$(2.1) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t) \frac{d}{dt} Q(x(t)) + b(t) Q(x(t)) = 0, \quad t \geq t_0 > 0,$$

²Note that the Lipschitz continuity of Q is only useful to establish the existence and uniqueness of a strong solution to (2.1) which in turn entails the almost everywhere differentiability of $t \mapsto Q(x(t))$.

where b is defined for all $t \geq t_0$ by

$$(2.2) \quad b(t) := (1 - \eta)\dot{\beta}(t) + \theta \frac{\beta(t)}{t}, \quad \text{where } \theta := (1 - \eta) + \eta(\alpha - 1),$$

with the additional parameter $\eta \in (0, 1)$. The system (2.1) recovers the choice in [20] by setting $\eta = \frac{1}{2}$, for which $\theta = \frac{\alpha}{2}$, and can be leveraged to approach solutions to (1.1) by setting $Q = I - T$, since indeed:

$$\text{zer } Q = \text{zer}(I - T) = \text{Fix } T.$$

We will need that (2.1) admits at least one strong global solution for every starting point $x(0) = x_0 \in H$, i.e., such that x and $\dot{x}$ are locally absolutely continuous on $[t_0, +\infty)$, and x satisfies (2.1) *almost everywhere* (a.e.) on $[t_0, +\infty)$. In turn, this will entail that for every locally absolutely continuous trajectory $t \mapsto x(t)$ the trajectory $t \mapsto Q(x(t))$ is also locally absolutely continuous. In fact, using [42, Proposition 6.2.1], Lipschitz continuity of Q and mild conditions on $\beta(t)$, it can be shown that the system (2.1) has a unique strong solution trajectory. In fact, Lipschitz continuity of Q can even be weakened to Lipschitz continuity over the bounded subsets of H to show the existence and uniqueness of a maximal solution. The passage to a global solution would use a classical contradiction argument and boundedness of the trajectory that we will show in our Lyapunov analysis.

2.1. Lyapunov analysis. To study the second order dynamical system (2.1), we introduce, for $x^* \in \text{zer } Q \neq \emptyset$, the energy function $E^\lambda: [t_0, +\infty) \rightarrow \mathbb{R}_+$ that will play a fundamental role in our analysis,

$$(2.3) \quad \begin{aligned} E^\lambda(t) &:= \delta(t) \langle Q(x(t)), x(t) - x^* \rangle + \frac{\xi(t)}{2} \|Q(x(t))\|^2 + \frac{c}{2} \|x(t) - x^*\|^2 + \frac{1}{2} \|v^\lambda(t)\|^2, \\ v^\lambda(t) &:= \lambda(x(t) - x^*) + t(\dot{x}(t) + (1 - \eta)\beta(t)Q(x(t))), \end{aligned}$$

for some constant $\lambda \geq 0$, and parameters given by

$$(2.4) \quad \delta(t) := \eta\lambda\beta(t)t, \quad \xi(t) := \eta(1 - \eta)\beta(t)^2t^2, \quad c := \lambda(\alpha - 1 - \lambda), \quad \text{for } t \geq t_0.$$

This choice of the parameters allows us to establish the following result, whose proof is postponed to Appendix B.

Lemma 2.1. *Let $x^* \in \text{zer } Q \neq \emptyset$, $x: [t_0, +\infty) \rightarrow H$ be a solution trajectory to (2.1), and E^λ the energy function defined in (2.3). Then, for almost all $t \geq t_0$, we have*

$$(2.5) \quad \begin{aligned} \frac{d}{dt} E^\lambda(t) &= \lambda t \eta \left(\dot{\beta}(t) + \frac{2 - \alpha}{t} \beta(t) \right) \langle Q(x(t)), x(t) - x^* \rangle - \frac{t}{2} (\alpha - 1 - \lambda) \|\dot{x}(t) + \beta(t)Q(x(t))\|^2 \\ &\quad + \beta(t)t^2 \left((1 - \eta)\eta \left(\dot{\beta}(t) + \frac{2 - \alpha}{t} \beta(t) \right) + \frac{1}{2} (\alpha - 1 - \lambda) \frac{\beta(t)}{t} \right) \|Q(x(t))\|^2 \\ &\quad - \eta t^2 \beta(t) \left\langle \dot{x}(t), \frac{d}{dt} Q(x(t)) \right\rangle - \frac{t}{2} (\alpha - 1 - \lambda) \|\dot{x}(t)\|^2. \end{aligned}$$

Lemma 2.1 naturally leads us to consider the following *growth condition* for $t \mapsto \beta(t)$:

$$(2.6) \quad 0 \leq \frac{\dot{\beta}(t)t}{\beta(t)} \leq \alpha - 2 - \varepsilon, \quad \text{for } \varepsilon \in [0, \alpha - 2], \text{ and all } t \geq t_0,$$

which with a suitable rearrangement, writes equivalently as

$$(2.7) \quad \dot{\beta}(t) + \frac{2 - \alpha}{t} \beta(t) \leq -\varepsilon \frac{\beta(t)}{t}, \quad \text{for all } t \geq t_0.$$

This makes the coefficient in front of the term $\|Q(x(t))\|^2$ negative and bounded away from zero at least for λ sufficiently close to $\alpha - 1$, hence leading to the desired energy decrease. This observation is crucial to establish the following convergence result.

Theorem 2.2. *Let $Q: H \rightarrow H$ be a monotone and Lipschitz continuous operator with $\text{zer } Q \neq \emptyset$. Let $x^* \in \text{zer } Q$, $x: [t_0, +\infty) \rightarrow H$ be a solution trajectory to (2.1) for some $\eta \in (0, 1)$ and $\alpha \geq 2$. Suppose that (2.6) holds with $\varepsilon \geq 0$. Then, it holds:*

$$(2.8) \quad \langle Q(x(t)), x(t) - x^* \rangle = \mathcal{O}\left(\frac{1}{t\beta(t)}\right) \quad \text{and} \quad \|Q(x(t))\| = \mathcal{O}\left(\frac{1}{t\beta(t)}\right) \quad \text{as } t \rightarrow +\infty.$$

Additionally, if $\alpha > 2$ and $\varepsilon > 0$, then the following holds:

(i) We have the following integral bounds:

$$\int_{t_0}^{\infty} \beta(t) \langle Q(x(t)), x(t) - x^* \rangle dt < +\infty, \quad \int_{t_0}^{\infty} t\beta(t)^2 \|Q(x(t))\|^2 dt < +\infty, \quad \int_{t_0}^{\infty} t \|\dot{x}(t)\|^2 dt < +\infty.$$

(ii) $x(t)$ converges weakly to a zero of Q as $t \rightarrow +\infty$.

(iii) The following fast rates of convergence hold:

$$\langle Q(x(t)), x(t) - x^* \rangle = o\left(\frac{1}{t\beta(t)}\right), \quad \|Q(x(t))\| = o\left(\frac{1}{t\beta(t)}\right) \quad \text{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{t}\right) \quad \text{as } t \rightarrow +\infty.$$

Remark 2.3 (The gap function rate). *The quantity $\langle Q(x(t)), x(t) - x^* \rangle$ is obviously nonnegative by monotonicity of Q . To get the significance of the rates related to this quantity, that we dub gap function, we first observe that in the simple case where $Q = \nabla f$, where f is a smooth convex function, then $f(x(t)) - \min_H f \leq \langle Q(x(t)), x(t) - x^* \rangle$ by convexity, whence we recover the same convergence rate on the objective. This remains true even when minimizing $f+g$ where f is an L -smooth function and g is a proper lower semicontinuous convex function with $\text{argmin}(f+g) \neq \emptyset$. We then take $Q(x) = \tau^{-1}(x - M(x))$, where $\tau \in (0, 2/L)$, $M(x) = \text{prox}_{\tau g}(x - \tau \nabla f(x))$, and $\text{prox}_{\tau g}$ is the proximal mapping of τg .*

We will see in Section 4 that this gap function plays also a key role in applications to primal-dual-type algorithms, where it serves as a bound for the primal-dual gap, see Example 4.1.

Proof. To lighten notation, we abbreviate (t) with t as a subscript. Additionally, we write Q_t instead of $Q(x(t))$ and ∂_t instead of $\frac{d}{dt}$. Setting $\lambda := \alpha - 1$ in (2.5), discarding the negative terms, we get using (2.7) and monotonicity of Q , that for almost all $t \geq t_0$:

$$\partial_t E_t^{\alpha-1} \leq -\varepsilon(\alpha-1)\eta\beta_t \langle Q_t, x_t - x^* \rangle - \eta t^2 \beta_t \langle \dot{x}_t, \partial_t Q_t \rangle \leq 0,$$

where the last inner product is non-positive since, using monotonicity of Q ,

$$(2.9) \quad -\langle \dot{x}_t, \partial_t Q_t \rangle = -\lim_{h \rightarrow 0} \left\langle \frac{x_{t+h} - x_t}{h}, \frac{Q(x_{t+h}) - Q(x_t)}{h} \right\rangle \leq 0.$$

Thus, $E_t^{\alpha-1}$ is nonincreasing and since all the terms in (2.3) are nonnegative, we get

$$(\alpha-1)\eta t \beta_t \langle Q_t, x_t - x^* \rangle + \frac{\eta(1-\eta)}{2} \beta_t^2 t^2 \|Q_t\|^2 \leq E_{t_0}^{\alpha-1}, \quad \text{for all } t \geq t_0,$$

hence (2.8).

Let us now turn to claims (i)-(iii). Suppose that (2.6) holds for some $\varepsilon > 0$. Thus, (2.5) and monotonicity of Q yield for all $0 \leq \lambda \leq \alpha - 1$ and almost all $t \geq t_0$:

$$\partial_t E_t^\lambda \leq -\varepsilon \lambda \eta \beta_t \langle Q_t, x_t - x^* \rangle - \frac{t}{2} (\alpha - 1 - \lambda) \|\dot{x}_t\|^2 + \beta_t^2 t \left(-(1-\eta)\eta\varepsilon + \frac{1}{2}(\alpha - 1 - \lambda) \right) \|Q_t\|^2.$$

Let $\bar{\lambda} := \alpha - 1 - \varepsilon\eta(1-\eta)$. Observe that $\bar{\lambda} \in [3(\alpha-2)/4, \alpha-1]$. We then have

$$0 \leq \alpha - 1 - \lambda \leq \varepsilon\eta(1-\eta), \quad \text{for all } \lambda \in [\bar{\lambda}, \alpha - 1].$$

It then follows that for all $\lambda \in [\bar{\lambda}, \alpha - 1]$ and almost all $t \geq t_0$:

$$(2.10) \quad \partial_t E_t^\lambda \leq -\lambda \eta \varepsilon \beta_t \langle Q_t, x_t - x^* \rangle - \frac{t}{2} (\alpha - 1 - \lambda) \|\dot{x}_t\|^2 - \frac{1}{2} \beta_t^2 t (1-\eta)\eta \|Q_t\|^2.$$

Integrating this inequality and using that the energy function E^λ is nonnegative, we get (i).

To show (ii), we start by expressing $E_t^\lambda - E_t^{\alpha-1}$ which reads

$$E_t^\lambda - E_t^{\alpha-1} = -\eta(\alpha-1-\lambda)\beta_t t \langle Q_t, x_t - x^* \rangle + \frac{\lambda}{2}(\alpha-1-\lambda)\|x_t - x^*\|^2 + \frac{1}{2}\|v_t^\lambda\|^2 - \frac{1}{2}\|v_t^{\alpha-1}\|^2.$$

Using the three point identity, we get

$$\begin{aligned} \frac{1}{2}\|v_t^\lambda\|^2 - \frac{1}{2}\|v_t^{\alpha-1}\|^2 &= -\frac{1}{2}\|v_t^\lambda - v_t^{\alpha-1}\|^2 + \langle v_t^\lambda, v_t^\lambda - v_t^{\alpha-1} \rangle \\ &= -\frac{(\alpha-1-\lambda)(\alpha-1+\lambda)}{2}\|x_t - x^*\|^2 - (\alpha-1-\lambda)t \langle x_t - x^*, \dot{x}_t + (1-\eta)\beta_t Q_t \rangle. \end{aligned}$$

Plugging this above we arrive at

$$\begin{aligned} (2.11) \quad E_t^\lambda - E_t^{\alpha-1} &= -\eta(\alpha-1-\lambda)\beta_t t \langle Q_t, x_t - x^* \rangle \\ &\quad - \frac{(\alpha-1)(\alpha-1-\lambda)}{2}\|x_t - x^*\|^2 - (\alpha-1-\lambda)t \langle x_t - x^*, \dot{x}_t + (1-\eta)\beta_t Q_t \rangle \\ &= -(\alpha-1-\lambda)\beta_t t \langle Q_t, x_t - x^* \rangle \\ &\quad - \frac{(\alpha-1)(\alpha-1-\lambda)}{2}\|x_t - x^*\|^2 - (\alpha-1-\lambda)t \langle x_t - x^*, \dot{x}_t \rangle. \end{aligned}$$

Therefore, we get for almost all $t \geq t_0$

$$(2.12) \quad E_t^\lambda - E_t^{\alpha-1} = -(\alpha-1-\lambda) \left[(\alpha-1)q(t) + t\dot{q}(t) - (\alpha-1)I(t) \right],$$

where we have denoted:

$$q(t) := \frac{1}{2}\|x_t - x^*\|^2 + I(t), \quad \text{and} \quad I(t) := \int_{t_0}^t \beta_s \langle Q_s, x_s - x^* \rangle ds.$$

Note that $I(t)$ admits a limit as $t \rightarrow +\infty$ by (i). Moreover, from (2.10) we also get that for each $\lambda \in [\bar{\lambda}, \alpha-1]$ the limit $\lim_{t \rightarrow +\infty} E_t^\lambda$ exists. In particular the limit of $E_t^\lambda - E_t^{\alpha-1}$ exists. Hence, we deduce from (2.12) that $\lim_{t \rightarrow +\infty} q(t) + \frac{t}{\alpha-1}\dot{q}(t)$ exists. This implies, thanks to [13, Lemma A.2] and since $\alpha > 2$, that also $\lim_{t \rightarrow +\infty} q(t)$ exists. In view of the definition of $q(t)$, $I(t)$ admitting a limit, we get that $\lim_{t \rightarrow +\infty} \|x_t - x^*\|$ exists. $x(t)$ is in particular bounded, and one can extract a weakly convergent subsequence. Using (2.8) and that $t \mapsto \beta(t)$ is nondecreasing, it is immediate to see that the weak cluster point of each such weakly converging subsequence is a zero of Q . We now invoke the standard Opial's lemma to get (ii).

Let us now establish the little- o rates in (iii). To do so, we first note that, since the limit of $t \mapsto q(t)$ exists, from (2.12) we get:

$$\lim_{t \rightarrow +\infty} t\dot{q}(t) = \lim_{t \rightarrow +\infty} t \langle \dot{x}_t + \beta_t Q_t, x_t - x^* \rangle \in \mathbb{R}.$$

Therefore, since for almost all $t \geq t_0$

$$\begin{aligned} E_t^{\alpha-1} &= \frac{1}{2}\|(\alpha-1)(x_t - x^*) + t(\dot{x}_t + (1-\eta)\beta_t Q_t)\|^2 + \eta(\alpha-1)\beta_t t \langle Q_t, x_t - x^* \rangle + \frac{\eta}{2}(1-\eta)\beta_t^2 t^2 \|Q_t\|^2 \\ &= \frac{1}{2}(\alpha-1)^2\|x_t - x^*\|^2 + (\alpha-1)t \langle x_t - x^*, \dot{x}_t + \beta_t Q_t \rangle \\ &\quad + \frac{t^2}{2}\|\dot{x}_t + (1-\eta)\beta_t Q_t\|^2 + \frac{\eta}{2}(1-\eta)\beta_t^2 t^2 \|Q_t\|^2, \end{aligned}$$

we deduce that

$$0 \leq \ell := \lim_{t \rightarrow +\infty} t^2 (\|\dot{x}_t + (1-\eta)\beta_t Q_t\|^2 + \eta(1-\eta)\|\beta_t Q_t\|^2)$$

exists. Suppose that $\ell > 0$. Then, there exists $s \geq t_0$ such that

$$\begin{aligned} &\int_{t_0}^{+\infty} t (\|\dot{x}_t + (1-\eta)\beta_t Q_t\|^2 + \eta(1-\eta)\|\beta_t Q_t\|^2) dt \\ &\geq \int_s^{+\infty} t^2 (\|\dot{x}_t + (1-\eta)\beta_t Q_t\|^2 + \eta(1-\eta)\|\beta_t Q_t\|^2) t^{-1} dt \geq \int_s^{+\infty} \frac{\ell}{2t} dt = +\infty, \end{aligned}$$

leading to a contradiction with the last two integral bounds in (i). This shows that $\ell = 0$ and combining this with Jensen's inequality proves the last two little- o rates in (iii). The first rate follows from the second by applying Cauchy–Schwarz and using the boundedness of the trajectories. $\square$

Remark 2.4 (On the growth of β). *The parameter function β can be understood as a time-scaling parameter [10]. As such, it directly impacts the rate of convergence, as shown in Theorem 2.2(iii). However, contrarily to classical autonomous first-order dynamical systems, where the time-scaling parameter can grow arbitrarily fast while maintaining convergence, the non-autonomous nature of our second-order dynamical system—particularly the fixed decay rate of the vanishing damping parameter $\frac{\alpha}{t}$ —imposes a fundamental constraint on the growth of β . Indeed, applying Grönwall's inequality to integrate (2.6), we get that $\beta(t)$ must satisfy*

$$\beta(t) \leq \beta_0 t^{\alpha-2}$$

for $\beta_0 > 0$. Clearly, the growth of $\beta(t)$ depends on the vanishing damping parameter α . Choosing $\beta(t) = t^{\alpha-2-\varepsilon}$ with $\varepsilon \in (0, \alpha-2)$, we can achieve arbitrarily fast decay rates $o(t^{-\nu})$, where $\nu := \alpha-1-\varepsilon$, by increasing α .

2.2. Edge cases and connection to Tikhonov regularization. It is instructive to inspect the edge cases in Theorem 2.2. We can recognize two different boundary regimes: $\theta \in \{1, \alpha-1\}$, which corresponds to $\eta \in \{0, 1\}$, and $\varepsilon = 0$.

2.2.1. Edge cases in θ . As observed by Attouch, Chbani and Riahi in [10], as well as in [20, 21], while the analysis of the continuous-time system (2.1) can be performed with general parameter functions β satisfying the growth condition (2.6), constant parameters β are the only choices currently compatible with explicit numerical schemes. For this reason, we focus here on this important setting, and, without loss of generality, we set $\beta = 1$. We have the two edge cases:

- $\theta = 1$: We get the system

$$(2.13) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \frac{d}{dt}Q(x(t)) + \frac{1}{t}Q(x(t)) = 0, \quad \text{for } t \geq t_0.$$

Multiplying (2.13) by t , we can rewrite (2.13) as

$$(2.14) \quad \frac{d}{dt}(t\dot{x}(t) + tQ(x(t))) + (\alpha-1)\dot{x}(t) = 0, \quad \text{for } t \geq t_0,$$

which, integrating from t_0 to t leads us to

$$(2.15) \quad \dot{x}(t) + Q(x(t)) + \frac{(\alpha-1)}{t}(x(t) - v) = 0, \quad \text{for } t \geq t_0,$$

where $v := x_0 + \frac{t_0}{\alpha-1}(\dot{x}(t_0) + Q(x_0))$. This is the so-called *anchored* or *Tikhonov regularization* of the flow (1.5), which is known to provide i) strong convergence to the orthogonal projection of v onto $\ker Q$ and ii) the rate $\|Q(x(t))\| = \mathcal{O}(t^{-1})$ for all $\alpha \geq 2$, see, e.g., [22, 19].

The rate on the residual cannot be improved to a little- o rate, as the following simple counterexample in dimension 2 shows: Pick $\alpha = 3$ and $Q: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ being the 90° counter-clockwise rotation, which is indeed monotone and Lipschitz. Identifying $x(t)$ with a complex number, and $Q(x(t))$ as $ix(t)$, the general solution $x: [t_0, +\infty) \rightarrow \mathbb{C} \cong \mathbb{R}^2 = H$ of (2.13) can be obtained explicitly and is defined for $c_1, c_2 \in \mathbb{R}$, as

$$(2.16) \quad x(t) = \frac{c_1(1-it)}{t^2} + \frac{c_2}{t^2}e^{-it}, \quad \text{for } t \geq t_0.$$

As it can be immediately observed, this only satisfies $t\|Q(x(t))\| = t\|ix(t)\| = t\|x(t)\| \geq \frac{c_1}{2}$ for all $t \geq t_1$ and $t_1 > t_0$ large enough. Therefore, the little- o rates *cannot* be achieved in this setting.

- $\theta = \alpha - 1$: It leads us to the system:

$$(2.17) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \frac{d}{dt}Q(x(t)) + \frac{\alpha-1}{t}Q(x(t)) = 0, \quad \text{for } t \geq t_0.$$

The same choice as in the first point above, i.e., with $Q(x(t)) = ix(t)$, $\alpha = 3$, and $H = \mathbb{R}^2 \cong \mathbb{C}$, leads us to a general solution $x: [t_0, +\infty) \rightarrow \mathbb{C}$ defined for $c_1, c_2 \in \mathbb{R}$, as,

$$(2.18) \quad x(t) = \frac{c_1}{t^2} + \frac{c_2(1+it)}{t^2} e^{-it}, \quad \text{for } t \geq t_0,$$

which, similarly as before, only satisfies $t\|Q(x(t))\| = t\|ix(t)\| = t\|x(t)\| \geq \frac{c_2}{2}$ for all $t \geq t_1$ and $t_1 > t_0$ large enough. Also in this edge case no little- o rates on the residual can be achieved. However, differently from the case $\theta = 1$, the dynamics (2.17) does not seem to reduce to a known first-order dynamical system and its asymptotic properties for general monotone Lipschitz operators remain an open problem.

2.2.2. Saturating the growth condition. Let us now consider a non-constant parameter β , but saturating the growth condition (2.6), i.e., with $\varepsilon = 0$, which yields, for $\alpha \geq 2$,

$$(2.19) \quad \dot{\beta}(t) + \frac{\beta(t)}{t}(2 - \alpha) = 0, \quad \text{which is solved for } \beta(t) = \beta_0 t^{\alpha-2}, \quad \text{for } \beta_0 > 0.$$

One can check that in this case $b(t) = (1 - \eta)\dot{\beta}(t) + \theta\frac{\beta(t)}{t} = \dot{\beta}(t) + \frac{\beta(t)}{t}$ and thus (2.1) reads as

$$(2.20) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\frac{d}{dt}Q(x(t)) + \left(\dot{\beta}(t) + \frac{1}{t}\beta(t)\right)Q(x(t)) = 0, \quad \text{for } t \geq t_0.$$

Multiplying (2.20) by t we shall write:

$$(2.21) \quad \begin{aligned} 0 &= t\ddot{x}(t) + \dot{x}(t) + t\beta(t)\frac{d}{dt}Q(x(t)) + \left(t\dot{\beta}(t) + \beta(t)\right)Q(x(t)) + (\alpha - 1)\dot{x}(t) \\ &= \frac{d}{dt} \left[t(\dot{x}(t) + \beta(t)Q(x(t))) \right] + (\alpha - 1)\dot{x}(t). \end{aligned}$$

Integrating (2.21) from t_0 to t , we unveil two interesting properties. First, if $\alpha = 1$ the system is fundamentally equivalent to a first-order system with perturbation and time-scaling (see [5]). Secondly, if $\alpha > 1$, we can rewrite it as

$$(2.22) \quad \dot{x}(t) + \beta_0 t^{\alpha-2}Q(x(t)) + \frac{\alpha-1}{t}(x(t) - v) = 0, \quad t \geq t_0,$$

where $v := \frac{t_0}{\alpha-1}(\dot{x}(t_0) + \beta(t_0)Q(x(t_0))) + x(t_0)$. We are once again in the setting of Tikhonov regularization and continuous-time version of Halpern method: The system (2.22) generates trajectories that converge *strongly* to the projection of v onto the set of zeros of Q with convergence rate $\|Q(x(t))\| = \mathcal{O}(t^{-(\alpha-1)})$, see, e.g., [34] and [22, Corollary 2.16.] for the rate.

3. CONVERGENCE ANALYSIS OF ALGORITHM 1

In this section, we present the convergence analysis of Algorithm 1. Let us first show that the latter can indeed be understood as a discretization of (2.1). Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1, $Q := I - T$ and, for all $k \geq -1$, set

$$(3.1) \quad z^{k+1} := T(x^k), \quad \text{and} \quad Q^{k+1} := Q(x^k) = (I - T)(x^k) = x^k - z^{k+1}.$$

Recall that Algorithm 1 reads, for all $k \geq 0$, as

$$(3.2) \quad x^{k+1} - x^k = \frac{\theta}{k+\sigma}(T(x^k) - x^k) + \left(1 - \frac{\alpha}{k+\sigma}\right)(T(x^k) - T(x^{k-1})).$$

Therefore, using the definition of Q and (3.1), we get:

$$(3.3) \quad z^{k+2} - z^{k+1} + Q^{k+2} - Q^{k+1} = -\frac{\theta}{k+\sigma}Q^{k+1} + \left(1 - \frac{\alpha}{k+\sigma}\right)(z^{k+1} - z^k).$$

Now, bringing every term to the left-hand side, we get:

$$(3.4) \quad z^{k+2} - 2z^{k+1} + z^k + \frac{\alpha}{k+\sigma}(z^{k+1} - z^k) + (Q^{k+2} - Q^{k+1}) + \frac{\theta}{k+\sigma}Q^{k+1} = 0.$$

Denoting by $t_k := k - 1 + \sigma$, we get for all $k \geq 0$:

$$(3.5) \quad (z^{k+2} - 2z^{k+1} + z^k) + \frac{\alpha}{t_{k+1}} (z^{k+1} - z^k) + (Q^{k+2} - Q^{k+1}) + \frac{\theta}{t_{k+1}} Q^{k+1} = 0,$$

which, when $z^{k+1} \simeq x^k$ (asymptotically the case), can be understood as an explicit discretization of (2.1) with $\beta \equiv 1$, step size 1, and $Q := I - T$, which is Lipschitz continuous and monotone. This explicit discretization differs from the one proposed in [21] in the specific case $\theta = \frac{\alpha}{2}$. The latter relies on a clever product-space reformulation of the second-order dynamical system, whereas (3.5) preserves the structural form of (2.1). As we will see in Section 3.1, this will allow us to construct a new energy function that almost exactly mirrors the continuous-time one in (2.3), all the while allowing for twice larger relaxation parameters.

Remark 3.1 (Resolvent viewpoint). *Consider the multivalued operator $\mathcal{Q}: H \rightarrow 2^H$ defined by $\mathcal{Q} := T^{-1} - I$. In this way, T can be represented as the resolvent operator of $\mathcal{Q}$, i.e., $T = J_{\mathcal{Q}} := (I + \mathcal{Q})^{-1}$. Building on Minty's classical result [51], nonexpansivity properties of T translate to monotonicity properties for $\mathcal{Q}$, and vice-versa. Specifically, if T is firmly nonexpansive, then $\mathcal{Q}$ is maximal monotone. More generally, if T is only nonexpansive, then $\mathcal{Q}$ is $\frac{1}{2}$ -comonotone, i.e.,*

$$(3.6) \quad \langle q' - q, z' - z \rangle \geq -\frac{1}{2} \|q' - q\|^2, \quad \text{for all } (z, q), (z', q') \in \text{gra } \mathcal{Q},$$

see [15] for a comprehensive discussion. Using this notation, we can write:

$$(3.7) \quad Q^{k+1} = (I - T)(x^k) \in \mathcal{Q}(T(x^k)) = \mathcal{Q}(z^{k+1}), \quad \text{for all } k \geq 0.$$

Comparing now with (3.5), despite $\mathcal{Q}$ not being necessarily Lipschitz continuous, not even single-valued, Algorithm 1 can be understood as a semi-implicit finite-difference discretization of (2.1)—implicit on the “Hessian”-driven damping term $\frac{d}{dt}Q(x(t))$ and explicit on the force term $\frac{\theta}{t}Q(x(t))$ —and with $\beta \equiv 1$, step size 1, and $Q := \mathcal{Q}$, now formulated on a single sequence $(z^k)_{k \in \mathbb{N}}$.

The main difference with the implicit discretization of the Fast-OGDA system proposed in [20] is that the latter, being implicit also on the force term, does not lead to an easily implementable numerical method as it involves computing the resolvent of $\mathcal{Q}$ with an iteration-dependent step size, which is hardly possible in practice. While being hard to implement in practice, the approach in [20] allows to consider more general time-scaling parameters β_k , yielding improved rates in the same fashion as in the classical Güler acceleration for proximal-point algorithms [41, 40], see also [20, Remark 2] for a more complete account on the role of β .

3.1. Lyapunov Energy. To analyze Algorithm 1, we denote finite differences with the Δ symbol, i.e., for a sequence $(a^k)_{k \in \mathbb{N}}$, we define $\Delta a_k := a^{k+1} - a^k$ for all $k \geq 0$. Second order differences are denoted by Δ^2 , i.e., $\Delta^2 a_k = \Delta(\Delta a)_k := \Delta a_{k+1} - \Delta a_k$. These notations will prove handy in our analysis since we follow closely the continuous-time Lyapunov analysis. In this sense, it is convenient to recall some discrete-time calculus rule listed in Appendix A.

Inspired by the continuous-time approach, we introduce an energy function that matches almost exactly³ (2.3). Specifically, we consider for all $k \geq -1$ the sequences $z^{k+1} := T(x^k)$ and $Q^{k+1} := Q(x^k)$, where $Q := I - T$, and for all $k \geq 1$, $\lambda \in (0, \alpha - 1]$, and any $z^* \in \text{zer } Q = \text{Fix } T$, the energy function:

$$(3.8) \quad \begin{aligned} E_k^\lambda &:= \delta_k \langle Q^k, z^k - z^* \rangle + \frac{\delta_k}{2} \|Q^k\|^2 + \frac{\xi_k}{2} \|Q^k\|^2 + \frac{c}{2} \|z^{k-1} - z^*\|^2 + \frac{1}{2} \|v^k\|^2, \\ v^k &:= \lambda(z^{k-1} - z^*) + t_{k-1} (\Delta z_{k-1} + (1 - \eta)Q^k), \end{aligned}$$

where $t_k := k - 1 + \sigma$, a constant $\lambda \geq 0$, and parameters given by

$$(3.9) \quad \delta_k := \eta \lambda t_{k-1}, \quad \xi_k := (1 - \eta) \eta t_{k-1}^2, \quad c := \lambda(\alpha - 1 - \lambda), \quad \text{for } k \geq 1.$$

Note that the only substantial difference with (2.3) is that the first inner product $\langle Q^k, z^k - z^* \rangle$ is no longer nonnegative, since $Q^k = Q(x^{k-1})$, i.e. Q is evaluated at x^{k-1} rather than at z^k . Observe that, in light of Remark 3.1, this inner product would be nonnegative as soon as T is firmly nonexpansive, in

³Yet another difference with prior work and in particular with [20, 7] is that the term in v^k is scaled differently.

which case $Q^k \in \mathcal{Q}(z^k)$, with $\mathcal{Q}$ maximal monotone. The second term allows us to make E_k^λ nonnegative even for general nonexpansive operators. Indeed, observing that $Q/2$ is firmly nonexpansive, it follows from [14, Proposition 4.4] that for every $x, y \in H$:

$$(3.10) \quad \langle Q(x) - Q(y), T(x) - T(y) \rangle = \langle Q(x) - Q(y), x - y \rangle - \|Q(x) - Q(y)\|^2 \geq -\frac{1}{2}\|Q(x) - Q(y)\|^2.$$

Using this with $x = x^{k-1}$ and $y = z^* \in \text{Fix } T$, hence $Q(z^*) = 0$, we get that $\langle Q^k, z^k - z^* \rangle + \frac{1}{2}\|Q^k\|^2 \geq 0$.

3.2. Energy Decrease. We start with the following preliminary result, which represents the discrete-time counterpart of Lemma 2.1, and our first main contribution:

Lemma 3.2. *Let $T: H \rightarrow H$ be a nonexpansive operator with $\text{Fix } T \neq \emptyset$. Let $(x^k)_{k \in \mathbb{N}}$ be a sequence generated by Algorithm 1 for some $\eta \in (0, 1)$, $\sigma > 0$, and $\alpha \geq 2$. Then there exists $\underline{\lambda} < \alpha - 1$ such that for all $\lambda \in (\underline{\lambda}, \alpha - 1]$ and all $k \geq 1$ such that $t_k \geq \lambda$ it holds:*

$$(3.11) \quad \begin{aligned} E_{k+1}^\lambda - E_k^\lambda &\leq -\lambda\eta(\alpha - 2) \left(\langle x^{k-1} - T(x^{k-1}), T(x^{k-1}) - z^* \rangle + \frac{1}{2}\|x^{k-1} - T(x^{k-1})\|^2 \right) \\ &\quad \left[-\eta(1 - \eta)(\alpha - 2)t_k - \frac{\eta}{2} \left(1 - \eta + \eta(\alpha - 1)^2 - \lambda(\alpha - 1) \right) \right] \|x^{k-1} - T(x^{k-1})\|^2 \\ &\quad + (\alpha - 1 - \lambda) \frac{t_k - (2\eta\theta + \eta + \theta)}{2} \|x^{k-1} - T(x^{k-1})\|^2 \\ &\quad - (\alpha - 1 - \lambda) \frac{t_k - (\alpha + 2 + \eta(\alpha - 1))}{2} \|T(x^{k-1}) - T(x^{k-2})\|^2. \end{aligned}$$

Observe that the first term on the right-hand side is negative thanks to (3.10). The proof of Lemma 3.2, which is postponed to Appendix B, follows exactly the continuous-time approach with $Q := I - T$. We are now ready to establish the discrete counterpart of Theorem 2.2.

Theorem 3.3. *Let $T: H \rightarrow H$ be a nonexpansive operator with $\text{Fix } T \neq \emptyset$. Let $(x^k)_{k \in \mathbb{N}}$ be a sequence generated by Algorithm 1 for some $\eta \in (0, 1)$, $\sigma > 0$, and $\alpha \geq 2$. Then:*

(i) *We have $\|x^k - T(x^k)\| = \mathcal{O}(k^{-1})$ as $k \rightarrow +\infty$.*

Suppose now that $\alpha > 2$ and let $z^ \in \text{Fix } T$. Then:*

(ii) *We have the summability properties:*

$$\begin{aligned} \sum_{k \in \mathbb{N}} \left(\langle x^k - T(x^k), T(x^k) - z^* \rangle + \frac{1}{2}\|x^k - T(x^k)\|^2 \right) &< +\infty, \\ \sum_{k \in \mathbb{N}} k\|x^k - T(x^k)\|^2 &< +\infty, \quad \text{and} \quad \sum_{k \in \mathbb{N}} k\|x^k - x^{k-1}\|^2 < +\infty. \end{aligned}$$

(iii) *The sequence $(x^k)_{k \in \mathbb{N}}$ converges weakly to a fixed point of T as $k \rightarrow +\infty$.*

(iv) *We have the following fast rates of convergence:*

$$\begin{aligned} \langle x^k - T(x^k), T(x^k) - z^* \rangle + \frac{1}{2}\|x^k - T(x^k)\|^2 &= o(k^{-1}), \\ \|x^k - x^{k-1}\| &= o(k^{-1}) \quad \text{and} \quad \|x^k - T(x^k)\| = o(k^{-1}) \quad \text{as } k \rightarrow +\infty. \end{aligned}$$

Proof. As in the proof of Lemma 3.2, it is convenient to introduce the additional sequence $(z^k)_{k \in \mathbb{N}}$. We recall that the latter is defined by $z^{k+1} := T(x^k)$ and thus satisfies (3.5) with $Q^{k+1} := Q(x^k) := (I - T)(x^k) = x^k - z^{k+1}$ for all $k \geq 0$.

(i) We set $\lambda = \alpha - 1$ and (3.11) becomes, after elementary computations,

$$(3.12) \quad \Delta E_k^{\alpha-1} \leq -\eta(\alpha - 1)(\alpha - 2) \left(\langle Q^k, z^k - z^* \rangle + \frac{1}{2}\|Q^k\|^2 \right) - \eta(1 - \eta)(\alpha - 2) \left(t_k - \frac{\alpha}{2} \right) \|Q^k\|^2,$$

for all $k \geq 1$ such that $t_k \geq \alpha - 1$, which also implies $t_k \geq \frac{\alpha}{2}$ as $\alpha \geq 2$. Therefore, using (3.10), there is $\bar{k} \geq 1$ large enough such that for all $k \geq \bar{k}$ we have $\Delta E_k^{\alpha-1} \leq 0$. Therefore, for all $k \geq \bar{k}$, we have $\frac{\xi_{k+1}}{2}\|Q^{k+1}\|^2 \leq E_{k+1}^{\alpha-1} \leq E_{\bar{k}}^{\alpha-1}$ whence we get the claim recalling the expressions of ξ_k and t_k .

Suppose from now on that $\alpha > 2$.

(ii) Setting again $\lambda = \alpha - 1$ and embarking from (3.12), we have after summing on both sides that

$$\begin{aligned} \eta(\alpha - 1)(\alpha - 2) \sum_{i=\bar{k}}^k \left(\langle Q^k, z^k - z^* \rangle + \frac{1}{2} \|Q^k\|^2 \right) + \eta(1 - \eta)(\alpha - 2) \sum_{i=\bar{k}}^k (t_i - t_{\bar{k}}) \|Q^i\|^2 \\ \leq E_{\bar{k}}^{\alpha-1} - E_{k+1}^{\alpha-1} \leq E_{\bar{k}}^{\alpha-1} < +\infty. \end{aligned}$$

Passing to the limit as $k \rightarrow +\infty$ we get the first two summability claims. For the last one, we set $\lambda = \alpha - 1 - \varepsilon$, for $\varepsilon > 0$ to be chosen small enough. Then, discarding the first negative term in (3.11), we have

$$(3.13) \quad \Delta E_k^{\alpha-1-\varepsilon} \leq \left[- \left(\eta(1 - \eta)(\alpha - 2) - \frac{\varepsilon}{2} \right) t_k + \frac{\eta(1 - \eta)}{2} ((\alpha - 1)^2 - 1) \right] \|Q^k\|^2 \\ - \varepsilon \frac{t_k - (\alpha + 2 + \eta(\alpha - 1))}{2} \|\Delta z_{k-1}\|^2.$$

Therefore, for $0 < \varepsilon < \eta(1 - \eta)(\alpha - 2)$, there exists $\tilde{k} \geq 1$ large enough such that for all $k \geq \tilde{k}$,

$$\Delta E_k^{\alpha-1-\varepsilon} \leq -\frac{\varepsilon}{2} (t_k - t_{\bar{k}}) \|\Delta z_{k-1}\|^2,$$

where we have discarded the first negative term in (3.13). Summing again we get $\sum_{k \in \mathbb{N}} k \|\Delta z_{k-1}\|^2 < +\infty$. Observe now that since $Q^{k+1} = x^k - z^{k+1}$ by definition, we have

$$\|x_{k-1}\|^2 = \|x^k - x^{k-1}\|^2 = \|\Delta z_k + \Delta Q_k\|^2 \leq 2\|\Delta z_k\|^2 + 4\|Q^{k+1}\|^2 + 4\|Q^k\|^2.$$

Multiplying by k , summing and using the summability claims proved above, we conclude.

(iii) We argue as in the continuous-time setting and set $\lambda = \alpha - 1 - \varepsilon$, for $\varepsilon > 0$ small enough as before. We will emphasize the dependency of $(v^k)_{k \in \mathbb{N}}$ on λ by writing explicitly $(v_\lambda^k)_{k \in \mathbb{N}}$. Let $k \geq 1$. We have

$$\begin{aligned} E_k^\lambda - E_k^{\alpha-1} = -\eta(\alpha - 1 - \lambda)t_{k-1} \left(\langle Q^k, z^k - z^* \rangle + \frac{1}{2} \|Q^k\|^2 \right) \\ + \frac{\lambda}{2} (\alpha - 1 - \lambda) \|z^{k-1} - z^*\|^2 + \underbrace{\frac{1}{2} \|v_\lambda^k\|^2 - \frac{1}{2} \|v_{\alpha-1}^k\|^2}_{(I)}. \end{aligned}$$

Let us now compute (I) separately. By the three point identity, we get

$$(I) := \frac{1}{2} \|v_\lambda^k\|^2 - \frac{1}{2} \|v_{\alpha-1}^k\|^2 = \frac{\lambda^2 - (\alpha - 1)^2}{2} \|z^{k-1} - z^*\|^2 - t_{k-1} (\alpha - 1 - \lambda) \langle \Delta z_{k-1} + (1 - \eta)Q^k, z^{k-1} - z^* \rangle.$$

Therefore,

$$\begin{aligned} E_k^\lambda - E_k^{\alpha-1} = -(\alpha - 1 - \lambda) \left[t_{k-1} \langle Q^k, z^k - z^* - (1 - \eta)\Delta z_{k-1} \rangle + \frac{\eta}{2} t_{k-1} \|Q^k\|^2 \right. \\ \left. + \frac{(\alpha - 1)}{2} \|z^{k-1} - z^*\|^2 + t_{k-1} \langle \Delta z_{k-1}, z^{k-1} - z^* \rangle \right]. \end{aligned}$$

Therefore, this can be written as

$$(3.14) \quad \begin{aligned} E_k^\lambda - E_k^{\alpha-1} = -(\alpha - 1 - \lambda) \left[(\alpha - 1)q_k + t_{k-1}\Delta q_k - (\alpha - 1)I_k \right. \\ \left. - (1 - \eta)t_{k-1} \langle Q^k, \Delta z_{k-1} \rangle - \frac{t_{k-1}}{2} (\|\Delta z_{k-1}\|^2 + (1 - \eta)\|Q^k\|^2) \right], \end{aligned}$$

where we have denoted

$$q_k := \frac{1}{2} \|z^{k-1} - z^*\|^2 + I_k, \quad \text{and} \quad I_k := \sum_{j=1}^{k-1} \left(\langle Q^j, z^j - z^* \rangle + \frac{1}{2} \|Q^j\|^2 \right).$$

Recall that I_k is nonnegative thanks to (3.10), hence so is q_k . As argued in the proof of (i), $\Delta E_k^{\alpha-1} \leq 0$ for all k large enough and thus $\frac{\varepsilon}{2} \|z^k - z^*\|^2 \leq E_k^{\alpha-1} < +\infty$, i.e. $(z^k)_{k \in \mathbb{N}}$ is a bounded sequence. It

then follows from the last two summability properties of (ii) that the terms in the second line of (3.14) converge to 0. (ii) also gives that I_k admits a limit. Moreover, E_k^λ and E_k^α are nonnegative and decreasing for k large enough as argued above. Therefore, we deduce that

$$(3.15) \quad \lim_{k \rightarrow +\infty} q_k + \frac{t_{k-1}}{\alpha - 1} \Delta q_k = \lim_{k \rightarrow +\infty} (E_k^\lambda - E_k^{\alpha-1}) \in \mathbb{R}.$$

We are then in position to apply [23, Lemma A.4] (see also [6, Lemma A.2] for a weaker variant that requires $\sigma \in \mathbb{N}$) to (3.15) to infer that $(q_k)_{k \in \mathbb{N}}$ admits a limit. Using again that I_k has a limit by (ii), we get that $(\|z^k - z^*\|)_{k \in \mathbb{N}}$ converges. Since $\|Q^k\| = \|x^{k-1} - T(x^{k-1})\| = \|x^{k-1} - z^k\| \rightarrow 0$ by (i), we have $\lim_{k \rightarrow +\infty} \|x^{k-1} - z^*\|^2 = \lim_{k \rightarrow +\infty} \|z^k - z^*\|^2$ and thus the limit of $\|x^k - z^*\|$ exists. The weak convergence claim of $(x^k)_{k \in \mathbb{N}}$ then follows from the standard Opial's lemma.

(iv) Let us first note that since the limit of $(q^k)_{k \in \mathbb{N}}$ exists (see the proof of (iii)), (3.15) yields

$$(3.16) \quad \mathbb{R} \ni \lim_{k \rightarrow +\infty} t_{k-1} \Delta q_k = \lim_{k \rightarrow +\infty} t_{k-1} \left[\langle Q^k, z^k - z^* \rangle + \frac{1}{2} \|Q^k\|^2 + \langle \Delta z_{k-1}, z^k - z^* \rangle - \frac{1}{2} \|\Delta z_{k-1}\|^2 \right].$$

Now, from (3.8), we have

$$\begin{aligned} E_k^{\alpha-1} &= \frac{1}{2} \|(\alpha - 1)(z^{k-1} - z^*) + t_{k-1} (\Delta z_{k-1} + (1 - \eta)Q^k)\|^2 \\ &\quad + \eta(\alpha - 1)t_{k-1} \left(\langle Q^k, z^k - z^* \rangle + \frac{1}{2} \|Q^k\|^2 \right) + \frac{(1 - \eta)\eta t_{k-1}^2}{2} \|Q^k\|^2 \\ &= \frac{(\alpha - 1)^2}{2} \|z^{k-1} - z^*\|^2 + \frac{t_{k-1}^2}{2} \|\Delta z_{k-1} + (1 - \eta)Q^k\|^2 + \frac{1}{2} (1 - \eta)\eta t_{k-1}^2 \|Q^k\|^2 \\ &\quad - \frac{1}{2} (\alpha - 1)(1 - \eta)t_{k-1} \|Q^k\|^2 - (\alpha - 1)(1 - \eta)t_{k-1} \langle Q^k, \Delta z_{k-1} \rangle \\ &\quad + (\alpha - 1)t_{k-1} \left[\langle Q^k, z^k - z^* \rangle + \frac{1}{2} \|Q^k\|^2 + \langle \Delta z_{k-1}, z^k - z^* \rangle - \frac{1}{2} \|\Delta z_{k-1}\|^2 \right]. \end{aligned}$$

Combining (3.16) (hence the last line has a limit), (ii) (hence the second line converges to 0), and that $\lim_{k \rightarrow \infty} \|z^{k-1} - z^*\|^2$ and $\lim_{k \rightarrow \infty} E_k^{\alpha-1}$ exist as proved above, we deduce that

$$\lim_{k \rightarrow +\infty} t_{k-1}^2 \left(\frac{1}{2} \|\Delta z_{k-1} + (1 - \eta)Q^k\|^2 + \frac{1}{2} (1 - \eta)\eta \|Q^k\|^2 \right) \in \mathbb{R}.$$

Observe now that the summability claims of (ii) already give us that

$$(3.17) \quad \liminf_{k \rightarrow +\infty} k^2 \|\Delta z_k\|^2 = 0, \quad \text{and} \quad \liminf_{k \rightarrow +\infty} k^2 \|Q^k\|^2 = 0.$$

Thus these are actually limits. To transfer the rates from z^k to x^k , we use the relation $x^k = Q^{k+1} + z^{k+1}$, yielding the last two rates in (iv). Using Cauchy–Schwarz and boundedness of $(z^k)_{k \in \mathbb{N}}$, we also deduce the first. This concludes the proof. $\square$

Remark 3.4 (The discrete gap function rate). *While the convergence rates for the discrete velocity $\|x^k - x^{k-1}\|$ and the fixed-point residual $\|x^k - T(x^k)\|$ are relatively natural in this context, the rate*

$$(3.18) \quad \langle x^k - T(x^k), T(x^k) - z^* \rangle + \frac{1}{2} \|x^k - T(x^k)\|^2 = o(k^{-1}) \quad \text{as } k \rightarrow +\infty,$$

may initially appear less intuitive. However, as we will see in Section 4, it plays a key role in applications to primal-dual-type algorithms, where it serves as a bound for the primal-dual gap, see Example 4.1. See also Remark 2.3 for its continuous-time counterpart.

Remark 3.5 (Non-asymptotic rates). *In this work, we primarily focus on asymptotic convergence rates, i.e., the behavior as $k \rightarrow +\infty$. Nevertheless, for each admissible parameter choice and each quantity of interest, we can also derive corresponding non-asymptotic rates with explicit constants. Let us also*

emphasize here the dependency of E_k^λ on η by writing $E_k^{\lambda,\eta}$. Embarking from (3.12), we have $E_{k+1}^{\alpha-1,\eta} \leq E_k^{\alpha-1,\eta}$ for all $k \geq 1$ such that $t_k \geq \alpha - 1$. If we now only consider $\sigma \geq \alpha - 1$, we obtain $E_{k+1}^{\alpha-1,\eta} \leq E_k^{\alpha-1,\eta}$ for all $k \geq 1$, leading to the following explicit bounds:

$$(3.19) \quad \begin{aligned} \langle x^{k-1} - T(x^{k-1}), T(x^{k-1}) - z^* \rangle + \frac{1}{2} \|x^{k-1} - T(x^{k-1})\|^2 &\leq \frac{E_1^{\alpha-1,\eta}}{\eta(\alpha-1)t_{k-1}}, \\ \|x^{k-1} - T(x^{k-1})\|^2 &\leq \frac{2E_1^{\alpha-1,\eta}}{\eta(1-\eta)t_{k-1}^2}, \end{aligned}$$

where $E_1^{\alpha-1,\eta}$ is a constant that only depends on the initial conditions, on z^* , and the algorithms parameters, including η —see (3.8) for the precise definition. A particularly illustrative situation occurs when $\alpha = 2$, $\eta = \frac{1}{2}$, and $\sigma = 1$, which yields $t_0 = 0$ and thus

$$(3.20) \quad \|x^{k-1} - T(x^{k-1})\|^2 \leq \frac{4\|z^0 - z^*\|^2}{t_{k-1}^2} = \frac{4\|T(x^{-1}) - x^*\|^2}{(k-1)^2} \quad \text{for all } k \geq 2.$$

A more in-depth investigation of the two bounds in (3.19), including the influence of the parameter η , is left for future work.

3.3. Connection to Existing Algorithms. Algorithm 1 offers a flexible and unifying framework that is connected to a broad class of existing methods. In this section, we highlight a few notable special cases and related algorithms to place them more clearly within the literature.

3.3.1. Connection to [21]. As we mentioned in the introduction, Algorithm 1 is a generalization of the scheme (2.8) in [21]. The latter can be recovered by setting $\sigma = \alpha + 1$ and $\theta = \frac{\alpha}{2}$, which gives, renaming x^k as x^{k+1} , for all $k \geq 1$:

$$(3.21) \quad x^{k+1} = \left(1 - \frac{\alpha}{2(k+\alpha)}\right) x^k + \frac{\alpha}{2(k+\alpha)} T(x^k) + \left(1 - \frac{\alpha}{k+\alpha}\right) (T(x^k) - T(x^{k-1})),$$

with initial points $x^0, x^1 \in H$. This is also a special case of [21, Algorithm 2.1], which can be recovered by introducing a step size $s \in (0, 1]$ and considering $T_s := (1-s)I + sT$, which is now s -averaged and has the same fixed points as T . Indeed, substituting T by T_s in (3.21) we get for all $k \geq 1$:

$$(3.22) \quad x^{k+1} = \left(1 - \frac{s\alpha}{2(k+\alpha)}\right) x^k + \frac{(1-s)k}{k+\alpha} (x^k - x^{k-1}) + \frac{s\alpha}{2(k+\alpha)} T(x^k) + \frac{sk}{k+\alpha} (T(x^k) - T(x^{k-1})).$$

Conversely, if T is ζ -averaged, we shall consider (3.22) with $s \in (0, 1/\zeta]$, which is equivalent to applying (3.21) with T replaced by $sT + (1-s)I$.

3.3.2. Connection to Halpern-type Methods. Similarly to the continuous-time dynamics discussed in Section 2.2.1, Algorithm 1 reduces to an *anchored* method when we set $\theta = 1$. Firstly, when $\alpha = 2$, $\sigma = \alpha$, and $x^0 = T(x^{-1})$, it is clear that Algorithm 1 coincides with the *momentum form* of the Optimal Halpern Method (OHM) [47, 56], and in turn with Kim's accelerated proximal point algorithm [44], since the latter two are equivalent, see [62, Appendix B]. In the edge case of setting θ equal to 1, it turns out that Algorithm 1 is indeed equivalent to anchored-type methods for any $\alpha \geq 2$ and $\sigma > 0$, as we will show in the following result.

Proposition 3.6. *Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1 with $\alpha \geq 2$, $\theta = 1$ and $\sigma > 0$. Then, $(x^k)_{k \in \mathbb{N}}$ satisfies:*

$$(3.23) \quad x^{k+1} = \varepsilon_k v + (1 - \varepsilon_k) T(x^k), \quad \text{for } k \geq 0,$$

where $\varepsilon_k := \frac{\alpha-1}{k+\sigma}$ and $v := \frac{t_0}{\alpha-1} (x^0 - T(x^{-1})) + T(x^{-1})$.

Proof. Multiplying the update rule of Algorithm 1 by $t_{k+1} := k + \sigma$, we get

$$\begin{aligned} (k + \sigma)(x^{k+1} - x^k) &= T(x^k) - x^k + (k + \sigma - \alpha)(T(x^k) - T(x^{k-1})) \\ \iff t_{k+1}(x^{k+1} - x^k) + x^k & \\ & - \left(T(x^k) + (k + \sigma - 1)(T(x^{k+1}) - T(x^k)) \right) + (\alpha - 1)(T(x^k) - T(x^{k-1})) = 0 \\ \iff (t_{k+1}x^{k+1} - t_kx^k) - (t_{k+1}T(x^k) - t_kT(x^{k-1})) + (\alpha - 1)(T(x^k) - T(x^{k-1})) &= 0. \end{aligned}$$

Summing the above equation from 0 to k , dividing the result by t_{k+1} and rearranging terms we get the claim. $\square$

Note that in the edge case $\theta = 1$ we cannot expect to obtain little- o rates for fixed-point residuals, and the convergence of the iterates is in fact in the strong topology, but is established with completely different techniques than those we have applied in Section 3.2: The former rely on the Grönwall's lemma (see, e.g., [19]), the latter on Opial's lemma.

3.3.3. Connection to Inertial KM and Nesterov's Method. The Fast-KM method is not the first attempt to introduce a Nesterov-type acceleration for iterative methods governed by nonexpansive operators. The earliest such attempt dates back to 2001, with the work of Attouch and Alvarez in [4]. Starting from an implicit discretization of the celebrated Heavy Ball with Friction (HBF) dynamical system governed by a maximal monotone operator, they arrived at the following numerical algorithm, for a firmly nonexpansive operator T :

$$(3.24) \quad x^{k+1} = T(x^k) + \alpha_k(T(x^k) - T(x^{k-1})), \quad \text{for all } k \geq 0,$$

with $\alpha_k \in (0, 1)$. Note that if $\alpha_k := 1 - \frac{\alpha}{k+\sigma}$ with $\sigma > 0$, $\alpha \geq 3$, and $T := I - \tau \nabla f$ for a L -smooth function $f: H \rightarrow \mathbb{R}$ and $\tau \in (0, \frac{1}{L}]$, then (3.24) coincides with the *Ravine* formulation of the celebrated Nesterov Accelerated Gradient method [52, 11], which is known to yield the optimal convergence rate in convex optimization. While (3.24) can still provide weakly converging sequences for general (firmly) nonexpansive operators—specifically if $\alpha_k \leq \bar{\alpha} < \frac{1}{3}$, cf. [4, Proposition 2.1]—whether it achieves fast convergence rates in this general setting remains unclear. A first negative answer was provided only recently in [50], in a broader context involving iteration-dependent operators and more general parameter choices. Specifically, in the context of (3.24), [50, Theorem 4(ii)] shows that if α_k satisfies a specific growth condition—implying, in particular, that $\sup_{k \in \mathbb{N}} \alpha_k < 1$ —then one only obtains

$$(3.25) \quad \min_{0 \leq k \leq n} \|x^k - T(x^k)\| \leq M \frac{\text{dist}(x^0, \text{Fix } T)}{\sqrt{n}}, \quad \text{for all } n \geq 1,$$

for some constant $M > 0$. This result is *ergodic* and of suboptimal order $\mathcal{O}(n^{-\frac{1}{2}})$.

Comparing (3.24) with Algorithm 1, we see that (3.24) could, in principle, be recovered from Algorithm 1 by replacing the vanishing regularization term $\frac{\theta}{k+\sigma}$ with 1, and by capping α_k at a value strictly smaller than 1. In the continuous-time model (2.1), this would correspond to taking $b \equiv 1$ and a *non*-vanishing damping function, respectively. However, doing so would worsen the rate to (3.25). Thus, we learn that letting $\alpha_k \rightarrow 1$ and using a vanishing regularization parameter $\frac{\theta}{k+\sigma}$ are both essential for achieving fast convergence rates for general nonexpansive operators.

3.3.4. Connection to Tran-Dinh's Method. In [61], Tran-Dinh proposed a method that is able to achieve a fast little- o rate for solving root-finding problems of the form:

$$(3.26) \quad \text{Find } x \in H \text{ such that: } 0 = G(x),$$

where $G: H \rightarrow H$ is $\frac{1}{L}$ -cocoercive, for $L > 0$. The method in [61, Theorem 3] is formulated as

$$(3.27) \quad x^{k+1} = x^k + \bar{\theta}_k(x^k - x^{k-1}) - \bar{\eta}_k(G(x^k) - \bar{\gamma}_k G(x^{k-1})), \quad \text{for } k \geq 0,$$

with parameter sequences defined, for $\bar{\gamma} \in (0, 1)$, $\omega > 2$, and all $k \geq 0$, as

$$(3.28) \quad \bar{\theta}_k := \frac{k+1}{k+2\omega+2}, \quad \bar{\eta}_k := \frac{\bar{\gamma}}{L} \frac{k+\omega+1}{k+2\omega+2}, \quad \text{and} \quad \bar{\gamma}_k := \frac{\bar{\gamma}}{L} \frac{\bar{\theta}_k}{\bar{\eta}_k}.$$

It turns out that (3.27) is a special case of Algorithm 1. Indeed, considering the operator $T := I - \frac{\bar{\gamma}}{L}G$, solving (3.26) is equivalent to finding a fixed point of T , T being $\frac{\bar{\gamma}}{2}$ -averaged. If we now apply Algorithm 1 to find a fixed point of T setting $\sigma := 2\omega + 2$, $\alpha := 2\omega + 1$ and $\theta := \omega$, we obtain exactly (3.27) with the parameter choices (3.28). Interestingly, Theorem 3.3 allows us to further extend the parameter choices proposed in [61] to $\omega \geq \frac{1}{2}$, instead of $\omega \geq 2$, and $\bar{\gamma} \in (0, 2]$ instead of $\bar{\gamma} \in (0, 1)$, as well as to establish a last-iterate rate on the gap function.

4. APPLICATION TO SPLITTING METHODS WITH PRECONDITIONERS

The goal here is apply our results to operator splitting methods for solving monotone inclusion problems, i.e. finding a zero of one or sum of maximally monotone operators. We follow the algorithmic framework introduced in [24]. For a maximal monotone operator $A: H \rightarrow 2^H$, we introduce a *preconditioner* M , i.e., a self-adjoint positive semidefinite bounded linear operator $M: H \rightarrow H$, which induces a Hilbertian (semi-)inner product $\langle u, v \rangle_M := \langle u, Mv \rangle$ and associated (semi) norm $\|u\|_M := \sqrt{\langle u, u \rangle_M}$ for all $u, v \in H$. Given A and M , we consider the resolvent:

$$(4.1) \quad J_{M^{-1}A} := (I + M^{-1}A)^{-1} = (M + A)^{-1}M.$$

It is worth noting that without further hypothesis, $J_{M^{-1}A}$ is not necessarily single-valued nor defined everywhere on H . On the other hand, this representation is helpful in practice, and evaluating $J_{M^{-1}A}$ can be facilitated by the right choice of M . As first remarked by He and Yuan in [43], this is in fact the case for primal-dual-type methods:

Example 4.1 (PDHG). *Let $f: H \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g: K \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, convex and lower semicontinuous functions on the Hilbert spaces H and K , and $L: H \rightarrow K$ a linear bounded map. The Primal-Dual-Hybrid-Gradient (PDHG) method by Chambolle and Pock [31] is a classical scheme to find a solution to the following saddle point problem*

$$(4.2) \quad \min_{x \in H} \max_{y \in K} \mathcal{L}(x, y) := f(x) + \langle Lx, y \rangle - g^*(y),$$

where g^* is the Legendre–Fenchel conjugate of g . This has enormous applications in applied mathematics, see, e.g., [32]. The PDHG iteration takes the form

$$(4.3) \quad \begin{cases} x^{k+1} = \text{prox}_{\tau_1 f_1}(x^k - \tau_1 L^T y^k), \\ y^{k+1} = \text{prox}_{\tau_2 f_2^*}(y^k + \tau_2 L(2x^{k+1} - x^k)). \end{cases}$$

Under a mild qualification condition, the generated sequence $u^k := (x^k, y^k)$ converges weakly to a saddle point of $\mathcal{L}$ as soon as $\tau_1 \tau_2 \|L\|^2 < 1$, see [43, 24]. Following He and Yuan [43], the method can be understood a fixed-point iterations with respect to $J_{M^{-1}A}$ with

$$(4.4) \quad A := \begin{bmatrix} \partial f & L^T \\ -L & \partial g^* \end{bmatrix}, \quad \text{and} \quad M := \begin{bmatrix} \frac{1}{\tau_1} I & -L^T \\ -L & \frac{1}{\tau_2} I \end{bmatrix}.$$

The operator $J_{M^{-1}A}$ can in fact be easily computed due to the block-lower triangular structure of $A + M$ (recall the formula (4.1) or see [24, Section 3] for the detailed computations), and the resulting fixed-point iteration coincides with (4.3).

The step-size condition $\tau_1 \tau_2 \|L\|^2 < 1$ is equivalent to the positive definiteness of M [24], which makes $M^{-1}A$ a maximal monotone operator on the Hilbert space H endowed with the norm induced by M . Being an instance of the (preconditioned) proximal point algorithm [24], the rate on the fixed-point residual is

$$(4.5) \quad \|u^k - J_{M^{-1}A}(u^k)\|_M = o(k^{-\frac{1}{2}}) \quad \text{as } k \rightarrow +\infty.$$

A second classical convergence measure for (4.3) is the primal-dual gap⁴, which we define as

$$(4.6) \quad \mathcal{G}_{u^*}(x, y) := \mathcal{L}(x, y^*) - \mathcal{L}(x^*, y), \quad \text{for all } (x, y) \in H \times K.$$

⁴More precisely, the primal-dual gap should be defined as $\sup_{u^* \in \text{zer } A} \mathcal{G}_{u^*}(x, y)$. However, as $\text{zer } A$ might be unbounded, this quantity is not easily controllable, leading to notions such as *restricted primal-dual gap* [31] or to further assumptions such as boundedness of $\text{zer } A$. In this example, we keep the presentation simple and avoid taking the sup.

The quantity that in Remark 3.4 we referred to as the gap function (3.20) is indeed an upper bound for (4.6). Specifically, using convexity and the definition of A in (4.4), we get

$$(4.7) \quad \begin{aligned} \mathcal{G}_{u^*}(x^{k+1}, y^{k+1}) &\leq \langle u^k - J_{M^{-1}A}(u^k), J_{M^{-1}A}(u^k) - u^* \rangle_M \\ &\leq \langle u^k - J_{M^{-1}A}(u^k), J_{M^{-1}A}(u^k) - u^* \rangle_M + \frac{1}{2} \|u^k - J_{M^{-1}A}(u^k)\|_M^2. \end{aligned}$$

Note that $u^k - J_{M^{-1}A}(u^k) = u^k - u^{k+1} \in M^{-1}A(u^{k+1})$, so the first inner product (4.7) is non-negative in this case. Secondly, while the right-hand side of (4.7) vanishes at the rate $o(k^{-\frac{1}{2}})$, it is also summable, leading, using convexity of $\mathcal{G}_{u^*}$ to the classical $\mathcal{O}(k^{-1})$ ergodic rate for (4.6), see [32].

By allowing the preconditioner M to be *degenerate*, i.e., with $\ker M \neq \emptyset$, Bredies and Sun in [26, 27], and later [24] recognized that the class of methods representable as iterations with respect to degenerate preconditioners can be greatly broadened to include methods such as the Douglas–Rachford Splitting (DRS) [48] and its extensions [25].

4.1. Preliminaries on Degenerate Proximal Point. The notion of monotonicity and cocoercivity can be extended to a *preconditioned* setting, even with a degenerate preconditioner $M: H \rightarrow H$. For a maximal monotone operator A and a preconditioner M we shall consider the operator $M^{-1}A$ as a composition of the multi-valued operators M^{-1} (pre-image) and A , defined as: $v \in M^{-1}A(u)$ if and only if $Mv \in A(u)$. It is easy to check that $M^{-1}A$ is monotone with respect to the semi-inner product induced by M . We then say that $M^{-1}A$ is *M-monotone*. Likewise, its resolvent $J_{M^{-1}A}$ is *M-cocoercive* in the sense that

$$(4.8) \quad \langle v - v', u - u' \rangle_M \geq \|v - v'\|_M^2, \quad \text{for all } (u, v), (u', v') \in \text{gra } J_{M^{-1}A}.$$

Due to the degeneracy of the preconditioner, the operator $J_{M^{-1}A}$ is not necessarily single-valued and could even have empty domain. To avoid such pathological cases, we suppose $\text{dom } J_{M^{-1}A} = H$. A further hypothesis we shall require is that the operator

$$(4.9) \quad (M + A)^{-1} \text{ is Lipschitz over } \text{Im } M,$$

which makes $J_{M^{-1}A}$ a well-defined Lipschitz operator on H , since indeed $J_{M^{-1}A} = (M + A)^{-1}M$. Additionally, it can be shown to satisfy the fundamental property:

$$(4.10) \quad \|J_{M^{-1}A}(u) - J_{M^{-1}A}(u')\|^2 \leq L \|u - u'\|_M^2, \quad \text{for all } u, u' \in H.$$

We refer to fixed-point iterations with respect to $J_{M^{-1}A}$ as *preconditioned proximal point algorithm*, which for $\theta \in (0, 2)$ and $u^0 \in H$, reads as

$$(4.11) \quad u^{k+1} = u^k + \theta(J_{M^{-1}A}(u^k) - u^k), \quad \text{for all } k \in \mathbb{N}.$$

The convergence guarantees of (4.11) are given in [24, Theorem 2.9]. Specifically, if (4.9) holds, then $(u^k)_{k \in \mathbb{N}}$ and $(J_{M^{-1}A}(u^k))_{k \in \mathbb{N}}$ converge weakly to the same zero of A with the rate:

$$(4.12) \quad \|u^k - J_{M^{-1}A}(u^k)\|_M = o(k^{-\frac{1}{2}}) \quad \text{as } k \rightarrow +\infty.$$

An appealing property of positive semidefinite preconditioners is that evaluating $J_{M^{-1}A}$ automatically produces a quotient space operation. Suppose that M has closed range and consider a factorization of M as $M = CC^*$, with $C: D \rightarrow H$ one-to-one from a Hilbert space D ($D \cong \text{Im } M$) to H . We say that CC^* is an *onto decomposition* of M . In this case, one can prove that C^* is onto [24, Proposition 2.3] and the operator $C^* \triangleright A := (C^* A^{-1} C)^{-1}$ is maximal monotone [24, Theorem 2.13]. The resolvents of $M^{-1}A$ and of $C^* \triangleright A$ are tightly connected: If M has closed range and $M = CC^*$ is onto, then

$$(4.13) \quad C^* J_{M^{-1}A}(u) = J_{C^* \triangleright A}(C^* u), \quad \text{for all } u \in H.$$

This is particularly useful as it reveals that the evaluation of $J_{M^{-1}A}$ is somewhat equivalent to evaluating $J_{C^* \triangleright A}$, which is a resolvent of a maximal monotone operator and is defined on D , which is isomorphic to $\text{Im } M$; see [33, Theorem 3.3.3] for a proof. This implies that the sequence $w^k := C^* u^k \in D$ is a classical proximal point method with respect to the maximal monotone operator $C^* \triangleright A$:

$$(4.14) \quad w^{k+1} = w^k + \theta(J_{C^* \triangleright A}(w^k) - w^k) \quad \text{as } k \rightarrow +\infty.$$

However, although $J_{M^{-1}A}(u^k) = (M + A)^{-1}(Cw^k)$ from (4.1) and $M = CC^*$, the weak convergence of $(u^k)_{k \in \mathbb{N}}$ and $(J_{M^{-1}A}(u^k))_{k \in \mathbb{N}}$, in infinite-dimensional spaces, does not follow directly from the that of $(w^k)_{k \in \mathbb{N}}$. For clarity, we present a prototypical example:

Example 4.2 (Douglas–Rachford [27, 24]). *Let $A_1, A_2: H \rightarrow 2^H$ be two maximal monotone operators on H and assume that $\text{zer}(A_1 + A_2) \neq \emptyset$. Following [24], the DRS method for finding a zero of $A_1 + A_2$ can be understood as iterating $J_{M^{-1}A}$ with the operators:*

$$(4.15) \quad A := \begin{bmatrix} A_1 & I & -I \\ -I & A_2 & I \\ I & -I & 0 \end{bmatrix}, \quad \text{and} \quad M := \begin{bmatrix} I & -I & I \\ -I & I & -I \\ I & -I & I \end{bmatrix} = \begin{bmatrix} I \\ -I \\ I \end{bmatrix} \underbrace{\begin{bmatrix} I & -I & I \end{bmatrix}}_{:=C^*}.$$

Indeed, since $M + A$ is block lower-triangular, evaluating $J_{M^{-1}A} = (M + A)^{-1}M$ only requires accessing the resolvents of A_1 and A_2 , we refer to [24] for the detailed computation. If $u^k := (x_1^k, x_2^k, v^k) \in H^3$ is such that $u^{k+1} := J_{M^{-1}A}(u^k)$ the DRS method can be obtained via the substitution $w^k := x_1^k - x_2^k + v^k$, which yields for all $k \in \mathbb{N}$:

$$(4.16) \quad \begin{cases} x_1^{k+1} = J_{A_1}(w^k), \\ x_2^{k+1} = J_{A_2}(2x_1^{k+1} - w^k), \\ w^{k+1} = w^k + x_2^{k+1} - x_1^{k+1}. \end{cases} \quad \text{and} \quad J_{M^{-1}A}(u^k) := \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ w^k + 2(x_2^{k+1} - x_1^{k+1}) \end{pmatrix}.$$

This corresponds to taking $w^k = C^*u^k$, which defines a fixed-point iteration with respect $J_{C^* \triangleright A}$, the operator $C^* \triangleright A$ now being maximal monotone. It follows from [24, Theorem 2.9] that $(w^k)_{k \in \mathbb{N}}$ weakly converges to some $w^* \in \text{zer } C^* \triangleright A$, with the rate

$$(4.17) \quad \|x_2^{k+1} - x_1^{k+1}\| = \|w^k - J_{C^* \triangleright A}(w^k)\| = \|u^k - J_{M^{-1}A}(u^k)\|_M = o(k^{-\frac{1}{2}}), \quad \text{as } k \rightarrow +\infty.$$

Additionally, $(J_{M^{-1}A}(u^k))_{k \in \mathbb{N}}$ weakly converges to a zero of A , yielding that the solution estimates $(x_1^{k+1})_{k \in \mathbb{N}}$ and $(x_2^{k+1})_{k \in \mathbb{N}}$ (a.k.a. shadow sequences) weakly converge to the same zero of $A_1 + A_2$.

Note that the weak convergence of the shadow sequences remained an open problem for almost 30 years until Svaiter proposed a convergence proof in [60]. The analysis of iterations associated to (4.1) with a degenerate preconditioner allows to recover this result from general principles applicable to a broad range of algorithms; see [24].

Example 4.3 (Graph Douglas–Rachford [25, 33, 2]). *The graph-Douglas–Rachford Splitting method (Graph-DRS) is an extension of the Douglas–Rachford method to solve instances of:*

$$(4.18) \quad \text{Find } x \in H : \quad 0 \in A_1(x) + \cdots + A_N(x),$$

with $A_1, \dots, A_N: H \rightarrow 2^H$ maximal monotone operators. The method only requires picking two real matrices $Z, \hat{Z} \in \mathbb{R}^{N \times (N-1)}$ and a step size $\tau > 0$, with $\ker Z^T = \text{span}\{\mathbf{1}\} \subset \ker \hat{Z}^T$ where $\mathbf{1} := (1, \dots, 1)^T \in \mathbb{R}^N$. Then, denoting by $L := ZZ^T$, $\hat{L} := \hat{Z}\hat{Z}^T$ and $d := \text{diag}(L + \hat{L})$, the update rule reads as:

$$(4.19) \quad \begin{cases} x_i^{k+1} = J_{\frac{\tau}{d_i} A_i} \left(-\frac{2}{d_i} \sum_{h=1}^{i-1} (L_{hi} + \hat{L}_{hi}) x_h^{k+1} + \frac{1}{d_i} \sum_{j=1}^{N-1} Z_{ij} w_j^k \right), & \text{for all } i \in \llbracket 1, N \rrbracket, \\ w_j^{k+1} = w_j^k - \sum_{i=1}^N Z_{ij} x_i^{k+1}, & \text{for all } j \in \llbracket 1, N-1 \rrbracket. \end{cases}$$

This method generalizes a number of splitting methods and can in fact be shown to encompass all averaged frugal resolvent splitting methods with minimal lifting in the sense of Ryu, see [2] and the references therein. Similarly to Example 4.2, it can be understood in the framework of degenerate proximal point methods with respect to $J_{C^* \triangleright A}$ for a maximal monotone operator $A: H \rightarrow 2^H$, and a degenerate preconditioner $M := CC^*$, this time with $C^*: H \rightarrow H^{N-1}$ and $H := H^{2N-1}$, cf. [25] for their precise (but technical) definition. Moreover, $x^{k+1} := (x_1^{k+1}, \dots, x_N^{k+1})$ are the first N components of $(M + A)^{-1}(Cw^k)$, where $w^k := (w_1^k, \dots, w_{N-1}^k)$.

While in (4.17) the fixed-point residual coincides with the difference of the two solution estimates, for (4.19) it controls a natural dispersion functional between the components of $\mathbf{x}^{k+1}$, i.e., their variance:

$$(4.20) \quad \text{Var}(\mathbf{x}^{k+1}) := \frac{1}{N} \sum_{i=1}^N \|x_i^{k+1} - \bar{x}^{k+1}\|^2 \leq \frac{1}{\lambda_1 N} \|\mathbf{w}^k - J_{C^* \triangleright A}(\mathbf{w}^k)\|^2 = o(k^{-1}), \quad \text{as } k \rightarrow +\infty.$$

where $\bar{x}^{k+1} := \frac{1}{N} \sum_{i=1}^N x_i^{k+1}$, and λ_1 is the smallest nonzero eigenvalue of L , see [25, Proposition 3.8] or [33, Lemma 5.2.1] for a proof of the inequality in (4.20). Additionally, thanks to [24, Theorem 2.9], all the solution estimates $(x_i^{k+1})_{k \in \mathbb{N}}$ for $i \in \llbracket 1, N \rrbracket$ converge weakly to the same solution to (4.18).

4.2. Fast Degenerate Preconditioned Proximal Point Algorithms. Once the preconditioned resolvent $J_{M^{-1}A}$ is formed, e.g., as in Examples 4.1, 4.2, and 4.3, we can employ it in conjunction to Algorithm 1 to study accelerated variants of degenerate preconditioned proximal point iterations (4.11). Denoting by $\theta_k := \frac{\theta}{k+\sigma}$ and $\alpha_k := 1 - \frac{\alpha}{k+\sigma}$, we can consider, for $s \in (0, 2]$, the M -averaged operator $T := I + s(J_{M^{-1}A} - I)$ and apply Algorithm 1, yielding, for all $k \geq 0$,

$$(4.21) \quad u^{k+1} = u^k + \alpha_k(1-s)(u^k - u^{k-1}) + s\theta_k J_{M^{-1}A}(u^k) + s\alpha_k \left(J_{M^{-1}A}(u^k) - J_{M^{-1}A}(u^{k-1}) \right).$$

Similarly to the non-accelerated setting, if M has closed range, using (4.13), the sequence $w^k := C^*u^k \in D$ can be easily seen to define an instance of Algorithm 1 with respect to $\tilde{T} := I + s(J_{C^* \triangleright A} - I)$, the latter being $\frac{s}{2}$ -averaged in the classical sense since $C^* \triangleright A$ is maximal monotone. Therefore, convergence rates and weak convergence of $(w^k)_{k \in \mathbb{N}}$ follow immediately from Theorem 3.3. However, using this argument, the weak convergence of $(J_{M^{-1}A}(u^k))_{k \in \mathbb{N}}$ does not follow from that of $(w^k)_{k \in \mathbb{N}}$ since the convergence of $(w^k)_{k \in \mathbb{N}}$ is in the weak topology and $J_{M^{-1}A}(u^k) = (M+A)^{-1}(Cw^k)$ is typically nonlinear. The following lemma captures the argument that allows us to close this important gap and extend the analysis to general preconditioners, with non-necessarily closed range. The mechanism underlying this result can be traced back to [24], although it was not explicitly established there.

Lemma 4.4 (Opial in the Degenerate Setting). *Let A be a maximal monotone operator with $\text{zer } A \neq \emptyset$, and M a self-adjoint positive semidefinite preconditioner such that $(M+A)^{-1}$ is Lipschitz over $\text{Im } M$ and $J_{M^{-1}A}$ has full domain. Let $(u^k)_{k \in \mathbb{N}}$ be a sequence, and suppose that:*

- (i) *For all $u^* \in \text{Fix } J_{M^{-1}A}$ the limit $\lim_{k \rightarrow +\infty} \|u^k - u^*\|_M$ exists;*
- (ii) *and $\lim_{k \rightarrow +\infty} \|u^k - J_{M^{-1}A}(u^k)\|_M = 0$.*

Then $J_{M^{-1}A}(u^k) \rightharpoonup u^ \in \text{zer } A$ as $k \rightarrow +\infty$. If, additionally, M has closed range and $M = CC^*$ is onto, then for $w^k := C^*u^k$:*

$$w^k \rightharpoonup w^* \in \text{zer } C^* \triangleright A, \quad \text{and} \quad (A+M)^{-1}(Cw^k) = J_{M^{-1}A}(u^k) \rightharpoonup u^* \in \text{zer } A, \quad \text{as } k \rightarrow +\infty.$$

Proof. For brevity, let us denote by $v^k := J_{M^{-1}A}(u^k)$ for all $k \in \mathbb{N}$. From (i), we deduce that the sequence $(Mu^k)_{k \in \mathbb{N}}$ is bounded. Therefore, since $(M+A)^{-1}C$ is Lipschitz and $J_{M^{-1}A} = (M+A)^{-1}M$, which is then single-valued everywhere on H , also $(v^k)_{k \in \mathbb{N}}$ is bounded. Since $M(u^k - v^k) \in A(v^k)$ for all $k \in \mathbb{N}$, $\lim_{k \rightarrow +\infty} \|M(u^k - v^k)\| = 0$ from point (ii), and A is maximal monotone (in particular its graph is weak-to-strong sequentially closed [14, Proposition 20.38]), every weak cluster point of $(v^k)_{k \in \mathbb{N}}$ belongs to $\text{zer } A$. Now we show that, in fact, it admits only one weak cluster point and thus $(v^k)_{k \in \mathbb{N}}$ converges weakly to it. Let $(v^{k_j})_{j \in \mathbb{N}}$ and $(v^{k_i})_{i \in \mathbb{N}}$ be two weakly converging subsequences and $v^*, v^{**} \in H$ their respective weak cluster points. Note that, from points (i) and (ii), for all $u^* \in \text{zer } A$, $\lim_{k \rightarrow +\infty} \|v^k - u^*\|_M = \lim_{k \rightarrow +\infty} \|u^k - u^*\|_M$, which we denote by $\ell(u^*)$. Suppose by contradiction that $\|v^* - v^{**}\|_M \neq 0$. Thus,

$$(4.22) \quad \begin{aligned} \ell(v^{**}) &= \lim_{i \rightarrow +\infty} \|v^{k_i} - v^{**}\|_M^2 = \lim_{i \rightarrow +\infty} (\|v^{k_i} - v^*\|_M^2 + 2\langle v^{k_i} - v^*, v^* - v^{**} \rangle_M + \|v^* - v^{**}\|_M^2) \\ &= \|v^* - v^{**}\|_M^2 + \lim_{i \rightarrow +\infty} \|v^{k_i} - v^*\|_M^2 > \ell(v^*), \end{aligned}$$

where we have used that $v^{k_i} \rightharpoonup v^*$. Reversing the role of v^* and v^{**} , we also get $\ell(v^{**}) < \ell(v^*)$, hence the desired contradiction, which yields $Mv^* = Mv^{**}$. Hence, $J_{M^{-1}A}(v^*) = J_{M^{-1}A}(v^{**})$, and, since $v^*, v^{**} \in \text{zer } A = \text{Fix } J_{M^{-1}A}$, we obtain $v^* = v^{**}$.

If M has closed range and $M = CC^*$ is onto, using (4.13), we obtain, since C^* is onto, that for each $w^* \in \text{zer } C^* \triangleright A$ there exists $u^* \in \text{zer } A$ such that $C^*u^* = w^*$, and thus

$$(4.23) \quad \|w^k - w^*\|^2 = \|C^*u^k - C^*u^*\|^2 = \|u^k - u^*\|_M^2$$

admits a limit as $k \rightarrow +\infty$. Similarly, $\lim_{k \rightarrow +\infty} \|w^k - J_{C^* \triangleright A}(w^k)\|^2 = 0$. The classical Opial's lemma yields that $w^k \rightharpoonup w^* \in \text{zer } C^* \triangleright A$. Using now that $(A + M)^{-1}(Cw^k) = J_{M^{-1}A}(u^k)$ for all $k \in \mathbb{N}$, which follows from (4.1), and that $J_{M^{-1}A}(u^k) \rightharpoonup u^*$, we conclude the proof. $\square$

Naturally, a similar result holds in continuous time. We now describe how this lemma can be used to establish the convergence of several accelerated splitting methods in a unified manner, thus demonstrating how Lemma 4.4 extends the analysis of [24] well beyond standard proximal point algorithms. It is important to emphasize, however, that Lemma 4.4 does not provide any information about the asymptotic behavior of the sequence $(u^k)_{k \in \mathbb{N}}$ itself, which requires a separate, method-specific analysis. We opted to present the adaptation of Theorem 3.3, i.e., the one for algorithmic schemes. The same can be done for Theorem 2.2, i.e., for continuous-time dynamics, but we leave this to the reader.

Theorem 4.5. *Let A be a maximal monotone operator with $\text{zer } A \neq \emptyset$, and M a self-adjoint positive semidefinite preconditioner such that $(M + A)^{-1}$ is Lipschitz over $\text{Im } M$ and $J_{M^{-1}A}$ has full domain. Let $(u^k)_{k \in \mathbb{N}}$ be the sequence⁵ generated by Algorithm 1 with $T = (1 - s)I + sJ_{M^{-1}A}$, $s \in (0, 2]$, $\alpha > 2$, $\theta \in (1, \alpha - 1)$, and $\sigma > 0$. Then:*

(i) *We have, for all $u^* \in \text{zer } A$,*

$$\begin{aligned} \sum_{k \in \mathbb{N}} \langle u^k - J_{M^{-1}A}(u^k), J_{M^{-1}A}(u^k) - u^* \rangle_M &< +\infty, \\ \sum_{k \in \mathbb{N}} k \|u^k - J_{M^{-1}A}(u^k)\|_M^2 &< \infty, \quad \sum_{k \in \mathbb{N}} k \|u^k - u^{k-1}\|_M^2 < +\infty. \end{aligned}$$

(ii) *$(J_{M^{-1}A}(u^k))_{k \in \mathbb{N}}$ converges weakly to some $u^* \in \text{zer } A$ as $k \rightarrow +\infty$. Additionally, if $s = 1$, also $(u^k)_{k \in \mathbb{N}}$ converges weakly to u^* .*

(iii) *We have the following fast rates of convergence:*

$$\begin{aligned} \langle u^k - J_{M^{-1}A}(u^k), J_{M^{-1}A}(u^k) - u^* \rangle_M &= o(k^{-1}), \\ \|u^k - u^{k-1}\|_M &= o(k^{-1}) \quad \text{and} \quad \|u^k - J_{M^{-1}A}(u^k)\|_M = o(k^{-1}) \quad \text{as } k \rightarrow +\infty. \end{aligned}$$

Suppose furthermore that M has closed range and that $M = CC^*$ is onto with $C: D \rightarrow H$. Then, $w^k := C^*u^k$ satisfies (i)-(iii) with $M^{-1}A$ replaced by $C^* \triangleright A$ and seminorms replaced by norms (in D), and $w^k \rightharpoonup C^*u^*$ as $k \rightarrow +\infty$.

Proof. The proof of point (i) and point (iii) follow exactly the same lines of Theorem 3.3 by replacing every norm by M -seminorms, every inner product with M -semi-inner products and setting $Q = s(I - J_{M^{-1}A})$. In fact, the only property utilized is the nonexpansiveness of T , which in this context translates as the M -nonexpansiveness of $sJ_{M^{-1}A} + (1 - s)I$ for $s \in (0, 2]$ see [24, Remark 2.7]. Specifically, (i) can be proven by following the arguments in the proof Theorem 3.3(ii). This yields, in particular, that

$$(4.24) \quad \lim_{k \rightarrow +\infty} \|u^k - J_{M^{-1}A}(u^k)\|_M = 0.$$

Similarly, statement (iii) can be shown following the proof of Theorem 3.3(iv).

Regarding (ii), following the proof of statement Theorem 3.3(iii), we get that $\lim_{k \rightarrow +\infty} \|u^k - u^*\|_M$ exists for any $u^* \in \text{zer } A$. We are now in the setting of the degenerate Opial lemma (Lemma 4.4), which yields weak convergence of $(J_{M^{-1}A}(u^k))_{k \in \mathbb{N}}$ to some $u^* \in \text{zer } A$. To conclude, we pick $s = 1$ and establish the weak convergence of $(u^k)_{k \in \mathbb{N}}$ to u^* by showing that $\varphi_k := \frac{1}{2} \|u^k - J_{M^{-1}A}(u^k)\|^2$ vanishes as $k \rightarrow +\infty$.

⁵We denote by $(u^k)_{k \in \mathbb{N}}$ the sequence generated by Algorithm 1 to avoid confusion with $\mathbf{x}^k = (x_1^k, \dots, x_N^k)$ generated by Graph-DRS discussed in Section 4.3.3.

Using (4.21) and denoting for brevity $y^k := u^k - J_{M^{-1}A}(u^k)$ and $\Delta_k := J_{M^{-1}A}(u^{k+1}) - J_{M^{-1}A}(u^k)$ for all $k \geq 0$, we get

$$\begin{aligned} y^{k+1} &= u^k + \theta_k(J_{M^{-1}A}(u^k) - u^k) + \alpha_k(J_{M^{-1}A}(u^k) - J_{M^{-1}A}(u^{k-1})) - J_{M^{-1}A}(u^{k+1}) \\ &= (1 - \theta_k)y^k + \alpha_k(J_{M^{-1}A}(u^k) - J_{M^{-1}A}(u^{k-1})) + J_{M^{-1}A}(u^k) - J_{M^{-1}A}(u^{k+1}). \\ &= (1 - \theta_k)y^k + \alpha_k\Delta_{k-1} - \Delta_k. \end{aligned}$$

Using the Young inequality with weight $\xi_k > 0$ and (4.10) we can write

$$\begin{aligned} (4.25) \quad \varphi_{k+1} - \varphi_k &\leq -(1 - (1 - \theta_k)^2(1 + \xi_k))\varphi_k + (1 + \xi_k^{-1})(\|\Delta_{k-1}\|^2 + \|\Delta_k\|^2) \\ &\leq -(1 - (1 - \theta_k)^2(1 + \xi_k))\varphi_k + (1 + \xi_k^{-1})L(\|u^k - u^{k-1}\|_M^2 + \|u^{k+1} - u^k\|_M^2). \end{aligned}$$

Picking now, for all k large enough such that $\theta_k < 2$,

$$(4.26) \quad \xi_k = \theta_k \frac{2 - \theta_k}{2(1 - \theta_k)^2}, \quad \text{we get: } \varepsilon_k := (1 - (1 - \theta_k)^2(1 + \xi_k)) = \frac{\theta_k}{2}(2 - \theta_k).$$

Therefore, since $(\varepsilon_k)_{k \in \mathbb{N}} \notin \ell^1$ by definition of θ_k and $(k\|u^k - u^{k-1}\|_M^2)_{k \in \mathbb{N}}$ is summable from point (i), it follows from (4.25) that $\varphi_k \rightarrow 0$. Thus, $u^k \rightharpoonup u^*$ weakly as well. $\square$

Remark 4.6. If $\alpha = \sigma = 2$, we are in the setting of the OHM method (cf. Section 3.3.2), which produces a sequence that converges strongly. In this case, if M has closed range, there is indeed no need to rely on a different analysis with degenerate preconditioners since the map $(M + A)^{-1}C$ is Lipschitz, and the convergence of $(u^k)_{k \in \mathbb{N}}$ would follow from that of $(w^k)_{k \in \mathbb{N}}$.

4.3. Application to Splitting Algorithms. In this section, we illustrate how Theorem 4.5 can be applied to show complete convergence results for accelerated splitting algorithms. We consider only three notable examples, but the analysis applies to a broad range of algorithms such as those in [33, Chapter 4] or in [24, Section 3]. See also the recent contribution [59] for an application to ADMM, using similar techniques to ours.

4.3.1. Fast PDHG. We start from the PDHG method by Chambolle–Pock [31] described in Example 4.1. Applying (4.21) with $s \in (0, 2]$, and $J_{M^{-1}A}$ build as in Example 4.1, with step sizes satisfying $\tau_1\tau_2\|L\|^2 \geq 1$, we obtain a variant of the PDHG algorithm (4.3) enhanced with momentum. Denoting by $\tilde{u}^{k+1} := (\tilde{x}^{k+1}, \tilde{y}^{k+1}) := J_{M^{-1}A}(u^k)$ and recalling the definition of the primal-dual gap $\mathcal{G}_{u^*}$ in (4.7), Theorem 4.5(iii), yields the following fast rates on the last iterate:

$$(4.27) \quad \|u^k - J_{M^{-1}A}(u^k)\|_M = o(k^{-1}), \quad \text{and} \quad \mathcal{G}_{u^*}(\tilde{x}^{k+1}, \tilde{y}^{k+1}) = o(k^{-1}) \quad \text{as } k \rightarrow +\infty,$$

for all $u^* := (x^*, y^*)$ saddle point of (4.2), thus improving the corresponding $o(k^{-\frac{1}{2}})$ rates for the standard PDHG method. Additionally, from Theorem 4.5(ii), we obtain the weak convergence of $(\tilde{u}^k)_{k \in \mathbb{N}}$ and (if $s = 1$) of $(u^k)_{k \in \mathbb{N}}$ to the same saddle point of $\mathcal{L}$. Note that in the non-degenerate setting $\tau_1\tau_2\|L\|^2 > 1$ the application of OHM, i.e., Algorithm 1 in the edge case $\alpha = \sigma = 2$, to PDHG was proposed in [44].

4.3.2. Fast Douglas–Rachford Splitting. In the special case of Douglas–Rachford iterations according to Example 4.2, applying (4.21) we obtain, for $\alpha > 2$, $\theta \in (1, \alpha - 1)$, the *Fast-Douglas–Rachford* method:

$$(4.28) \quad \begin{cases} x_1^{k+1} = J_{A_1}(w^k), \\ x_2^{k+1} = J_{A_2}(2x_1^{k+1} - w^k), \\ w^{k+1} = w^k + \alpha_k(w^k - w^{k-1}) + s\theta_k(x_2^{k+1} - x_1^{k+1}) + s\alpha_k(x_2^{k+1} - x_2^k - (x_1^{k+1} - x_1^k)). \end{cases}$$

Theorem 4.5, together with the degenerate preconditioned proximal point formulation of the method described in Example 4.2, allows us to improve the worst-case rate (4.17) from $o(k^{-\frac{1}{2}})$ to

$$(4.29) \quad \|x_2^{k+1} - x_1^{k+1}\| = o(k^{-1}) \quad \text{as } k \rightarrow +\infty,$$

while keeping the weak convergence of the shadow sequences $(x_1^{k+1})_{k \in \mathbb{N}}$ and $(x_2^{k+1})_{k \in \mathbb{N}}$ to the same zero of $A_1 + A_2$. Let us underscore that while (4.29) and the weak convergence of $(w^k)_{k \in \mathbb{N}}$ follow from classical

theory and were established in [21] in the corresponding special case, the weak convergence of the shadow sequences requires dedicated analysis. Here, it follows from Theorem 4.5(ii).

4.3.3. Fast Graph Douglas–Rachford Splitting. In this section, we show the application of Theorem 4.5 to Graph-DRS presented in Example 4.3. Applying (4.21), we get, for $k \geq 0$ and $\mathbf{w}^{-1}, \mathbf{w}^0 \in H^{N-1}$, $\alpha > 2$, and $\theta \in (0, 1)$:

$$(4.30) \quad \begin{cases} x_i^{k+1} = J_{\frac{\tau}{d_i} A_i} \left(-\frac{2}{d_i} \sum_{h=1}^{i-1} (L_{hi} + \hat{L}_{hi}) x_h^{k+1} + \frac{1}{d_i} \sum_{j=1}^{N-1} Z_{ij} w_j^k \right), & \text{for all } i \in \llbracket 1, N \rrbracket, \\ w_j^{k+1} = w_j^k + \alpha_k (w_j^k - w_j^{k-1}) - s\theta_k \sum_{i=1}^N Z_{ij} x_i^{k+1} - s\alpha_k \sum_{i=1}^N Z_{ij} (x_i^{k+1} - x_i^k), & \text{for all } j \in \llbracket 1, N-1 \rrbracket. \end{cases}$$

Also in this case, Theorem 4.5, together with the degenerate preconditioned proximal point representation of the method mentioned in Example 4.3, allows us to improve the rates in (4.20) yielding

$$(4.31) \quad \text{Var}(\mathbf{x}^{k+1}) = o(k^{-2}) \quad \text{as } k \rightarrow +\infty,$$

keeping the weak convergence of the solution estimates $(x_i^{k+1})_{i \in \mathbb{N}}$ to the same solution $x^* \in H$ of (4.18).

5. NUMERICAL EXPERIMENTS

In this section, we present our numerical experiments, which were performed in Python on a 12th-Gen. Intel(R) Core(TM) i7-1255U, 1.70–4.70 GHz laptop with 16 Gb of RAM. The code is available for reproducibility at <https://github.com/echenn/fast-km>.

5.1. Toy problems. In these first numerical experiments, we consider two main instances of T :

- We consider an optimization problem that can be written as

$$(5.1) \quad \min_{x \in \mathbb{R}^d} f_1(x) + \cdots + f_N(x),$$

for $f_1, \dots, f_N: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ proper, convex and lower semicontinuous such that $\text{prox}_{\tau f_i}$ can be easily computed for each $\tau > 0$. In this case, we take T as given in (4.19) applied to the first-order optimality conditions of (5.1), and thus use Fast Graph-DRS initialized with $\mathbf{w}^{-1} = \mathbf{w}^0 = 0$.

- As in [21], for dimension $d \in \mathbb{N}$ even, we consider $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined by $T(x) := J_{\tau \Sigma}(x)$, where $\tau = 10^{-1}$ and Σ is the skew-symmetric (hence monotone) matrix given block-wise by:

$$\Sigma := \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix},$$

where I is now the identity operator in $\mathbb{R}^{d/2}$ and 0 represents the zero operator in $\mathbb{R}^{d/2}$. In this case, we use Algorithm 1 initialized with $x^{-1} = x^0 = \mathbf{1}$.

5.2. The influence of parameters. As we introduced two degrees of flexibility in Algorithm 1 consisting in the parameter $\theta \in (1, \alpha - 1)$ and $\sigma > 0$, we first numerically investigate the influence of these two parameters separately. We consider the two instances of T described in Section 5.1. In the first one, we set $d = 2$, $N = 2$ and thus apply the Fast-Douglas–Rachford method according to Section 4.3.2 to:

$$f_1(x) := 10^{-3}|x|, \quad \text{and} \quad f_2(x) := \frac{1}{2} \text{dist}^2(x, \mathbb{B}), \quad x \in \mathbb{R}^2,$$

where $\mathbb{B}$ is the unit euclidean ball centered at $(1, 1)^T$. The two proximity operators are easily accessed through, soft-thresholding and averaged projection onto $\mathbb{B}$, respectively; see [14, Example 14.5 and Proposition 24.8(vii)]. In the second one, we set $d = 10$.

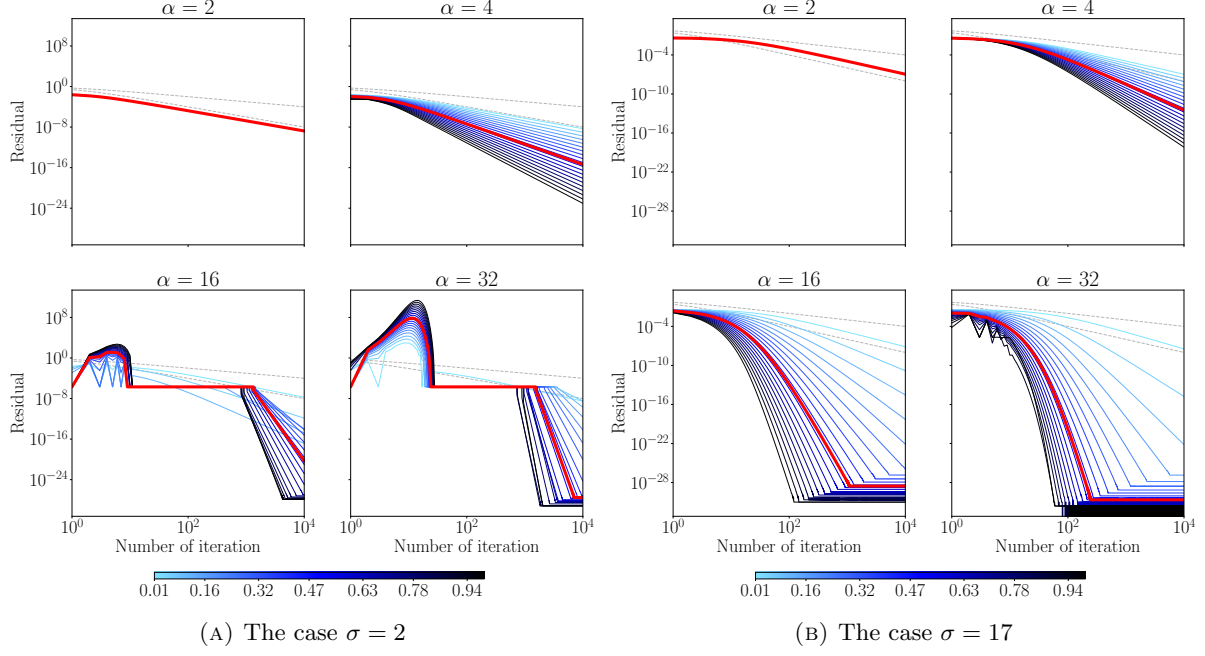


FIGURE 1. Results of Section 5.2.1 with the first instance of T described in Section 5.1: Comparison of different choices of $\eta \in (0, 1)$, with corresponding colors indicated in the colorbar. The case $\eta = \frac{1}{2}$ is highlighted in red.

5.2.1. *Influence of θ .* In this section, we perform a systematic comparison between $\eta = \frac{1}{2}$, which was introduced (when $\sigma = \alpha + 1$) in [21] and the choices $\eta \in (0, 1)$, interpolating the parameter ranges $\theta \in (1, \alpha - 1)$. Specifically, we consider $\alpha = 2, 4, 16, 32$ and two different choices of σ , $\sigma = 2$ and $\sigma = 17$. Note in particular that the special case of [21] is considered with $\alpha = 32$. When $\alpha = 2$ the methods with different values of η all coincide (since $\theta = 1$ for all $\eta \in (0, 1)$) and, when $\sigma = 2$, the resulting method reduces to OHM, see Section 3.3.2. We plot the results in Figure 1 and Figure 2.

Our numerical experiments demonstrate that Algorithm 1 performs particularly well compared to the baseline choice $\eta = \frac{1}{2}$. As shown in Figure 1, performances improve noticeably when $\theta > \frac{\alpha}{2}$; in fact, the larger the relaxation parameter, the faster the method tends to converge. However, this behavior does not occur in Figure 2, which corresponds to the case where T is the resolvent of a skew-symmetric matrix. In that setting, the baseline choice $\eta = \frac{1}{2}$ yields the best performance.

We suspect that this contrasting behavior stems from an intrinsic trade-off between minimizing the gap function and minimizing the fixed-point residual. An inspection of the constants in (3.19) suggests that larger relaxation parameters favor gap-function reduction, while their effect on the fixed-point residual appears to be strongly problem-specific. In particular, when T is the resolvent of a skew-symmetric operator, the gap function is identically zero, making the residual the sole quantity of interest.

Eventually, we can also observe that while OHM ($\alpha = \sigma = 2$), i.e., the accelerated Douglas–Rachford method in the sense of Kim [44], achieves an *optimal* non-asymptotic convergence rate [62], Algorithm 1, which exhibits *order-optimal* but little- o rates, can perform much better in practice.

5.2.2. *The influence of σ .* Keeping in mind (3.5), the parameter σ is related to the starting time t_0 in the continuous dynamics (2.1). Recall in fact that the discrete time is defined by $t_k = k - 1 + \sigma$ so $t_0 = \sigma - 1$. We test 20 different choices of σ spaced evenly from 1 to 100 and take the four choices $\alpha = 2, 4, 16, 32$. We put particular emphasis on the choice $\sigma = \alpha$, which, for $\alpha = 2$ leads to OHM [47, 62]. We consider only the operator T corresponding to the second instance of Section 5.1. The results are depicted in Figure 3. We refrain from displaying the results for the first instance of T in Section 5.1 since the results

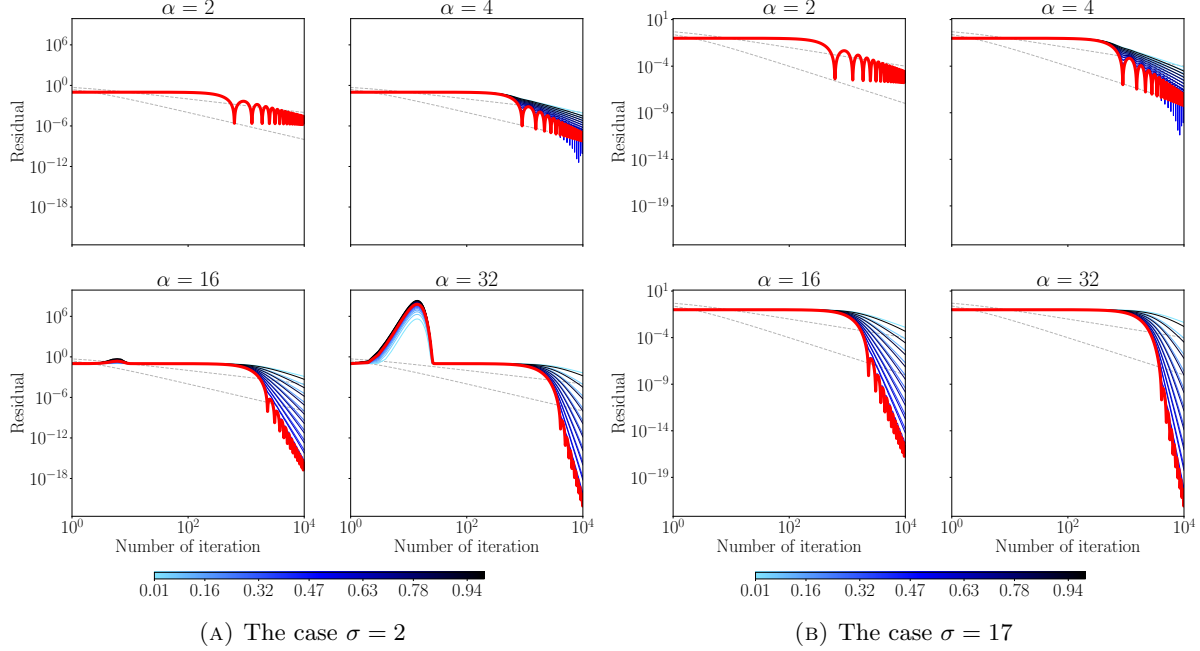


FIGURE 2. Results of Section 5.2.1 with T as described the second instance of T described in Section 5.1: Comparison of different choices of $\eta \in (0, 1)$, with corresponding colors indicated in the colorbar. $\eta = \frac{1}{2}$ is highlighted in red.

therein are almost indistinguishable. Regarding the tuning of the parameter σ , our experiments suggest indeed setting $\sigma = \alpha$.

5.3. Cooling Heuristics. Motivated by the insights of Sections 5.2.1, and 5.2.2, we propose a “cooling” heuristics⁶, which consists in gradually increasing the parameter α along the iterations until a certain upper bound $\alpha_{\max}$ is reached: $\alpha_0 < \alpha_1 < \dots < \alpha_{\lfloor \text{maxit}/2 \rfloor} = \alpha_{\max}$. Specifically, we consider two different strategies: a *linear* and a *logarithmic* parameter scaling. In either case, we set $\alpha_0 > 0$, $\alpha_{\max} = 100\alpha_0$, and we fix a maximum number of iterations **maxit**. Then, in first case, at each iteration we increase α with a fixed step size, in such a way that $\alpha_{\max}$ is reached for $k = \lfloor \text{maxit}/2 \rfloor$. In the second case, the increase is defined in a logarithmic scale.

To set up our numerical experiment, we consider $\alpha = 2, 4, 16, 32$, and the operator T as given in the second choice in Section 5.1. For each α , we compare the performance of the vanilla Krasnoselskii–Mann (KM) method, Fast-KM with $\eta = \frac{1}{2}$ and $\alpha = \sigma$, and Fast-KM with linear and logarithmic scaling in the cooling heuristics and $\eta = \frac{1}{2}$, $\alpha_0 = \alpha$, abbreviated (Fast-KM-LNC) and (Fast-KM-LGC), respectively. Recall that $\alpha_{\max} = 100\alpha_0$. The results are shown in Figure 4. A notable insight from these experiment is the beneficial convergence speed of Algorithm 1 with large parameters α , but without incurring into the undesirable initial instability state; see also [1] who thoroughly studied the choice of α for the ODE (1.4) when $\beta(t) \equiv 0$, $b(t) \equiv 1$ and Q is the gradient of a convex function.

5.4. Application to Realistic Problems. In this section, we test Algorithm 1 to solve optimization problems arising from realistic settings. We start with an application to Optimal Transport (OT), and in particular, to the *Beckmann minimal flow formulation* of the Wasserstein 1 distance, where DRS proved particularly effective [53, 17, 29]. In the second experiment, we consider the geometric median problem, a standard nonsmooth benchmark especially for distributed optimization.

⁶The name is borrowed from Consensus Based Optimization literature [39].

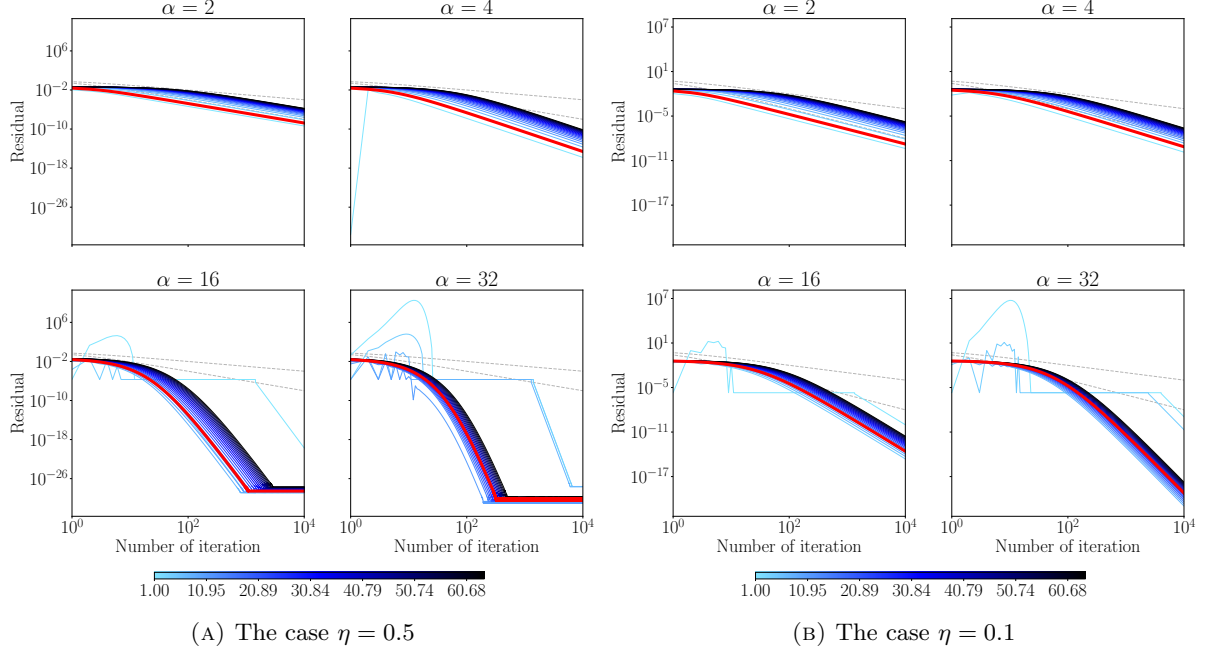


FIGURE 3. Results of Section 5.2.2: Comparison between different choices of $\sigma \in (0, 100)$. $\sigma = \alpha$ is highlighted in red.

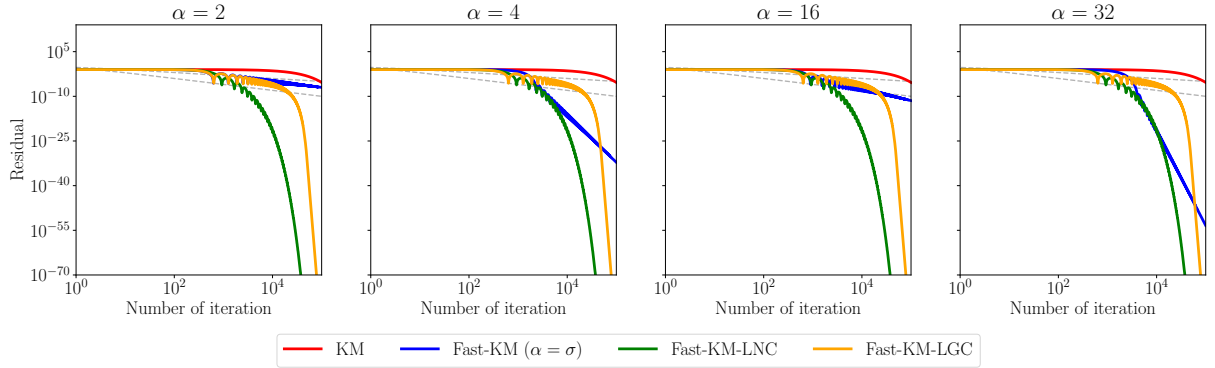


FIGURE 4. Results of Section 5.3: Testing cooling heuristics.

In these two experiments, we compare the performance of i) the vanilla Krasnoselskii–Mann (KM) method, ii) Algorithm 1 with $\eta = 0.5$, iii) Algorithm 1 with $\eta = 0.9$ and iii) Algorithm 1 with a linear cooling heuristics as introduced in Section 5.3. The results are reported in Figure 5.

5.4.1. Optimal Transport. In this experiment, we tackle a high dimensional OT problem, which is usually solved in its regularized primal form using the celebrated Sinkhorn Algorithm [18]. Consider a square grid $\Omega := \{(i, j) \mid i, j \in \llbracket 1, p \rrbracket\}$, with $p \in \mathbb{N}$, and two discrete probability measures μ , and ν on Ω , i.e., $\mu, \nu \in \mathbb{R}_+^n$ with $n = p^2$ such that $\sum_{i=1}^n \mu_i = \sum_{i=1}^n \nu_i = 1$. Denoted by div the discrete divergence operator with Neumann boundary conditions on Ω defined trough finite differences (see, e.g., [29]), we aim at solving the following convex problem:

$$(5.2) \quad \min_{\sigma \in \mathbb{R}^{n \times 2}} \|\sigma\|_{2,1}, \quad \text{s.t.:} \quad \text{div } \sigma = \mu - \nu,$$

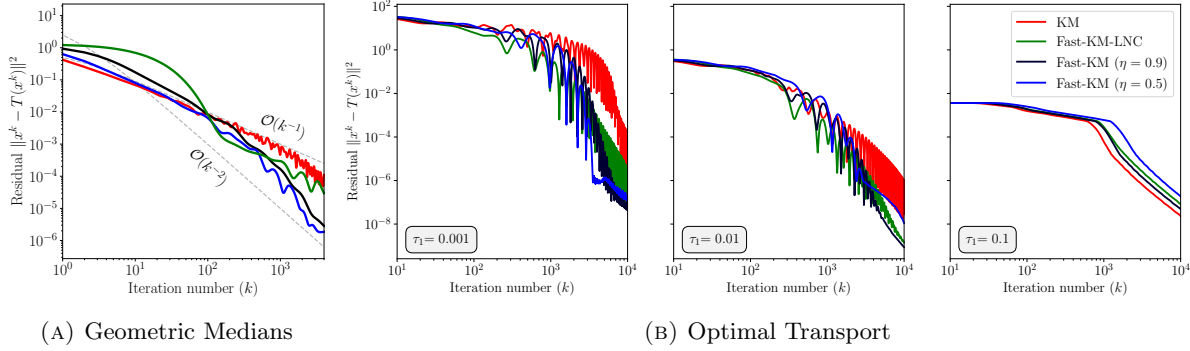


FIGURE 5. Results of Section 5.4. In Figure 5a, we solve a geometric median problem (Section 5.4.2), using Graph-DRS as the fixed-point operator T , cf. Section 4.3.3 and Example 4.3. In Figure 5b, we solve an ℓ_1 large scale optimal transportation problem, using Chambolle–Pock as the fixed-point operator T , cf. Section 4.3.1 and Example 4.1.

where $\|\sigma\|_{2,1}$ is the *group-lasso* penalty, i.e., $\|\sigma\|_{2,1} := \sum_{i=1}^n \|\sigma_i\|$, and for each $i \in \llbracket 1, n \rrbracket$, $\sigma_i \in \mathbb{R}^2$. Specifically, we consider $p = 100$, which represents a relatively large scale instance for OT.

This *basis pursuit*-type problem can be addressed by applying the Chambolle–Pock method (cf. Example 4.1) with $f(\sigma) := \|\sigma\|_{2,1}$, $g(\sigma) := \iota_{\{\mu - \nu\}}(\sigma)$, and $L := \text{div}$. These two functions are indeed *prox-friendly*. $\text{prox}_{\tau f}$ is the group soft-thresholding operator (see [14, Example 14.5]), while $\text{prox}_{\tau g}$ is just the constant function mapping to $\mu - \nu$. We apply the Fast-Chambolle–Pock solver (Section 4.3.1) to solve (5.2) with $\tau_1 \tau_2 = 10^{-1}$, which is a lower bound for $\|L\|^2$ and $\tau_1 = 10^{-3}, 10^{-2}$ and 10^{-1} .

As can be seen from Figure 5b, the Fast-Chambolle–Pock method emerges here as a great practical alternative to the plain Chambolle–Pock: the method outperformed the standard approach with KM, especially in regimes with high oscillations, and confirmed once again that large relaxation parameters are indeed beneficial in practice.

5.4.2. Solving the Geometric Median Problem. We now turn to computing the geometric median of N sample points in $\mathbb{R}^d$. Given a sample $\{x_i\}_i$ of N points in $\mathbb{R}^d$ we aim at solving

$$(5.3) \quad \min_{x \in \mathbb{R}^d} \|x - x_1\| + \cdots + \|x - x_N\|.$$

To solve (5.3), we consider the graph-DRS operator in (4.19) splitting (5.3) in N terms. We consider, $N = 100$ and $d = 100$. All the terms $x \mapsto \|x - x_i\|$ are in fact prox-friendly with proximity operators that can be computed by simply applying the soft-thresholding operator suitably shifted by x_i . We apply (4.30) with $\hat{Z} = 0$ and build an arbitrary Z such that $\ker Z^* = \text{span}\{\mathbf{1}\}$ and show the results in Figure 5a. The same conclusions drawn from the OT experiment in Section 5.4.2 remain true here.

6. CONCLUSION

In this paper, we introduced a generalized version of the Fast-KM method proposed in [21], incorporating two additional degrees of flexibility through the parameters $\theta \in (1, \alpha - 1)$ and $\sigma > 0$. We provided a new convergence proof based on a semi-implicit discretization of a continuous-time model. Notably, our approach encompasses the case $\alpha = 2$, which corresponds to the Optimal Halpern Method, thereby unifying the convergence analysis of these two acceleration mechanisms. We demonstrated how these results can be applied to operator splitting methods potentially with preconditioners, offering critical convergence guarantees for the so-called shadow sequences. Furthermore, our numerical experiments showcased the benefits afforded by the added flexibility of the method. This paper opens up new intriguing questions, among which: i) what is the tight non-asymptotic rate for Algorithm 1? ii) Does Algorithm 1 converge when $\theta = \alpha - 1$? We leave this for future research.

APPENDICES

A. Discrete Calculus Rules. Let $a := (a^k)_{k \in \mathbb{N}}$, $b := (b^k)_{k \in \mathbb{N}}$, and $c := (c^k)_{k \in \mathbb{N}}$, be three real sequences. The product of two sequences a and b is denoted by $ab := (a^k b^k)_{k \in \mathbb{N}}$. As usual, if $a = b$, we denote $a^2 = ab$. Recall that $\Delta a_k := a^{k+1} - a^k$ for all $k \in \mathbb{N}$. Therefore, for all $k \geq 0$ it holds

$$\begin{aligned} \Delta(ab)_k &= \Delta a_k b^k + a^{k+1} \Delta b_k, \\ \Delta(abc)_k &= \Delta a_k b^k c^k + a^{k+1} \Delta b_k c^k + a^{k+1} b^{k+1} \Delta c_k, \\ \Delta a_k^2 &= 2(a^{k+1} - a^k)a^k + (\Delta a_k)^2 = 2(a^{k+1} - a^k)a^{k+1} - (\Delta a_k)^2. \end{aligned} \quad (6.1)$$

This formalism can be readily extended to sequences in a Hilbert space H , with products replaced by inner products. For instance, if $(z^k)_{k \in \mathbb{N}}$ is a sequence in H , and $\zeta^k := \frac{1}{2}\|z^k\|^2$ for all $k \geq 0$, can write

$$\Delta \zeta_k = \langle \Delta z_k, z^k \rangle + \frac{1}{2} \|\Delta z_k\|^2 = \langle \Delta z_k, z^{k+1} \rangle - \frac{1}{2} \|\Delta z_k\|^2. \quad (6.2)$$

Similar rules as those in (6.1) apply for products of sequences in H and in $\mathbb{R}$, for instance if $(\varepsilon^k)_{k \in \mathbb{N}}$ is a real sequence, then for all $k \geq 0$

$$\Delta(\varepsilon z)_k = \Delta \varepsilon_k z^k + \varepsilon^{k+1} \Delta z_k = \Delta z_k \varepsilon^k + z^{k+1} \Delta \varepsilon_k. \quad (6.3)$$

B. Missing Proofs. The following proofs were postponed from the main part of the paper.

Proof of Lemma 2.1. For the sake of notation, we also abbreviate $E_t^\lambda := E^\lambda(t)$, $\dot{x}_t := \dot{x}(t)$, $x_t := x(t)$, $v_t^\lambda := v^\lambda(t)$, $\delta_t := \delta(t)$, and $\xi_t := \xi(t)$. Additionally, instead of writing $\frac{d}{dt}$ we write ∂_t . Using Leibniz rule, we get for almost all $t \geq t_0$:

$$\begin{aligned} \partial_t E_t^\lambda &= \dot{\delta}_t \langle Q_t, x_t - x^* \rangle + \delta_t \langle \partial_t Q_t, x_t - x^* \rangle + \delta_t \langle Q_t, \dot{x}_t \rangle \\ &\quad + \frac{\dot{\xi}_t}{2} \|Q_t\|^2 + \xi_t \langle \partial_t Q_t, Q_t \rangle + c \langle x_t - x^*, \dot{x}_t \rangle + \langle v_t^\lambda, \dot{v}_t^\lambda \rangle. \end{aligned} \quad (6.4)$$

Let us compute the last inner product $\langle v_t^\lambda, \dot{v}_t^\lambda \rangle$. For $\dot{v}_t^\lambda$ we have, using constitutive equation (2.1), that for almost all $t \geq t_0$:

$$\begin{aligned} \dot{v}_t^\lambda &= (\lambda + 1)\dot{x}_t + (1 - \eta)\beta_t Q_t + t \left(-\frac{\alpha}{t} \dot{x}_t - \beta_t \partial_t Q_t - b_t Q_t + (1 - \eta)\dot{\beta}_t Q_t + (1 - \eta)\beta_t \partial_t Q_t \right) \\ &= (\lambda + 1 - \alpha)\dot{x}_t + (1 - \eta)\beta_t Q_t - t\eta\beta_t \partial_t Q_t + t((1 - \eta)\dot{\beta}_t - b_t)Q_t \\ &= -(\alpha - 1 - \lambda)\dot{x}_t - t \left(b_t - (1 - \eta)\dot{\beta}_t - (1 - \eta)\frac{\beta_t}{t} \right) Q_t - t\beta_t \eta \partial_t Q_t. \end{aligned}$$

Denoting by $w_t := b_t - (1 - \eta)\dot{\beta}_t - (1 - \eta)\frac{\beta_t}{t}$, the inner product read for almost all $t \geq t_0$ as

$$\begin{aligned} \langle v_t^\lambda, \dot{v}_t^\lambda \rangle &= -\lambda(\alpha - 1 - \lambda) \langle x_t - x^*, \dot{x}_t \rangle - (\alpha - 1 - \lambda)t \|\dot{x}_t\|^2 - (1 - \eta)(\alpha - 1 - \lambda)t\beta_t \langle Q_t, \dot{x}_t \rangle \\ &\quad - \lambda t w_t \langle x_t - x^*, Q_t \rangle - t^2 w_t \langle \dot{x}_t, Q_t \rangle - t^2 \beta_t w_t (1 - \eta) \|Q_t\|^2 \\ &\quad - \eta t \lambda \beta_t \langle x_t - x^*, \partial_t Q_t \rangle - \eta t^2 \beta_t \langle \dot{x}_t, \partial_t Q_t \rangle - t^2 \beta_t^2 (1 - \eta) \eta \langle Q_t, \partial_t Q_t \rangle. \end{aligned} \quad (6.5)$$

Plugging (6.5) into (6.4), using the definition of δ_t , ξ_t and c from (2.4), and canceling out the appropriate terms, we obtain for almost all $t \geq t_0$ that

$$\begin{aligned} \partial_t E_t^\lambda &= \left(\dot{\delta}_t - t\lambda w_t \right) \langle Q_t, x_t - x^* \rangle - \eta t^2 \beta_t \langle \dot{x}_t, \partial_t Q_t \rangle \\ &\quad + \left(\delta_t - (1 - \eta)(\alpha - 1 - \lambda)t\beta_t - t^2 w_t \right) \langle Q_t, \dot{x}_t \rangle + \left(\frac{\dot{\xi}_t}{2} - (1 - \eta)t^2 \beta_t w_t \right) \|Q_t\|^2 - (\alpha - 1 - \lambda)t \|\dot{x}_t\|^2. \end{aligned} \quad (6.6)$$

Let us study the quadratic form in the second line of (6.6) separately⁷. Plugging in the definitions of δ_t , ξ_t in (2.4) and of b in (2.2), we get, first, $w_t = (\alpha - 1)\eta \frac{\beta_t}{t}$, hence,

$$\begin{aligned} \delta_t - (1 - \eta)(\alpha - 1 - \lambda)t\beta_t - t^2w_t &= -t\beta_t(\alpha - 1 - \lambda), \\ \frac{\dot{\xi}_t}{2} - (1 - \eta)t^2\beta_tw_t &= \beta_t t^2(1 - \eta)\eta \left(\dot{\beta}_t + \frac{(2 - \alpha)}{t}\beta_t \right), \end{aligned}$$

so the quadratic form simplifies to:

$$\begin{aligned} &(\delta_t - (1 - \eta)(\alpha - 1 - \lambda)t\beta_t - t^2w_t) \langle Q_t, \dot{x}_t \rangle + \left(\frac{\dot{\xi}_t}{2} - (1 - \eta)t^2\beta_tw_t \right) \|Q_t\|^2 - (\alpha - 1 - \lambda)t\|\dot{x}_t\|^2 \\ &= -t\beta_t(\alpha - 1 - \lambda) \langle Q_t, \dot{x}_t \rangle + \beta_t t^2(1 - \eta)\eta \left(\dot{\beta}_t + \frac{(2 - \alpha)}{t}\beta_t \right) \|Q_t\|^2 - (\alpha - 1 - \lambda)t\|\dot{x}_t\|^2 \\ &= -\frac{t}{2}(\alpha - 1 - \lambda)\|\dot{x}_t + \beta_t Q_t\|^2 - \frac{t}{2}(\alpha - 1 - \lambda)\|\dot{x}_t\|^2 \\ &\quad + \beta_t t^2 \left((1 - \eta)\eta \left(\dot{\beta}_t + \frac{\beta_t}{t}(2 - \alpha) \right) + \frac{1}{2}(\alpha - 1 - \lambda)\frac{\beta_t}{t} \right) \|Q_t\|^2. \end{aligned}$$

Plugging this and the definition of ξ_t and of w_t into (6.6), we get the claim. $\square$

Proof of Lemma 3.2. Let $k \geq 1$. Apply the discrete Leibniz rule (see Appendix A) to get:

$$\begin{aligned} \Delta E_k^\lambda &= \Delta \delta_k \langle Q^k, z^k - z^* \rangle + \delta_{k+1} \langle \Delta Q_k, z^k - z^* \rangle + \delta_{k+1} \langle Q^{k+1}, \Delta z_k \rangle + \Delta \left(\frac{\delta}{2} \|Q\|^2 \right)_k \\ &\quad + \frac{1}{2} \Delta \xi_k \|Q^k\|^2 + \xi_{k+1} \langle \Delta Q_k, Q^{k+1} \rangle - \frac{1}{2} \xi_{k+1} \|\Delta Q_k\|^2 \\ (6.7) \quad &\quad + c \langle \Delta z_{k-1}, z^k - z^* \rangle - \frac{c}{2} \|\Delta z_{k-1}\|^2 \\ &\quad + \langle \Delta v_k, v^{k+1} \rangle - \frac{1}{2} \|\Delta v_k\|^2. \end{aligned}$$

Following the continuous-time approach, we first write Δv_k only in terms of $z^{k-1} - z^*$, Q^k and ΔQ_k , then compute $\langle \Delta v_k, v^{k+1} \rangle$ and finally plug everything into ΔE_k^λ . This will give the necessary cancellations. Using that $\Delta t_k = 1$ and the update of Algorithm 1 (see (3.5)), we get

$$\begin{aligned} \Delta v_k &= \lambda \Delta z_{k-1} + \Delta t_{k-1} (\Delta z_{k-1} + (1 - \eta)Q^k) + t_k (\Delta^2 z_{k-1} + (1 - \eta)\Delta Q_k) \\ (6.8) \quad &= \lambda \Delta z_{k-1} + (\Delta z_{k-1} + (1 - \eta)Q^k) + t_k \left(-\frac{\alpha}{t_k} \Delta z_{k-1} - \Delta Q_k - \frac{\theta}{t_k} Q^k + (1 - \eta)\Delta Q_k \right) \\ &= -(\alpha - 1 - \lambda)\Delta z_{k-1} - (\theta - (1 - \eta))Q^k - t_k \eta \Delta Q_k \\ &= -(\alpha - 1 - \lambda)\Delta z_{k-1} - \eta(\alpha - 1)Q^k - t_k \eta \Delta Q_k. \end{aligned}$$

Let us now evaluate $\langle \Delta v_k, v^{k+1} \rangle$, we get:

$$\begin{aligned} (6.9) \quad &\langle \Delta v_k, v^{k+1} \rangle \\ &= -\lambda(\alpha - 1 - \lambda) \langle \Delta z_{k-1}, z^k - z^* \rangle - t_k(\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta z_k \rangle - (\alpha - 1 - \lambda)t_k(1 - \eta) \langle \Delta z_{k-1}, Q^{k+1} \rangle \\ &\quad - \lambda\eta(\alpha - 1) \langle Q^k, z^k - z^* \rangle - t_k\eta(\alpha - 1) \langle Q^k, \Delta z_k \rangle - t_k\eta(1 - \eta)(\alpha - 1) \langle Q^k, Q^{k+1} \rangle \\ &\quad - \lambda t_k \eta \langle \Delta Q_k, z^k - z^* \rangle - t_k^2 \eta \langle \Delta Q_k, \Delta z_k \rangle - \eta(1 - \eta)t_k^2 \langle \Delta Q_k, Q^{k+1} \rangle. \end{aligned}$$

⁷The coefficient b_t in (2.2) is in fact the only parameter that (i) does not depend on λ (artificial parameter introduced in the Lyapunov analysis), and (ii) can make this quadratic form negative semidefinite for all $\lambda \in (0, \alpha - 1]$. To see this, consider $\lambda = \alpha - 1$. Then, δ_t must be equal to $t^2 w_t$, which yields (2.2).

Let us now inspect $\|\Delta v_k\|^2$:

$$(6.10) \quad \begin{aligned} -\frac{1}{2}\|\Delta v_k\|^2 = & -\frac{1}{2}\eta^2(\alpha-1)^2\|Q^k\|^2 - \frac{t_k^2}{2}\eta^2\|\Delta Q_k\|^2 - \frac{1}{2}(\alpha-1-\lambda)^2\|\Delta z_{k-1}\|^2 \\ & - (\alpha-1)\eta^2 t_k \langle Q^k, \Delta Q_k \rangle - \eta(\alpha-1-\lambda)(\alpha-1) \langle Q^k, \Delta z_{k-1} \rangle \\ & - \eta t_k(\alpha-1-\lambda) \langle \Delta z_{k-1}, \Delta Q_k \rangle . \end{aligned}$$

Plugging (6.9) and (6.10) into (6.7) and rearranging suitably we get

$$(6.11) \quad \begin{aligned} \Delta E_k^\lambda = & -\lambda\eta(\alpha-2) \langle Q^k, z^k - z^* \rangle - t_k^2\eta \langle \Delta Q_k, \Delta z_k \rangle \\ & - \eta t_k(\alpha-1-\lambda) \langle \Delta z_{k-1}, \Delta Q_k \rangle \end{aligned}$$

$$(6.12) \quad \begin{aligned} & + \frac{1}{2}\eta(1-\eta)\Delta t_{k-1}^2\|Q^k\|^2 - t_k\eta(1-\eta)(\alpha-1) \langle Q^k, Q^{k+1} \rangle \\ & - \frac{\eta^2}{2}(\alpha-1)^2\|Q^k\|^2 - (\alpha-1)\eta^2 t_k \langle Q^k, \Delta Q_k \rangle \end{aligned}$$

$$(6.13) \quad \begin{aligned} & + \lambda\eta t_k \langle Q^{k+1}, \Delta z_k \rangle - (\alpha-1-\lambda)t_k(1-\eta) \langle \Delta z_{k-1}, Q^{k+1} \rangle - t_k\eta(\alpha-1) \langle Q^k, \Delta z_k \rangle \\ & - \eta(\alpha-1-\lambda)(\alpha-1) \langle Q^k, \Delta z_{k-1} \rangle \end{aligned}$$

$$(6.14) \quad -t_k(\alpha-1-\lambda) \langle \Delta z_{k-1}, \Delta z_k \rangle$$

$$(6.15) \quad \begin{aligned} & - \frac{1}{2}(\alpha-1-\lambda)^2\|\Delta z_{k-1}\|^2 \\ & - \frac{1}{2}\eta t_k^2\|\Delta Q_k\|^2 - \frac{\lambda}{2}(\alpha-1-\lambda)\|\Delta z_{k-1}\|^2 + \Delta \left(\frac{\delta}{2}\|Q\|^2 \right)_k , \end{aligned}$$

where in the last line we have used $\eta(1-\eta) + \eta^2 = \eta$. To deal with the various nonnegative off-sets in the inner products (6.12), (6.13) and (6.14), we make a discrete version of the continuous-time quadratic form appear and control the remainder with Young's inequality. This gives

$$\begin{aligned} (6.12) & := \frac{1}{2}\eta(1-\eta)\Delta t_{k-1}^2\|Q^k\|^2 - t_k\eta(1-\eta)(\alpha-1) \langle Q^k, Q^{k+1} \rangle \\ & \quad - \frac{\eta^2}{2}(\alpha-1)^2\|Q^k\|^2 - (\alpha-1)\eta^2 t_k \langle Q^k, \Delta Q_k \rangle \\ & = \eta \left(\frac{(1-\eta)}{2}\Delta t_{k-1}^2 - (1-\eta)t_k(\alpha-1) - \frac{\eta}{2}(\alpha-1)^2 \right) \|Q^k\|^2 - t_k(\alpha-1)\eta \langle Q^k, \Delta Q_k \rangle \\ & = - \underbrace{\left[\eta(1-\eta) \left((\alpha-2)t_k + \frac{1}{2} \right) + \frac{1}{2}\eta^2(\alpha-1)^2 \right]}_{(Q.1)} \|Q^k\|^2 - t_k(\alpha-1)\eta \langle Q^k, \Delta Q_k \rangle , \end{aligned}$$

where in the coefficient that multiplies $\|Q^k\|^2$ we have used, by the definition of t_k ,

$$(6.16) \quad \frac{1}{2}\Delta t_{k-1}^2 = \frac{1}{2}((k-1+\sigma)^2 - (k-2+\sigma)^2) = t_k - \frac{1}{2} ,$$

hence $\frac{1}{2}\Delta t_{k-1}^2 - t_k(\alpha - 1) = -\frac{1}{2} - t_k(\alpha - 2)$. The first term of our quadratic form involving $\|Q^k\|^2$ appears. We now force $\langle Q^k, \Delta z_{k-1} \rangle$ to appear. This comes from the fact that

$$\begin{aligned}
(6.13) &:= \lambda \eta t_k \langle Q^{k+1}, \Delta z_k \rangle - (\alpha - 1 - \lambda) t_k (1 - \eta) \langle \Delta z_{k-1}, Q^{k+1} \rangle - t_k \eta (\alpha - 1) \langle Q^k, \Delta z_k \rangle \\
&\quad - \eta (\alpha - 1 - \lambda) (\alpha - 1) \langle Q^k, \Delta z_{k-1} \rangle \\
&= \lambda \eta t_k \langle Q^k, \Delta z_{k-1} \rangle - (\alpha - 1 - \lambda) t_k (1 - \eta) \langle Q^k, \Delta z_{k-1} \rangle - t_k \eta (\alpha - 1) \langle Q^k, \Delta z_{k-1} \rangle \\
&\quad + \lambda \eta t_k \langle Q^{k+1}, \Delta^2 z_{k-1} \rangle - (\alpha - 1 - \lambda) t_k (1 - \eta) \langle \Delta Q_k, \Delta z_{k-1} \rangle - t_k \eta (\alpha - 1) \langle Q^k, \Delta^2 z_{k-1} \rangle \\
&\quad + \lambda \eta t_k \langle \Delta Q_k, \Delta z_{k-1} \rangle - \eta (\alpha - 1 - \lambda) (\alpha - 1) \langle Q^k, \Delta z_{k-1} \rangle \\
&= - \underbrace{(\alpha - 1 - \lambda) t_k \langle Q^k, \Delta z_{k-1} \rangle}_{(Q.2)} - t_k \eta (\alpha - 1 - \lambda) \langle Q^k, \Delta^2 z_{k-1} \rangle \\
&\quad + \lambda \eta t_k \langle \Delta Q_k, \Delta^2 z_{k-1} \rangle - (\alpha - 1 - \lambda) t_k (1 - \eta) \langle \Delta Q_k, \Delta z_{k-1} \rangle \\
&\quad + \lambda \eta t_k \langle \Delta Q_k, \Delta z_{k-1} \rangle - \eta (\alpha - 1 - \lambda) (\alpha - 1) \langle Q^k, \Delta z_{k-1} \rangle \\
&= (Q.2) \\
&\quad - \underbrace{t_k \eta (\alpha - 1 - \lambda) \langle Q^k, \Delta^2 z_{k-1} \rangle}_{(II)} + \lambda \eta t_k \langle \Delta Q_k, \Delta z_k \rangle - (\alpha - 1 - \lambda) t_k (1 - \eta) \langle \Delta Q_k, \Delta z_{k-1} \rangle \\
&\quad - \underbrace{\eta (\alpha - 1 - \lambda) (\alpha - 1) \langle Q^k, \Delta z_{k-1} \rangle}_{(N.1)}.
\end{aligned}$$

We will handle (II) separately later. To conclude, we make the term in $\|\Delta z_{k-1}\|^2$ appear. Starting with (6.14) we have:

$$\begin{aligned}
(6.14) &:= -t_k (\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta z_k \rangle - \frac{1}{2} (\alpha - 1 - \lambda)^2 \|\Delta z_{k-1}\|^2 \\
&= \underbrace{-t_k (\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2}_{(Q.3)} \underbrace{- t_k (\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta^2 z_{k-1} \rangle}_{(III)} \underbrace{- \frac{1}{2} (\alpha - 1 - \lambda)^2 \|\Delta z_{k-1}\|^2}_{(N.2)}.
\end{aligned}$$

Once again, we will deal with (III) later. Gathering all the terms together, we get

$$\begin{aligned}
\Delta E_k^\lambda &= -\lambda \eta (\alpha - 2) \langle Q^k, z^k - z^* \rangle - t_k^2 \eta \langle \Delta Q_k, \Delta z_k \rangle \\
&\quad - \eta t_k (\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta Q_k \rangle \\
&\quad + (Q.1) - t_k (\alpha - 1) \eta \langle Q^k, \Delta Q_k \rangle \\
&\quad + (Q.2) + (II) + \lambda \eta t_k \langle \Delta Q_k, \Delta z_k \rangle - (\alpha - 1 - \lambda) t_k (1 - \eta) \langle \Delta Q_k, \Delta z_{k-1} \rangle + (N.1) \\
&\quad + (Q.3) + (III) + (N.2) \\
&\quad - \frac{1}{2} \eta t_k^2 \|\Delta Q_k\|^2 - \frac{\lambda}{2} (\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2 + \Delta \left(\frac{\delta}{2} \|Q\|^2 \right)_k \\
(6.17) \quad &= -\lambda \eta (\alpha - 2) \langle Q^k, z^k - z^* \rangle - (t_k^2 - \lambda t_k) \eta \langle \Delta Q_k, \Delta z_k \rangle \\
&\quad - t_k (\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta Q_k \rangle \\
&\quad + (Q.1) - t_k (\alpha - 1) \eta \langle Q^k, \Delta Q_k \rangle \\
&\quad + (Q.2) + (II) + (N.1) \\
&\quad + (Q.3) + (III) + (N.2) - \frac{1}{2} \eta t_k^2 \|\Delta Q_k\|^2 - \frac{\lambda}{2} (\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2 + \Delta \left(\frac{\delta}{2} \|Q\|^2 \right)_k,
\end{aligned}$$

Let us now collect all the terms in (Q.2) and (Q.3) and argue analogously to the continuous-time setting to get

$$\begin{aligned}
(6.18) \quad (Q.2) + (Q.3) &= -(\alpha - 1 - \lambda)t_k \langle Q^k, \Delta z_{k-1} \rangle - t_k(\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2 \\
&= -(\alpha - 1 - \lambda)t_k \left(\langle Q^k, \Delta z_{k-1} \rangle + \|\Delta z_{k-1}\|^2 \right) \\
&= -(\alpha - 1 - \lambda)t_k \left(\frac{1}{2} \|Q^k + \Delta z_{k-1}\|^2 + \frac{1}{2} \|\Delta z_{k-1}\|^2 - \frac{1}{2} \|Q^k\|^2 \right) \\
&= (\alpha - 1 - \lambda) \frac{t_k}{2} \|Q^k\|^2 - (\alpha - 1 - \lambda) \frac{t_k}{2} \|\Delta z_{k-1}\|^2 - (\alpha - 1 - \lambda) \frac{t_k}{2} \|Q^k + \Delta z_{k-1}\|^2.
\end{aligned}$$

Note that the positive term $(\alpha - 1 - \lambda) \frac{t_k}{2}$ in the last display is not a concern when $\alpha > 2$ since we shall take λ arbitrarily close to $\alpha - 1$ in our Lyapunov analysis in such a way that when combined with (Q.1), the coefficient in front of $\|Q^k\|^2$ will be negative. The case $\alpha = 2$ will be treated with $\lambda = \alpha - 1$. We now deal with (II) and (III), while the terms labeled with a capital “N” will be handled properly later. We have

$$\begin{aligned}
(6.19) \quad (II) &= -(\alpha - 1 - \lambda)t_k \eta \langle Q^k, \Delta^2 z_{k-1} \rangle \\
&= -(\alpha - 1 - \lambda)t_k \eta \left\langle Q^k, -\frac{\alpha}{t_k} \Delta z_{k-1} - \Delta Q_k - \frac{\theta}{t_k} Q^k \right\rangle \\
&= (\alpha - 1 - \lambda) \eta t_k \langle Q^k, \Delta Q_k \rangle + \underbrace{(\alpha - 1 - \lambda) [\alpha \eta \langle Q^k, \Delta z_{k-1} \rangle + \eta \theta \|Q^k\|^2]}_{(N.3)}.
\end{aligned}$$

As far as (III) is concerned, we have

$$\begin{aligned}
(6.20) \quad (III) &= -t_k(\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta^2 z_{k-1} \rangle \\
&= -t_k(\alpha - 1 - \lambda) \left\langle \Delta z_{k-1}, -\frac{\alpha}{t_k} \Delta z_{k-1} - \Delta Q_k - \frac{\theta}{t_k} Q^k \right\rangle \\
&= t_k(\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta Q_k \rangle + \underbrace{\alpha(\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2 + \theta(\alpha - 1 - \lambda) \langle \Delta z_{k-1}, Q^k \rangle}_{(N.4)}.
\end{aligned}$$

Plugging now (II) and (III) into (6.17) we get

$$\begin{aligned}
(6.21) \quad \Delta E_k^\lambda &= -\lambda \eta (\alpha - 2) \langle Q^k, z^k - z^* \rangle - (t_k^2 - \lambda t_k) \eta \langle \Delta Q_k, \Delta z_k \rangle \\
&\quad + (Q.1) \underbrace{-t_k \lambda \eta \langle Q^k, \Delta Q_k \rangle}_{(IV)} \\
&\quad + (Q.2) + (N.1) + (N.3) \\
&\quad + (Q.3) + (N.2) + (N.4) \\
&\quad - \frac{1}{2} \eta t_k^2 \|\Delta Q_k\|^2 - \underbrace{\frac{\lambda}{2} (\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2}_{(N.5)} + \Delta \left(\frac{\delta}{2} \|Q\|^2 \right)_k,
\end{aligned}$$

where the term $t_k(\alpha - 1 - \lambda) \langle \Delta z_{k-1}, \Delta Q_k \rangle$ canceled out with the corresponding one in (III), and we simplified the coefficient that multiplies $\langle Q^k, \Delta Q_k \rangle$ using the contribution from (II). Let us also define for brevity: $(N) := (N.1) + (N.2) + (N.3) + (N.4) + (N.5)$ and inspect (IV): Using the polarization

identity and recalling that $\delta_k = t_{k-1}\lambda\eta$, we get:

$$\begin{aligned}
 (IV) &= -t_k\lambda\eta \langle Q^k, \Delta Q_k \rangle \\
 &= -t_k\lambda\eta \left(\frac{1}{2}\|Q^{k+1}\|^2 - \frac{1}{2}\|Q^k\|^2 - \frac{1}{2}\|\Delta Q_k\|^2 \right) \\
 (6.22) \quad &= -\frac{1}{2}t_k\lambda\eta\|Q^{k+1}\|^2 + \frac{1}{2}t_{k-1}\lambda\eta\|Q^k\|^2 + \frac{\lambda}{2}\eta\|Q^k\|^2 + \frac{t_k}{2}\lambda\eta\|\Delta Q_k\|^2 \\
 &= -\Delta \left(\frac{\delta}{2}\|Q\|^2 \right)_k + \frac{\lambda}{2}\eta\|Q^k\|^2 + \frac{t_k}{2}\lambda\eta\|\Delta Q_k\|^2.
 \end{aligned}$$

Inserting (IV), (Q.1), (Q.2), and (Q.3) into (6.21), simplifying and grouping terms yields

$$\begin{aligned}
 \Delta E_k^\lambda &= -(t_k^2 - \lambda t_k)\eta \left(\langle \Delta Q_k, \Delta z_k \rangle + \frac{1}{2}\|\Delta Q_k\|^2 \right) - \lambda\eta(\alpha - 2) \left(\langle Q^k, z^k - z^* \rangle + \frac{1}{2}\|Q^k\|^2 \right) \\
 (6.23) \quad &+ \left[\frac{\lambda}{2}\eta(\alpha - 2) + \frac{\lambda}{2}\eta \right] \|Q^k\|^2 - \left[\eta(1 - \eta) \left((\alpha - 2)t_k + \frac{1}{2} \right) + \frac{1}{2}\eta^2(\alpha - 1)^2 \right] \|Q^k\|^2 + (N) \\
 &+ (\alpha - 1 - \lambda) \frac{t_k}{2} \|Q^k\|^2 - (\alpha - 1 - \lambda) \frac{t_k}{2} \|\Delta z_{k-1}\|^2 - (\alpha - 1 - \lambda) \frac{t_k}{2} \|Q^k + \Delta z_{k-1}\|^2.
 \end{aligned}$$

We now make use of (3.10) to infer that

$$(6.24) \quad \langle \Delta Q_k, \Delta z_k \rangle \geq -\frac{1}{2}\|\Delta Q_k\|^2.$$

This allows us to get for $k \geq 1$ and such that $t_k \geq \lambda$

$$\begin{aligned}
 \Delta E_k^\lambda &\leq -\frac{\eta}{2} \left((1 - \eta) + \eta(\alpha - 1)^2 - \lambda(\alpha - 1) \right) \|Q^k\|^2 - \left(\eta(1 - \eta)(\alpha - 2) - \frac{\alpha - 1 - \lambda}{2} \right) t_k \|Q^k\|^2 \\
 (6.25) \quad &- (\alpha - 1 - \lambda) \frac{t_k}{2} \|\Delta z_{k-1}\|^2 - (\alpha - 1 - \lambda) \frac{t_k}{2} \|Q^k + \Delta z_{k-1}\|^2 + (N).
 \end{aligned}$$

Now, invoking the Young's inequality on all the inner products appearing in (N), we obtain

$$\begin{aligned}
 (N) &= -\eta(\alpha - 1 - \lambda)(\alpha - 1) \langle Q^k, \Delta z_{k-1} \rangle - \frac{1}{2}(\alpha - 1 - \lambda)^2 \|\Delta z_{k-1}\|^2 \\
 &+ (\alpha - 1 - \lambda) [\alpha\eta \langle Q^k, \Delta z_{k-1} \rangle + \eta\theta \|Q^k\|^2] \\
 (6.26) \quad &+ \alpha(\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2 + \theta(\alpha - 1 - \lambda) \langle \Delta z_{k-1}, Q^k \rangle - \frac{\lambda}{2}(\alpha - 1 - \lambda) \|\Delta z_{k-1}\|^2 \\
 &= (\alpha - 1 - \lambda) \left(\eta\theta \|Q^k\|^2 + (\eta + \theta) \langle Q^k, \Delta z_{k-1} \rangle + \frac{\alpha + 1}{2} \|\Delta z_{k-1}\|^2 \right) \\
 &\leq (\alpha - 1 - \lambda) \left(\eta\theta + \frac{\eta + \theta}{2} \right) \|Q^k\|^2 + (\alpha - 1 - \lambda) \frac{\alpha + 2 + \eta(\alpha - 1)}{2} \|\Delta z_{k-1}\|^2.
 \end{aligned}$$

Inserting (6.26) into (6.25) and rearranging we get the claim. $\square$

REFERENCES

- [1] S. Adly and H. Attouch. “Complexity Analysis Based on Tuning the Viscosity Parameter of the Su–Boyd–Candès Inertial Gradient Dynamics”. In: *Set-Valued and Variational Analysis* 32.2 (2024), p. 17.
- [2] A. Åkerman, E. Chenchene, P. Giselsson, and E. Naldi. *Splitting the Forward-Backward Algorithm: A Full Characterization*. 2025.
- [3] F. Alvarez, H. Attouch, J. Bolte, and P. Redont. “A second-order gradient-like dissipative dynamical system with Hessian-driven damping: Application to optimization and mechanics”. In: *Journal de Mathématiques Pures et Appliquées* 81.8 (2002), pp. 747–779.
- [4] F. Alvarez and H. Attouch. “An Inertial Proximal Method for Maximal Monotone Operators via Discretization of a Nonlinear Oscillator with Damping”. In: *Set-Valued Analysis* 9.1 (Mar. 2001), pp. 3–11.
- [5] H. Attouch, R. I. Bot, and D.-K. Nguyen. “Fast convex optimization via time scale and averaging of the steepest descent”. In: *Mathematics of Operations Research* (2024).

- [6] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. “Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping”. In: *Optimization* 72.5 (2023), pp. 1199–1238.
- [7] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. “First-order optimization algorithms via inertial systems with Hessian driven damping”. In: *Mathematical Programming* 193.1 (2022), pp. 113–155.
- [8] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. “Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$ ”. In: *ESAIM: Control, Optimisation and Calculus of Variations* 25 (2019), p. 2.
- [9] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. “Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity”. In: *Mathematical Programming* 168.1 (2018), pp. 123–175.
- [10] H. Attouch, Z. Chbani, and H. Riahi. “Fast Proximal Methods via Time Scaling of Damped Inertial Dynamics”. In: *SIAM Journal on Optimization* 29.3 (2019), pp. 2227–2256.
- [11] H. Attouch and J. Fadili. “From the Ravine Method to the Nesterov Method and Vice Versa: A Dynamical System Perspective”. In: *SIAM Journal on Optimization* 32.3 (2022), pp. 2074–2101.
- [12] H. Attouch, P.-E. Maingé, and P. Redont. “A second-order differential system with Hessian-driven damping; Application to non-elastic shock laws”. In: *Differential Equations and Applications* 4 (2012), pp. 27–65.
- [13] H. Attouch, J. Peypouquet, and P. Redont. “Fast convex optimization via inertial dynamics with Hessian driven damping”. In: *Journal of Differential Equations* 261.10 (2016), pp. 5734–5783.
- [14] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Second. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, 2017.
- [15] H. H. Bauschke, W. M. Moursi, and X. Wang. “Generalized monotone operators and their averaged resolvents”. In: *Mathematical Programming* 189.1 (2021), pp. 55–74.
- [16] A. Beck and M. Teboulle. “A Fast Iterative Shrinkage-Thresholding algorithm for linear inverse problems”. In: *SIAM Journal On Imaging Sciences* 2 (2009), pp. 183–202.
- [17] J.-D. Benamou and G. Carlier. “Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations”. In: *Journal of Optimization Theory and Applications* 167.1 (2015), pp. 1–26.
- [18] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. “Iterative Bregman Projections for Regularized Transportation Problems”. In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138.
- [19] R. I. Boţ and E. Chenchene. *Extra-Gradient Method with Flexible Anchoring: Strong Convergence and Fast Residual Decay*. 2024.
- [20] R. I. Boţ, E. R. Csetnek, and D.-K. Nguyen. “Fast Optimistic Gradient Descent Ascent (OGDA) Method in Continuous and Discrete Time”. In: *Foundations of Computational Mathematics* (2023).
- [21] R. I. Boţ and D.-K. Nguyen. “Fast Krasnoselskii–Mann Algorithm with a Convergence Rate of the Fixed Point Iteration of $o\left(\frac{1}{k}\right)$ ”. In: *SIAM Journal on Numerical Analysis* 61.6 (2023), pp. 2813–2843.
- [22] R. I. Boţ and D.-K. Nguyen. *Tikhonov regularization of monotone operator flows not only ensures strong convergence of the trajectories but also speeds up the vanishing of the residuals*. arXiv:2406.00852. 2024.
- [23] R. I. Boţ, E. Chenchene, E. R. Csetnek, and D. A. Hulett. *Accelerating Diagonal Methods for Bilevel Optimization: Unified Convergence via Continuous-Time Dynamics*. 2025.
- [24] K. Bredies, E. Chenchene, D. A. Lorenz, and E. Naldi. “Degenerate Preconditioned Proximal Point Algorithms”. In: *SIAM Journal on Optimization* 32.3 (2022), pp. 2376–2401.
- [25] K. Bredies, E. Chenchene, and E. Naldi. “Graph and Distributed Extensions of the Douglas–Rachford Method”. In: *SIAM Journal on Optimization* 34.2 (2024), pp. 1569–1594.
- [26] K. Bredies and H. Sun. “A proximal point analysis of the preconditioned alternating direction method of multipliers”. In: *Journal of Optimization Theory and Applications* 173.3 (2017), pp. 878–907.
- [27] K. Bredies and H. Sun. “Preconditioned Douglas–Rachford splitting methods for convex-concave saddle-point problems”. In: *SIAM Journal on Numerical Analysis* 53.1 (2015), pp. 421–444.
- [28] F. E. Browder. “Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces”. In: *Proceedings of Symposia in Pure Mathematics*. Vol. 18, part 2. Providence, RI: American Mathematical Society, 1976.
- [29] G. Carlier, E. Chenchene, and K. Eichinger. “Wasserstein Medians: Robustness, PDE Characterization, and Numerics”. In: *SIAM Journal on Mathematical Analysis* 56.5 (2024), pp. 6483–6520.
- [30] A. Chambolle and C. Dossal. “On the Convergence of the Iterates of the “Fast Iterative Shrinkage/Thresholding Algorithm””. In: *Journal of Optimization Theory and Applications* 166.3 (2015), pp. 968–982.
- [31] A. Chambolle and T. Pock. “A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging”. In: *Journal of Mathematical Imaging and Vision* 40.1 (2011), pp. 120–145.

- [32] A. Chambolle and T. Pock. “An introduction to continuous optimization for imaging”. In: *Acta Numerica* 25 (2016), 161–319.
- [33] E. Chenchene. “Splitting methods in nonsmooth convex degenerate optimization”. Dissertation (PhD). Graz: University of Graz, 2023.
- [34] R. Cominetti, J. Peypouquet, and S. Sorin. “Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization”. In: *Journal of Differential Equations* 245.12 (2008), pp. 3753–3763.
- [35] J. P. Contreras and R. Cominetti. “Optimal error bounds for non-expansive fixed-point iterations in normed spaces”. In: *Mathematical Programming* 199.1 (2023), pp. 343–374.
- [36] D. Davis and W. Yin. “A three-operator splitting scheme and its optimization applications”. In: *Set-Valued and Variational Analysis. Theory and Applications* 25.4 (2017), pp. 829–858.
- [37] D. Davis and W. Yin. “Convergence Rate Analysis of Several Splitting Schemes”. In: *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer International Publishing, 2016, pp. 115–163.
- [38] J. Douglas Jr. and H. H. Rachford Jr. “On the numerical solution of heat conduction problems in two and three space variables”. In: *Transactions of the American Mathematical Society* 82 (1956), pp. 421–439.
- [39] M. Fornasier, H. Huang, L. Pareschi, and P. Sünnen. “Anisotropic Diffusion in Consensus-Based Optimization on the Sphere”. In: *SIAM Journal on Optimization* 32.3 (2022), pp. 1984–2012.
- [40] O. Güler. “New Proximal Point Algorithms for Convex Minimization”. In: *SIAM Journal on Optimization* 2.4 (1992), pp. 649–664.
- [41] O. Güler. “On the Convergence of the Proximal Point Algorithm for Convex Minimization”. In: *SIAM Journal on Control and Optimization* 29.2 (1991), pp. 403–419.
- [42] A. Haraux. *Systèmes dynamiques dissipatifs et applications*. Vol. 17. Recherches en Mathématiques Appliquées. Paris: Masson, 1991.
- [43] B. He and X. Yuan. “Convergence Analysis of Primal-Dual Algorithms for a Saddle-Point Problem: From Contraction Perspective”. In: *SIAM Journal on Imaging Sciences* 5 (2012), pp. 119–149.
- [44] D. Kim. “Accelerated proximal point method for maximally monotone operators”. In: *Mathematical Programming* 190.1 (2021), pp. 57–87.
- [45] M. A. Krasnoselskii. “Two remarks on the method of successive approximations”. In: *Uspekhi Matematicheskikh Nauk* 10 (1955), pp. 123–127.
- [46] J. Liang, J. Fadili, and G. Peyré. “Convergence rates with inexact non-expansive operators”. In: *Mathematical Programming* 159.1 (2016), pp. 403–434.
- [47] F. Lieder. “On the convergence rate of the Halpern-iteration”. In: *Optimization Letters* 15.2 (2021), pp. 405–418.
- [48] P. L. Lions and B. Mercier. “Splitting Algorithms for the Sum of Two Nonlinear Operators”. In: *SIAM Journal on Numerical Analysis* 16.6 (1979), pp. 964–979.
- [49] W. R. Mann. “Mean value methods in iteration”. In: *Proceedings of the American Mathematical Society* 4 (1953), pp. 506–510.
- [50] J. J. Maulén, I. Fierro, and J. Peypouquet. “Inertial Krasnoselskii-Mann Iterations”. In: *Set-Valued and Variational Analysis* 32.2 (Mar. 2024), p. 10.
- [51] G. J. Minty. “Monotone (nonlinear) operators in Hilbert space”. In: *Duke Math. J.* 29 (1962), pp. 341–346.
- [52] Y. Nesterov. “A method for solving the convex programming problem with convergence rate $O(1/k^2)$ ”. In: *Proceedings of the USSR Academy of Sciences* 269 (1983), pp. 543–547.
- [53] N. Papadakis, G. Peyré, and E. Oudet. “Optimal transport with proximal splitting”. In: *SIAM Journal on Imaging Sciences* 7.1 (2014), pp. 212–238.
- [54] H. Raguet, J. Fadili, and G. Peyré. “A generalized forward-backward splitting”. In: *SIAM Journal on Imaging Sciences* 6.3 (2013), pp. 1199–1226.
- [55] S. Reich. “Nonlinear evolution equations and nonlinear ergodic theorems”. In: *Nonlinear Analysis: Theory, Methods & Applications* 1.4 (1977), pp. 319–330.
- [56] S. Sabach and S. Shtern. “A First Order Method for Solving Convex Bilevel Optimization Problems”. In: *SIAM Journal on Optimization* 27.2 (2017), pp. 640–660.
- [57] W. Su, S. Boyd, and E. J. Candès. “A Differential Equation for Modeling Nesterov’s Accelerated Gradient Method: Theory and Insights”. In: *Journal of Machine Learning Research* 17.153 (2016), pp. 1–43.
- [58] J. Suh, J. Park, and E. K. Ryu. “Continuous-time Analysis of Anchor Acceleration”. In: *Advances in Neural Information Processing Systems*. Vol. 36. 2023, pp. 32782–32866.
- [59] D. Sun, Y. Yuan, G. Zhang, and X. Zhao. “Accelerating Preconditioned ADMM via Degenerate Proximal Point Mappings”. In: *SIAM Journal on Optimization* 35.2 (2025), pp. 1165–1193.

- [60] B. F. Svaiter. “On weak convergence of the Douglas-Rachford method”. In: *SIAM Journal on Control and Optimization* 49.1 (2011), pp. 280–287.
- [61] Q. Tran-Dinh. “From Halpern’s fixed-point iterations to Nesterov’s accelerated interpretations for root-finding problems”. In: *Computational Optimization and Applications* 87 (2024), pp. 181–218.
- [62] T. Yoon, J. Kim, J. J. Suh, and E. K. Ryu. “Optimal Acceleration for Minimax and Fixed-Point Problems is Not Unique”. In: *Proceedings of the 41st International Conference on Machine Learning*. Vol. 235. Proceedings of Machine Learning Research. PMLR, 2024, pp. 57244–57314.

*FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA, OSKAR-MORGENSTERN-PLATZ 1, 1090 VIENNA, AUSTRIA
Email address: `radu.bot@univie.ac.at`, `enis.chenchene@univie.ac.at`

†NORMANDIE UNIV, ENSICAEN, CNRS, GREYC, FRANCE.
Email address: `Jalal.Fadili@ensicaen.fr`