SIAM J. Optim.
Vol. 30, No. 4, pp. 3252–3283

NEWTON-LIKE INERTIAL DYNAMICS AND PROXIMAL ALGORITHMS GOVERNED BY MAXIMALLY MONOTONE OPERATORS*

HEDY ATTOUCH† and SZILÁRD CSABA LÁSZLÓ‡

Abstract. The introduction of the Hessian damping in the continuous version of Nesterov's accelerated gradient method provides, by temporal discretization, fast proximal gradient algorithms where the oscillations are significantly attenuated. We will extend these results to the maximally monotone case. We rely on the technique introduced by Attouch and Peypouquet [*Math. Program.*, 174 (2019), pp. 391–432], where the maximally monotone operator is replaced by its Yosida approximation with an appropriate adjustment of the regularization parameter. In a general Hilbert framework, we obtain the weak convergence of the iterates to equilibria, and the rapid convergence of the discrete velocities to zero. By specializing these algorithms to convex minimization, we obtain the convergence rate $o\left(1/k^2\right)$ of the values, and the rapid convergence of the gradients toward zero.

Key words. damped inertial dynamics, Hessian damping, large step proximal method, Lyapunov analysis, maximally monotone operators, Newton method, time-dependent viscosity, vanishing viscosity, Yosida regularization

AMS subject classifications. 37N40, 46N10, 49M30, 65K05, 65K10, 90B50, 90C25

DOI. 10.1137/20M1333316

1. Introduction.

Let $\mathcal{H}$ be a real Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. Given a general maximally monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$, based on Newton's method, we want to design rapidly converging proximal algorithms to solve the monotone inclusion

$$0 \in Ax. \tag{1.1}$$

Solving (1.1), i.e., finding a zero of A, is a difficult problem of fundamental importance in optimization, equilibrium theory, economics and game theory, partial differential equations, and statistics, among other subjects (see, for instance, [22, 25, 26, 30, 31, 33, 36]). As a guide to our study, the algorithms will be derived from the temporal implicit discretization of the second-order differential equation

$$(\mathrm{DIN}-\mathrm{AVD})_{\alpha,\beta} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\frac{d}{dt}\left(A_{\lambda(t)}(x(t))\right) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0,$$

where α, β are positive damping parameters, and $J_{\lambda A} = (I + \lambda A)^{-1}$, $A_\lambda = \frac{1}{\lambda}(I - J_{\lambda A})$ stand respectively for the resolvent of A and the Yosida regularization of A of index $\lambda > 0$. According to the Lipschitz continuity property of A_λ, $(\mathrm{DIN}-\mathrm{AVD})_{\alpha,\beta}$ is a well-posed evolution equation which enjoys nice asymptotic convergence properties. The object of our study is the proximal regularized inertial Newton algorithm for

*Received by the editors April 21, 2020; accepted for publication (in revised form) September 24, 2020; published electronically December 3, 2020.
https://doi.org/10.1137/20M1333316
Funding: This work was supported by COST Action, CA16228. The second author was also supported by a grant of the Ministry of Research and Innovation, CNCS–UEFISCDI, project PN-III-P1-1.1-TE-2016-0266.
†IMAG, Université Montpellier, CNRS, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France (hedy.attouch@univ-montp2.fr).
‡Department of Mathematics, Technical University of Cluj-Napoca, Cluj-Napoca, Romania (szilard.laszlo@math.utcluj.ro).

monotone operator, called (PRINAM) for short, and which can be viewed as a discrete temporal version of $(\text{DIN} - \text{AVD})_{\alpha,\beta}$. It is written as follows: given $x_1, x_2 \in \mathcal{H}$

$$\begin{cases} y_k = \left(1 - \beta\left(\frac{1}{\lambda_k} - \frac{1}{\lambda_{k-1}}\right)\right) x_k + \left(\alpha_k - \frac{\beta}{\lambda_{k-1}}\right)(x_k - x_{k-1}) + \frac{\beta}{\lambda_k} J_{\lambda_k A}(x_k) \\ \qquad - \frac{\beta}{\lambda_{k-1}} J_{\lambda_{k-1} A}(x_{k-1}), \\ x_{k+1} = \frac{\lambda_{k+1}}{\lambda_{k+1} + s} y_k + \frac{s}{\lambda_{k+1} + s} J_{(\lambda_{k+1}+s)A}(y_k) \text{ for all } k \geq 2. \end{cases}$$

This algorithm includes both extrapolation and relaxation steps. Compared to the extrapolation step in the accelerated gradient method of Nesterov, its main characteristic is to include an additional correction term which is equal to the difference of the resolvents computed at two consecutive iterates. As a main result, we will prove that for an appropriate adjustment of the parameters, any sequence (x_k) generated by this algorithm converges weakly to a zero of A. Moreover, when $A = \partial f$ specializes in the subdifferential of a proper, convex lower semicontinuous function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, we obtain the convergence rate $o\left(\frac{1}{k^2}\right)$ of the values, and the fast convergence of the gradients toward zero. Our study is based on several recent advances in the study of inertial dynamics and algorithms for solving optimization problems and monotone inclusions. We describe them briefly in the following sections. Our main contribution is to show how to put them together.

1.1. Asymptotic vanishing damping.

The inertial system

$$(\text{AVD})_\alpha \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0,$$

was introduced in the context of convex optimization by Su, Boyd, and Candès in [41]. For a general convex differentiable function f, it provides a continuous version of the accelerated gradient method of Nesterov. For $\alpha \geq 3$, each trajectory $x(\cdot)$ of $(\text{AVD})_\alpha$ satisfies the asymptotic convergence rate of the values $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(1/t^2\right)$. As a specific feature, the viscous damping coefficient $\frac{\alpha}{t}$ vanishes (tends to zero) as time t goes to infinity, hence the terminology. The case $\alpha = 3$, which corresponds to Nesterov's historical algorithm, is critical. In the case $\alpha = 3$, the question of the convergence of the trajectories remains an open problem (except in one dimension where convergence holds [10]). For $\alpha > 3$, it has been shown by Attouch et al. [9] that each trajectory converges weakly to a minimizer. For $\alpha > 3$, it is shown in [15] and [35] that the asymptotic convergence rate of the values is actually $o(1/t^2)$. These rates are optimal, that is, they can be reached or approached arbitrarily close. The corresponding inertial algorithms

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}), \\ x_{k+1} = y_k - s\nabla f(y_k) \end{cases}$$

are in line with the Nesterov accelerated gradient method. They enjoy similar properties to the continuous case; see Chambolle and Dossal [24] and [7, 9, 15] for further results.

1.2. Hessian damping. The inertial system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0$$

combines asymptotic vanishing damping with Hessian-driven damping. It was considered by Attouch, Peypouquet, and Redont in [16] (see also [3, 12]). At first glance, the presence of the Hessian may seem to entail numerical difficulties. However, this is not the case as the Hessian intervenes in the form $\nabla^2 f(x(t))\dot{x}(t)$, which is nothing but the derivative with respect to time of the function $t \mapsto \nabla f(x(t))$. So, the temporal discretization of this dynamic provides first-order algorithms of the form

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) - \beta_k\left(\nabla f(x_k) - \nabla f(x_{k-1})\right), \\ x_{k+1} = y_k - s\nabla f(y_k). \end{cases}$$

As a specific feature, and by comparison with the accelerated gradient method of Nesterov, these algorithms contain a correction term which is equal to the difference of the gradients at two consecutive steps. While preserving the convergence properties of the Nesterov accelerated method, they provide fast convergence to zero of the gradients and reduce the oscillatory aspects. Several recent studies have been devoted to this subject; see Attouch et al. [8], Boţ, Csetnek, and László [23], Kim [31], Lin and Jordan [32], and Shi et al. [40].

1.3. Inertial dynamics and cocoercive operators. Let's come to the case of maximally monotone operators. Álvarez and Attouch [2] and Attouch and Maingé [11] studied the equation

$$\ddot{x}(t) + \gamma\dot{x}(t) + A(x(t)) = 0 \tag{1.2}$$

when A is a λ-cocoercive[1] (and hence maximally monotone) operator. An extension of (1.2) with variable damping parameter and also a variable relaxation parameter was studied by Boţ and Csetnek in [22]. Cocoercivity plays an important role in the study of (1.2), not only to ensure the existence of solutions, but also to analyze their long-term behavior. In [2] and [11] it has been shown that each trajectory of (1.2) converges weakly to a zero of A if the cocoercivity parameter λ and the damping coefficient γ satisfy the inequality $\lambda\gamma^2 > 1$. Since A_λ is λ-cocoercive and $A_\lambda^{-1}(0) = A^{-1}(0)$, we immediately deduce that, under the condition $\lambda\gamma^2 > 1$, given a general maximally monotone operator A, each trajectory of

$$\ddot{x}(t) + \gamma\dot{x}(t) + A_\lambda(x(t)) = 0$$

converges weakly to a zero of A. In the quest for a faster convergence, the analysis of

$$(\mathrm{DIN-AVD})_{\alpha,0} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0,$$

leads us to introduce a time-dependent parameter $\lambda(\cdot)$ satisfying $\lambda(t) \times \frac{\alpha^2}{t^2} > 1$; see Attouch and Peypouquet [14]. Temporal discretization of this dynamic gives the relaxed inertial proximal algorithm

$$(\mathrm{RIPA}) \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}), \\ x_{k+1} = (1-\rho_k)y_k + \rho_k J_{\mu_k A}(y_k), \end{cases}$$

whose convergence properties have been analyzed by Attouch and Peypouquet [14] Attouch and Cabot [5].

[1] $A : \mathcal{H} \to \mathcal{H}$ is λ-cocoercive (λ is a positive parameter) if for all $x, y \in \mathcal{H}$ $\langle Ay - Ax, y - x\rangle \geq \lambda\|Ay - Ax\|^2$.

1.4. Link with Newton-like methods for solving monotone inclusions. Let us specify the link between our study and Newton's method for solving (1.1). To overcome the ill-posed character of the continuous Newton method, the following first-order evolution system was studied by Attouch and Svaiter (see [18]), for a general maximally monotone operator A

$$\begin{cases} v(t) \in A(x(t)), \\ \gamma(t)\dot{x}(t) + \beta\dot{v}(t) + v(t) = 0. \end{cases}$$

This system can be considered as a continuous version of the Levenberg–Marquardt method, which acts as a regularization of the Newton method. Remarkably, under a fairly general assumption on the regularization parameter $\gamma(t)$, this system is well posed and generates trajectories that converge weakly to equilibria. Parallel results have been obtained for the associated proximal algorithms obtained by implicit temporal discretization; see [1, 13, 17, 20]. Formally, this system is written

$$\gamma(t)\dot{x}(t) + \beta\frac{d}{dt}\left(A(x(t))\right) + A(x(t)) = 0.$$

Thus, $(\mathrm{DIN}-\mathrm{AVD})_{\alpha,\beta}$ can be considered as an inertial and regularized version of this system.

1.5. Organization of the paper. The (PRINAM) algorithm is studied in section 2. In section 3 we examine the case $A = \partial f$ where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is convex lower semicontinuous. Each section is completed by numerical illustrations. Finally, we outline some perspectives.

2. Convergence of the associated proximal relaxed algorithm. The (PRINAM) algorithm will be introduced by implicit temporal discretization of $(\mathrm{DIN}-\mathrm{AVD})_{\alpha,\beta}$. In view of the Lipschitz continuity property of A_λ, the explicit discretization might work well too. In fact, the implicit discretization tends to follow the continuous-time trajectories more closely. In addition, the implicit and explicit discretizations have a comparable iteration complexity.

2.1. Regularized inertial proximal algorithms. Take a fixed time step $h > 0$, and set $t_k = kh$, $x_k = x(t_k)$, $\lambda_k = \lambda(t_k)$. Consider the implicit finite-difference scheme for $(\mathrm{DIN}-\mathrm{AVD})_{\alpha,\beta}$

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{kh^2}(x_k - x_{k-1}) + \frac{\beta}{h}(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})) + A_{\lambda_{k+1}}(x_{k+1}) = 0, \tag{2.1}$$

with centered second-order variation. After expanding (2.1), we obtain

$$x_{k+1} + h^2 A_{\lambda_{k+1}}(x_{k+1}) = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) - \beta h(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})). \tag{2.2}$$

Set $s = h^2$. Keeping the notation β for βh, and setting $\alpha_k := (1 - \frac{\alpha}{k})$, we have

$$x_{k+1} + sA_{\lambda_{k+1}}(x_{k+1}) = y_k, \tag{2.3}$$

where

$$y_k := x_k + \alpha_k(x_k - x_{k-1}) - \beta(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})). \tag{2.4}$$

From (2.3) we get

$$x_{k+1} = \left(I + sA_{\lambda_{k+1}}\right)^{-1}(y_k), \tag{2.5}$$

where $(I + sA_{\lambda_{k+1}})^{-1}$ is the resolvent of index $s > 0$ of the maximally monotone operator $A_{\lambda_{k+1}}$.

Putting (2.4) and (2.5) together, we obtain the following algorithm: given $x_1, x_2 \in \mathcal{H}$, for all $k \geq 2$

$$\begin{cases} y_k \; = \; x_k + \alpha_k(x_k - x_{k-1}) - \beta(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})), \\ x_{k+1} \; = \; \left(I + sA_{\lambda_{k+1}}\right)^{-1}(y_k). \end{cases} \tag{2.6}$$

Let us give some equivalent formulations of this algorithm. According to the resolvent equation (formulated as a semigroup property) $(A_\lambda)_s = A_{\lambda+s}$, we have

$$(I + sA_\lambda)^{-1} = I - s\,(A_\lambda)_s = I - sA_{\lambda+s}.$$

Thus, we obtain the following formulation, which makes use of the Yosida approximations of A. Given $x_1, x_2 \in \mathcal{H}$, for all $k \geq 2$

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) - \beta(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})), \\ x_{k+1} = y_k - sA_{\lambda_{k+1}+s}(y_k). \end{cases} \tag{2.7}$$

According to $A_\lambda = \frac{1}{\lambda}(I - J_{\lambda A})$, let us reformulate (2.7) using the resolvents of A. We have

$$\begin{aligned} A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1}) &= \frac{1}{\lambda_k}x_k - \frac{1}{\lambda_{k-1}}x_{k-1} - \left(\frac{1}{\lambda_k}J_{\lambda_k A}(x_k) - \frac{1}{\lambda_{k-1}}J_{\lambda_{k-1}A}(x_{k-1})\right) \\ &= \frac{1}{\lambda_{k-1}}(x_k - x_{k-1}) + \left(\frac{1}{\lambda_k} - \frac{1}{\lambda_{k-1}}\right)x_k \\ &\quad - \left(\frac{1}{\lambda_k}J_{\lambda_k A}(x_k) - \frac{1}{\lambda_{k-1}}J_{\lambda_{k-1}A}(x_{k-1})\right) \\ y_k - sA_{\lambda_{k+1}+s}(y_k) &= y_k - \frac{s}{\lambda_{k+1}+s}\left(y_k - J_{(\lambda_{k+1}+s)A}(y_k)\right) \\ &= \frac{\lambda_{k+1}}{\lambda_{k+1}+s}y_k + \frac{s}{\lambda_{k+1}+s}J_{(\lambda_{k+1}+s)A}(y_k). \end{aligned}$$

This gives the (PRINAM) algorithm. It is formulated below in terms of the resolvents of A.

(PRINAM)

Take $x_1 \in \mathcal{H}$, $x_2 \in \mathcal{H}, s > 0, \beta \geq 0$.

$$\text{Step } k: \begin{cases} y_k = \left(1 - \beta\left(\frac{1}{\lambda_k} - \frac{1}{\lambda_{k-1}}\right)\right)x_k + \left(\alpha_k - \frac{\beta}{\lambda_{k-1}}\right)(x_k - x_{k-1}) \\ \qquad + \beta\Big(\frac{1}{\lambda_k}J_{\lambda_k A}(x_k) - \frac{1}{\lambda_{k-1}}J_{\lambda_{k-1}A}(x_{k-1})\Big), \\ x_{k+1} = \dfrac{\lambda_{k+1}}{\lambda_{k+1}+s}y_k + \dfrac{s}{\lambda_{k+1}+s}J_{(\lambda_{k+1}+s)A}(y_k). \end{cases}$$

We are now in position to prove the main result of this section, the following theorem.

THEOREM 2.1. *Let* $A : \mathcal{H} \to 2^{\mathcal{H}}$ *be a maximally monotone operator such that* $S = A^{-1}(0) \neq \emptyset$. *Consider the algorithm* (PRINAM) *where, for all* $k \geq 1, \alpha_k = \frac{t_k - 1}{t_{k+1}}, t_k = rk + q$, $r > 0, q \in \mathbb{R}$, *and*

$$\lambda_k = \lambda k^2 \ \ \textit{with} \ \ \lambda > \frac{(2\beta + s)^2 r^2}{s}.$$

Then, for any sequences (x_k), (y_k) *generated by* (PRINAM), *the following properties are satisfied:*

(i) *The speed* $(x_{k+1} - x_k)_{k \geq 1}$ *tends to zero, and we have the estimates*

$$\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right) \ \ \textit{as } k \to +\infty, \quad \sum_{k \geq 2} k \|x_k - x_{k-1}\|^2 < +\infty,$$

$$\|A_{\lambda_k}(x_k)\| = o\left(\frac{1}{k^2}\right) \ \ \textit{as } k \to +\infty, \quad \sum_{k \geq 1} k^3 \|A_{\lambda_k}(x_k)\|^2 < +\infty.$$

(ii) *The sequence* (x_k) *converges weakly to some* $\hat{x} \in S$, *as* $k \to +\infty$.

(iii) *The sequence* (y_k) *converges weakly to* $\hat{x} \in S$, *as* $k \to +\infty$.
Precisely, $\|y_k - x_k\| = \mathcal{O}\left(\frac{1}{k}\right)$, *as* $k \to +\infty$, *and so* $y_k - x_k$ *converges strongly to zero.*

2.2. Geometric interpretation. In algorithm (PRINAM), the proximal parameter λ_k tends to infinity in a controlled way, namely $\lambda_k = \lambda k^2$ with λ sufficiently large. This property balances the vanishing property of the damping coefficient. As a classical property of the resolvents [19, Theorem 23.44], for any $x \in \mathcal{H}$, $J_{\lambda A} x \to \mathrm{proj}_S(x)$ as $\lambda \to +\infty$, where S is the set of zeros of A. Thus the algorithm is written

$$x_{k+1} = \theta_k y_k + (1 - \theta_k) J_{(\lambda_{k+1} + s)A}(y_k) = y_k + \frac{s}{\lambda_{k+1} + s}\left(J_{(\lambda_{k+1}+s)A}(y_k) - y_k\right)$$

with $\lambda_k \sim +\infty$, $\theta_k = \frac{\lambda_{k+1}}{\lambda_{k+1} + s} \sim 1$, $\frac{s}{\lambda_{k+1}+s} \sim 0$, $J_{(\lambda_{k+1}+s)A}(y_k) \sim \mathrm{proj}_S(y_k)$ as $k \to +\infty$. At step k, after reaching y_k, the direction $J_{(\lambda_{k+1}+s)A}(y_k) - y_k \sim \mathrm{proj}_S(y_k) - y_k$ is well oriented in the direction of S, but we are allowed to take only a small step in this direction. This is illustrated in Figure 2.1.

Remark 1. Following [5, 6, 7], we could develop our theory with a general sequence (α_k) of extrapolation coefficients which satisfy $0 \leq \alpha_k \leq 1$. A particularly interesting situation is the case $\alpha_k \to 1$, which corresponds to the asymptotic vanishing damping in the associated dynamic system. The sequence (t_k) which is linked to the sequence (α_k) by the relation

$$\alpha_k = \frac{t_k - 1}{t_{k+1}} \tag{2.8}$$

plays a central role. To simplify the presentation, in Theorem 2.1 we limit our study to the case

$$t_k = rk + q, \ r > 0, \ q \in \mathbb{R}, \tag{2.9}$$

which contains most interesting situations. In particular, when $\alpha_k = 1 - \frac{\alpha}{k}$, we have $t_k = \frac{k-1}{\alpha - 1}$, which corresponds to $r = \frac{1}{\alpha-1}$, $q = -\frac{1}{\alpha-1}$. The critical value $\alpha = 3$ corresponds to $r = \frac{1}{2}$.

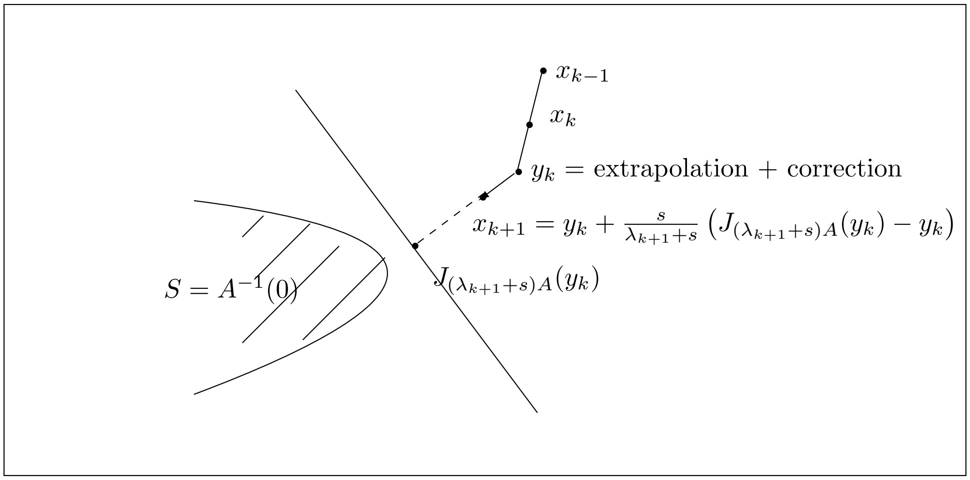


FIG. 2.1. (PRINAM) *algorithm.*

2.3. Proof of Theorem 2.1.

The discrete energy. Take $z \in S$. For each $k \geq 2$, let us define the discrete energy

(2.10)
$$\mathcal{E}_{a,b}^k := at_{k-1}\left\langle A_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\right\rangle + \frac{1}{2}\|b(x_{k-1} - z) + t_k(x_k - x_{k-1} + sA_{\lambda_k}(x_k))\|^2 + \frac{b(1-b)}{2}\|x_{k-1} - z\|^2.$$

Note that if we assume that $a \geq 0$ and $0 < b < 1$, then, due to the monotonicity of $A_{\lambda_{k-1}}$ and the fact that t_{k-1} is positive after an index k', we obtain that for all $k \geq k'$ the discrete energy $\mathcal{E}_{a,b}^k$ is a sum of nonnegative terms. We will show that by adjusting the real parameters a and b, the following Lyapunov property is satisfied: there exist $\epsilon_1, \epsilon_2 > 0$ and an index $N \in \mathbb{N}$ such that

$$\mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^k + \epsilon_1 k^3\|A_{\lambda_k}(x_k)\|^2 + \epsilon_2 k\|x_k - x_{k-1}\|^2 \leq 0 \quad \text{for all} \quad k \geq N.$$

Specifically, in what follows, we will take

$$0 < b < 1 \quad \text{and} \quad b\beta < a < b\beta + bs, \quad \text{whenever } \beta > 0; \tag{2.11}$$

$$0 < b < 1 \quad \text{and} \quad a = 0 \quad \text{for } \beta = 0. \tag{2.12}$$

For each $k \geq 2$, we briefly write $\mathcal{E}_{a,b}^k$ as follows:

$$\mathcal{E}_{a,b}^k = at_{k-1}\left\langle A_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\right\rangle + \frac{1}{2}\|v_k\|^2 + \frac{b(1-b)}{2}\|x_{k-1} - z\|^2,$$

with $v_k := b(x_{k-1} - z) + t_k(x_k - x_{k-1} + sA_{\lambda_k}(x_k))$.

Using successively the definition of v_k, (2.3), (2.4), and (2.8) we obtain

$$\begin{aligned} v_{k+1} &= b(x_k - z) + t_{k+1}(x_{k+1} - x_k + sA_{\lambda_{k+1}}(x_{k+1})) \\ &= b(x_k - z) + t_{k+1}(y_k - x_k) \\ &= b(x_k - z) + t_{k+1}\left(\alpha_k(x_k - x_{k-1}) - \beta\left(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})\right)\right) \\ &= b(x_k - z) + (t_k - 1)(x_k - x_{k-1}) - \beta t_{k+1}(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})). \end{aligned} \tag{2.13}$$

Further, v_k can be written as

$$v_k = b(x_k - z) + (t_k - b)(x_k - x_{k-1}) + st_k A_{\lambda_k}(x_k). \tag{2.14}$$

Therefore, for all $k \geq 2$, we have

$$\begin{aligned}
&\frac{1}{2}\|v_{k+1}\|^2 - \frac{1}{2}\|v_k\|^2 \\
&\quad = \frac{1}{2}\|b(x_k - z) + (t_k - 1)(x_k - x_{k-1}) - \beta t_{k+1}(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1}))\|^2 \\
&\qquad - \frac{1}{2}\|b(x_k - z) + (t_k - b)(x_k - x_{k-1}) + st_k A_{\lambda_k}(x_k)\|^2 \\
&\quad = \frac{1}{2}((t_k - 1)^2 - (t_k - b)^2)\|x_k - x_{k-1}\|^2 \\
&\qquad + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2)\|A_{\lambda_k}(x_k)\|^2 - \beta^2 t_{k+1}^2 \langle A_{\lambda_k}(x_k), A_{\lambda_{k-1}}(x_{k-1})\rangle \\
&\qquad + \frac{1}{2}\beta^2 t_{k+1}^2 \|A_{\lambda_{k-1}}(x_{k-1})\|^2 + b(b-1)\langle x_k - x_{k-1}, x_k - z\rangle \\
&\qquad - b(\beta t_{k+1} + st_k)\langle A_{\lambda_k}(x_k), x_k - z\rangle + b\beta t_{k+1}\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - z\rangle \\
&\qquad - (\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle A_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
&\qquad + \beta t_{k+1}(t_k - 1)\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - x_{k-1}\rangle.
\end{aligned} \tag{2.15}$$

According to the elementary identities

$$\begin{aligned}
b(b-1)\langle x_k - x_{k-1}, x_k - z\rangle &= b(b-1)\|x_k - x_{k-1}\|^2 + b(b-1)\langle x_k - x_{k-1}, x_{k-1} - z\rangle, \\
b\beta t_{k+1}\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - z\rangle &= b\beta t_{k+1}\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - x_{k-1}\rangle \\
&\quad + b\beta t_{k+1}\langle A_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\rangle
\end{aligned}$$

formula (2.15) becomes

$$\begin{aligned}
&\frac{1}{2}\|v_{k+1}\|^2 - \frac{1}{2}\|v_k\|^2 \\
&\quad = \frac{1}{2}(b-1)(2t_k + b - 1)\|x_k - x_{k-1}\|^2 \\
&\qquad + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2)\|A_{\lambda_k}(x_k)\|^2 - \beta^2 t_{k+1}^2 \langle A_{\lambda_k}(x_k), A_{\lambda_{k-1}}(x_{k-1})\rangle \\
&\qquad + \frac{1}{2}\beta^2 t_{k+1}^2 \|A_{\lambda_{k-1}}(x_{k-1})\|^2 + b(b-1)\langle x_k - x_{k-1}, x_{k-1} - z\rangle \\
&\qquad - b(\beta t_{k+1} + st_k)\langle A_{\lambda_k}(x_k), x_k - z\rangle + b\beta t_{k+1}\langle A_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\rangle \\
&\qquad - (\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle A_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
&\qquad + \beta t_{k+1}(t_k + b - 1)\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - x_{k-1}\rangle.
\end{aligned} \tag{2.16}$$

Moreover, we have for all $k \geq 2$

$$\begin{aligned}
&\frac{b(1-b)}{2}\|x_k - z\|^2 - \frac{b(1-b)}{2}\|x_{k-1} - z\|^2 \\
&\quad = \frac{b(1-b)}{2}\|(x_k - x_{k-1}) + (x_{k-1} - z)\|^2 - \frac{b(1-b)}{2}\|x_{k-1} - z\|^2 \\
&\quad = \frac{b(1-b)}{2}\|x_k - x_{k-1}\|^2 + b(1-b)\langle x_k - x_{k-1}, x_{k-1} - z\rangle.
\end{aligned} \tag{2.17}$$

By combining the above results (the terms $\langle x_k - x_{k-1}, x_{k-1} - z\rangle$ cancel out), we get for all $k \geq 2$

(2.18)
$$\begin{aligned}
\mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^{k} = {} & (at_k - b(\beta t_{k+1} + st_k)) \langle A_{\lambda_k}(x_k), x_k - z\rangle \\
& + (b\beta t_{k+1} - at_{k-1}) \left\langle A_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\right\rangle \\
& + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2)\|A_{\lambda_k}(x_k)\|^2 - \beta^2 t_{k+1}^2 \langle A_{\lambda_k}(x_k), A_{\lambda_{k-1}}(x_{k-1})\rangle \\
& + \frac{1}{2}\beta^2 t_{k+1}^2 \|A_{\lambda_{k-1}}(x_{k-1})\|^2 \\
& - (\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle A_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
& + \beta t_{k+1}(t_k + b - 1)\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - x_{k-1}\rangle + \frac{1}{2}(b-1)(2t_k - 1)\|x_k - x_{k-1}\|^2.
\end{aligned}$$

According to the assumptions (2.11) and (2.12) on the parameters a and b, there exists $k_1 \geq 2$ such that for all $k \geq k_1$

$$at_k - b(\beta t_{k+1} + st_k) < 0 \text{ and } b\beta t_{k+1} - at_{k-1} \leq 0,$$

where in the last relation the equality holds only in case $a = 0$, $\beta = 0$. According to the cocoerciveness of A_{λ_k} and $A_{\lambda_{k-1}}$ and $z \in S$, we deduce from the above inequalities that, for all $k \geq k_1$,

$$(at_k - b(\beta t_{k+1} + st_k)) \langle A_{\lambda_k}(x_k), x_k - z\rangle \leq (at_k - b(\beta t_{k+1} + st_k))\lambda_k \|A_{\lambda_k}(x_k)\|^2,$$
$$(b\beta t_{k+1} - at_{k-1}) \left\langle A_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\right\rangle \leq (b\beta t_{k+1} - at_{k-1})\lambda_{k-1}\|A_{\lambda_{k-1}}(x_{k-1})\|^2.$$

Therefore, (2.18) yields, for all $k \geq k_1$,

(2.19)
$$\begin{aligned}
\mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^{k} \leq {} & \left((at_k - b(\beta t_{k+1} + st_k))\lambda_k + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2)\right) \|A_{\lambda_k}(x_k)\|^2 \\
& + \left((b\beta t_{k+1} - at_{k-1})\lambda_{k-1} + \frac{1}{2}\beta^2 t_{k+1}^2\right) \|A_{\lambda_{k-1}}(x_{k-1})\|^2 \\
& + \frac{1}{2}(b-1)(2t_k - 1)\|x_k - x_{k-1}\|^2 - \beta^2 t_{k+1}^2 \langle A_{\lambda_k}(x_k), A_{\lambda_{k-1}}(x_{k-1})\rangle \\
& - (\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle A_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
& + \beta t_{k+1}(t_k + b - 1)\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - x_{k-1}\rangle.
\end{aligned}$$

Further, for all $p_1, p_2 > 0$ and $k \geq k_1$ we have the elementary inequalities

(2.20)
$$-\beta^2 t_{k+1}^2 \langle A_{\lambda_k}(x_k), A_{\lambda_{k-1}}(x_{k-1})\rangle \leq \frac{\beta^2}{2} t_{k+1}^2 \left(\|A_{\lambda_k}(x_k)\|^2 + \|A_{\lambda_{k-1}}(x_{k-1})\|^2\right);$$

(2.21)
$$\begin{aligned}
& -(\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle A_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
& \quad \leq |\beta t_{k+1}(t_k - 1) + st_k(t_k - b)| \left(p_1 k\|A_{\lambda_k}(x_k)\|^2 + \frac{1}{4p_1 k}\|x_k - x_{k-1}\|^2\right);
\end{aligned}$$

(2.22)
$$\begin{aligned}
& \beta t_{k+1}(t_k + b - 1)\langle A_{\lambda_{k-1}}(x_{k-1}), x_k - x_{k-1}\rangle \\
& \quad \leq |\beta t_{k+1}(t_k + b - 1)| \left(p_2 k\|A_{\lambda_{k-1}}(x_{k-1})\|^2 + \frac{1}{4p_2 k}\|x_k - x_{k-1}\|^2\right).
\end{aligned}$$

Combining (2.19) with (2.20)–(2.21)–(2.22), we deduce that, for all $k \geq k_1$,

$$\mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^{k} \leq P(k)\|A_{\lambda_k}(x_k)\|^2 + Q(k)\|A_{\lambda_{k-1}}(x_{k-1})\|^2 + R(k)\|x_k - x_{k-1}\|^2, \tag{2.23}$$

where

$$\begin{aligned}
P(k) &= (at_k - b(\beta t_{k+1} + st_k))\lambda_k + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2) + \frac{\beta^2}{2}t_{k+1}^2 + |\beta t_{k+1}(t_k - 1)\\
&\qquad + st_k(t_k - b)|p_1 k;\\
Q(k) &= (b\beta t_{k+1} - at_{k-1})\lambda_{k-1} + \frac{1}{2}\beta^2 t_{k+1}^2 + \frac{\beta^2}{2}t_{k+1}^2 + |\beta t_{k+1}(t_k + b - 1)|p_2 k;\\
R(k) &= \frac{1}{2}(b-1)(2t_k - 1) + \frac{|\beta t_{k+1}(t_k - 1) + st_k(t_k - b)|}{4p_1 k} + \frac{|\beta t_{k+1}(t_k + b - 1)|}{4p_2 k}.
\end{aligned}$$

Elementary computation gives the following asymptotic developments:

$$\begin{aligned}
P(k) &= ((a - b\beta - bs)r\lambda + (\beta + s)r^2 p_1)k^3 + r_1(k),\\
Q(k) &= ((b\beta - a)r\lambda + \beta r^2 p_2)k^3 + r_2(k),\\
R(k) &= \left((b-1)r + \frac{(\beta+s)r^2}{4p_1} + \frac{\beta r^2}{4p_2}\right)k + r_3(k),
\end{aligned}$$

where $r_1(k), r_2(k) = \mathcal{O}(k^2)$, and $r_3(k) = \mathcal{O}(1)$ as $k \to +\infty$. Let us adjust the parameters by taking

$$\begin{aligned}
&b = \frac{1}{2} \in (0,1), \quad a = \frac{\beta(\beta+s)}{2\beta+s} \in (b\beta, b(\beta+s)) \ \text{ whenever } \beta > 0;\\
&b = \frac{1}{2} \ \text{ and } a = 0 \text{ whenever } \beta = 0.
\end{aligned}$$

In this last case $Q \equiv 0$. For the rest of the proof we do not need to distinguish the cases $\beta > 0$ and $\beta = 0$. Further, take

$$p_1 = p_2 = \frac{\lambda s}{2\epsilon(2\beta+s)r} \ \text{ with } \ 1 < \epsilon < \frac{\lambda s}{(2\beta+s)^2 r^2}.$$

This is possible, thanks to our basic assumption on λ, namely $\lambda > \frac{(2\beta+s)^2 r^2}{s}$. With this choice of parameters, we have for all $k \geq k_1$

$$\begin{aligned}
P(k) &= \frac{(\beta+s)sr\lambda}{2(2\beta+s)}\left(\frac{1}{\epsilon} - 1\right)k^3 + r_1(k),\\
Q(k) &= \frac{\beta sr\lambda}{2(2\beta+s)}\left(\frac{1}{\epsilon} - 1\right)k^3 + r_2(k),\\
R(k) &= \left(-\frac{1}{2}r + \frac{(2\beta+s)^2\epsilon r^3}{2s\lambda}\right)k + r_3(k).
\end{aligned}$$

Since $\epsilon > 1$, we have $\frac{(\beta+s)sr\lambda}{2(2\beta+s)}(\frac{1}{\epsilon} - 1) < 0$ and $\frac{\beta sr\lambda}{2(2\beta+s)}(\frac{1}{\epsilon} - 1) \leq 0$. Since $\epsilon < \frac{\lambda s}{(2\beta+s)^2 r^2}$, we have $-\frac{1}{2}r + \frac{(2\beta+s)^2\epsilon r^3}{2s\lambda} < 0$. So, there exists $\epsilon_1, \epsilon_2 > 0$ such that

$$\frac{(\beta+s)sr\lambda}{2(2\beta+s)}\left(\frac{1}{\epsilon} - 1\right) + \epsilon_1 < 0 \ \text{ and } \ \left(-\frac{1}{2}r + \frac{(2\beta+s)^2\epsilon r^3}{2s\lambda}\right) + \epsilon_2 < 0.$$

According to the above inequalities, (2.23) leads to

$$
\begin{aligned}
&\mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^{k} + \epsilon_1 k^3 \|A_{\lambda_k}(x_k)\|^2 + \epsilon_2 k \|x_k - x_{k-1}\|^2 \\
&\quad \le \left(\left(\frac{(\beta+s)sr\lambda}{2(2\beta+s)} \left(\frac{1}{\epsilon} - 1 \right) + \epsilon_1 \right) k^3 + r_1(k) \right) \|A_{\lambda_k}(x_k)\|^2 \\
&\qquad + \left(\frac{\beta sr\lambda}{2(2\beta+s)} \left(\frac{1}{\epsilon} - 1 \right) k^3 + r_2(k) \right) \|A_{\lambda_{k-1}}(x_{k-1})\|^2 \\
&\qquad + \left(\left(-\frac{1}{2}r + \frac{(2\beta+s)^2 \epsilon r^3}{2s\lambda} + \epsilon_2 \right) k + r_3(k) \right) \|x_k - x_{k-1}\|^2.
\end{aligned} \tag{2.24}
$$

Take $N \ge k_1$ such that, for all $k \ge N$

$$
\begin{aligned}
&\left(\frac{(\beta+s)sr\lambda}{2(2\beta+s)} \left(\frac{1}{\epsilon} - 1 \right) + \epsilon_1 \right) k^3 + r_1(k) \le 0, \\
&\frac{\beta sr\lambda}{2(2\beta+s)} \left(\frac{1}{\epsilon} - 1 \right) k^3 + r_2(k) \le 0, \\
&\left(-\frac{1}{2}r + \frac{(2\beta+s)^2 \epsilon r^3}{2s\lambda} + \epsilon_2 \right) k + r_3(k) \le 0.
\end{aligned}
$$

Then, for all $k \ge N$

$$
\mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^{k} + \epsilon_1 k^3 \|A_{\lambda_k}(x_k)\|^2 + \epsilon_2 k \|x_k - x_{k-1}\|^2 \le 0. \tag{2.25}
$$

Estimates. According to (2.25), the sequence of nonnegative numbers $(\mathcal{E}_{a,b}^{k})_{k\in\mathbb{N}}$ is nonincreasing and therefore converges. In particular, it is bounded. From this and by adding the inequalities (2.25) we obtain

$$
\sup_k t_k \left\langle A_{\lambda_k}(x_k), x_k - z \right\rangle < +\infty, \tag{2.26}
$$

$$
\sup_k \|b(x_k - z) + t_{k+1}(x_{k+1} - x_k + sA_{\lambda_{k+1}}(x_{k+1}))\|^2 < +\infty, \tag{2.27}
$$

$$
\sup_k \|x_k - z\|^2 < +\infty, \tag{2.28}
$$

$$
\sum_{k=1}^{+\infty} k^3 \|A_{\lambda_k}(x_k)\|^2 < +\infty, \tag{2.29}
$$

$$
\sum_{k=2}^{+\infty} k \|x_k - x_{k-1}\|^2 < +\infty. \tag{2.30}
$$

Since the general term of a convergent series goes to zero, we deduce from (2.29) that

$$
\|A_{\lambda_k}(x_k)\| = o\left(\frac{1}{k^{\frac{3}{2}}} \right) \text{ as } k \to +\infty. \tag{2.31}
$$

In fact, we will get better estimates a little further. According to (2.28), the sequence $(\|x_k - z\|)$ is bounded, and so is the sequence (x_k). Combining the above results with (2.27), we deduce that

$$
\|x_k - x_{k-1}\| = \mathcal{O}\left(\frac{1}{k} \right) \text{ as } k \to +\infty. \tag{2.32}
$$

From (x_k) bounded, and A_{λ_k} $\frac{1}{\lambda_k}$- Lipschitz continuous, we obtain the existence of $M > 0$ such that

$$\|\lambda_k A_{\lambda_k}(x_k)\| = \|\lambda_k A_{\lambda_k}(x_k) - \lambda_k A_{\lambda_k}(z)\| \leq \lambda_k \frac{1}{\lambda_k}\|x_k - z\| \leq M, \tag{2.33}$$

which yields

$$\|A_{\lambda_k}(x_k)\| = \mathcal{O}\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty. \tag{2.34}$$

Let us show the following better estimate which will play a key role in the rest of the proof:

$$\|A_{\lambda_k}(x_k)\| = o\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty.$$

To obtain it, we follow the line of proof of [14, Theorem 3.6]. From Lemma A.4 in [14], for all $k \geq 2$

$$\|\lambda_k A_{\lambda_k}(x_k) - \lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\| \leq 2\|x_k - x_{k-1}\| + 2\|x_k - z\|\frac{|\lambda_k - \lambda_{k-1}|}{\lambda_k}. \tag{2.35}$$

According to $\|x_k - x_{k-1}\| = \mathcal{O}\left(\frac{1}{k}\right)$ as $k \to +\infty$, (x_k) is bounded, and $\lambda_k = \lambda k^2$, we conclude that there exists $C > 0$ such that

$$\|\lambda_k A_{\lambda_k}(x_k) - \lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\| \leq \frac{C}{k} \text{ for all } k \geq 2. \tag{2.36}$$

According to (2.33) and (2.36), we deduce that

$$\begin{aligned}
&\left| \|\lambda_k A_{\lambda_k}(x_k)\|^2 - \|\lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\|^2 \right| \\
&\quad = \Big(\|\lambda_k A_{\lambda_k}(x_k)\| + \|\lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\|\Big)\Big|\|\lambda_k A_{\lambda_k}(x_k)\| - \|\lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\|\Big| \\
&\quad \leq 2M\|\lambda_k A_{\lambda_k}(x_k) - \lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\| \leq \frac{2MC}{k} \text{ for all } k \geq 2.
\end{aligned}$$

Consequently, by using (2.29) we get

$$\begin{aligned}
&\sum_{k\geq 2} \left| \|\lambda_k A_{\lambda_k}(x_k)\|^4 - \|\lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\|^4 \right| \\
&\quad = \sum_{k\geq 2} (\|\lambda_k A_{\lambda_k}(x_k)\|^2 + \|\lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\|^2)\Big|\|\lambda_k A_{\lambda_k}(x_k)\|^2 - \|\lambda_{k-1} A_{\lambda_{k-1}}(x_{k-1})\|^2\Big| \\
&\quad \leq \sum_{k\geq 2} \frac{2MC\lambda^2 k^4}{k}\|A_{\lambda_k}(x_k)\|^2 + \sum_{k\geq 2} \frac{2MC\lambda^2 (k-1)^4}{k}\|A_{\lambda_{k-1}}(x_{k-1})\|^2 < +\infty.
\end{aligned}$$

From this, by a telescopic argument we conclude that $\lim_{k\to+\infty} \|\lambda_k A_{\lambda_k}(x_k)\|^4$ exists. But then $\lim_{k\to+\infty} \|\lambda_k A_{\lambda_k}(x_k)\|^2$ and $\lim_{k\to+\infty} \|\lambda_k A_{\lambda_k}(x_k)\|$ also exist. Set

$$\lim_{k\to+\infty} k^4\|A_{\lambda_k}(x_k)\|^2 := L \geq 0.$$

According to (2.29) we will have $\sum_{k\geq 1} \frac{1}{k}(k^4\|A_{\lambda_k}(x_k)\|^2) = \sum_{k\geq 1} k^3\|A_{\lambda_k}(x_k)\|^2 < +\infty$, which implies that $L = 0$. Hence, $\lim_{k\to+\infty} k^2\|A_{\lambda_k}(x_k)\| = 0$, that is,

$$\|A_{\lambda_k}(x_k)\| = o\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty. \tag{2.37}$$

Convergence of (x_k). Using the Opial lemma, let us prove that the sequence (x_k) converges weakly toward an element of S. Take $z \in S$, and consider the anchor sequence (h_k) defined by $h_k = \frac{1}{2}\|x_k - z\|^2$ for $k \geq 1$. Elementary algebra gives

$$h_{k+1} - h_k = \frac{1}{2}\|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x_k, x_k - z\rangle. \tag{2.38}$$

According to (2.6) we have for all $k \geq 2$

$$\begin{aligned} \langle x_{k+1} - x_k, x_k - z\rangle &= \langle y_k - x_k - sA_{\lambda_{k+1}}(x_{k+1}), x_k - z\rangle \\ &= \alpha_k \langle x_k - x_{k-1}, x_k - z\rangle - \beta\langle A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1}), x_k - z\rangle \\ &\quad - s\langle A_{\lambda_{k+1}}(x_{k+1}), x_k - z\rangle. \end{aligned} \tag{2.39}$$

Let us examine the terms involved in the above equality. We have for all $k \geq 2$

$$\begin{aligned} &x_k - x_{k-1}, x_k - z\rangle = h_k - h_{k-1} + \frac{1}{2}\|x_k - x_{k-1}\|^2; \\ &-s\langle A_{\lambda_{k+1}}(x_{k+1}), x_k - z\rangle = s\langle A_{\lambda_{k+1}}(x_{k+1}), x_{k+1} - x_k\rangle - s\langle A_{\lambda_{k+1}}(x_{k+1}), x_{k+1} - z\rangle. \end{aligned}$$

Combining these relations with (2.38) and (2.39), and neglecting the term $-s\langle A_{\lambda_{k+1}}(x_{k+1}), x_{k+1} - z\rangle$ which is nonpositive, we obtain for all $k \geq 2$

$$\begin{aligned} h_{k+1} - h_k &\leq \alpha_k(h_k - h_{k-1}) + \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 \\ &\quad - \beta\langle A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1}), x_k - z\rangle + s\langle A_{\lambda_{k+1}}(x_{k+1}), x_{k+1} - x_k\rangle. \end{aligned} \tag{2.40}$$

According to $\|x_k - z\|$ bounded, $\|A_{\lambda_{k+1}}(x_{k+1})\| = o\left(\frac{1}{k^2}\right)$, and $\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right)$, we obtain the existence of a constant $M > 0$ such that for all $k \geq 2$

$$\begin{aligned} h_{k+1} - h_k &\leq \alpha_k(h_k - h_{k-1}) + \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 \\ &\quad + M\|A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})\| + M\frac{1}{k^3}. \end{aligned} \tag{2.41}$$

In addition, by (2.36) and by the fact that $\lambda_k = \lambda k^2$, we get

$$\|\lambda k^2 A_{\lambda_k}(x_k) - \lambda(k-1)^2 A_{\lambda_{k-1}}(x_{k-1})\| \leq \frac{C}{k} \text{ for all } k \geq 2.$$

Equivalently,

$$\|(2\lambda k - \lambda)A_{\lambda_k}(x_k) + \lambda(k-1)^2(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1}))\| \leq \frac{C}{k} \text{ for all } k \geq 2.$$

Using again that $\|A_{\lambda_{k+1}}(x_{k+1})\| = o\left(\frac{1}{k^2}\right)$, we deduce that, for some $K > 0$,

$$\|A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})\| \leq \frac{K}{k^3}. \tag{2.42}$$

Therefore, (2.41) leads to

$$h_{k+1} - h_k \leq \alpha_k(h_k - h_{k-1}) + \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 + MK\frac{1}{k^3} + M\frac{1}{k^3}. \tag{2.43}$$

Let us analyze this inequality with the help of Lemma A.1. Set

$$\omega_k := \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 + MK\frac{1}{k^3} + M\frac{1}{k^3}.$$

As a direct result of the estimates we have already obtained, we have $\sum_{k\geq 2} t_{k+1}\omega_k < +\infty$. Therefore, by applying Lemma A.1 to the sequence $a_k = [h_k - h_{k-1}]_+$ we obtain

$$\sum_{k\geq 2}[h_k - h_{k-1}]_+ < +\infty.$$

Since h_k is nonnegative, this property classically gives the existence of $\lim_{k\to+\infty} h_k$, and therefore the existence of $\lim_{k\to+\infty}\|x_k - z\|$. This shows item (i) of the Opial lemma.

It remains to show that every weak cluster point of the sequence (x_k) belongs to S. Let x^* be a weak cluster point of (x_k) and consider a subsequence (x_{k_n}) of (x_k) such that $x_{k_n} \rightharpoonup x^*$, $n \to +\infty$. According to (2.37) we have

$$\lim_{k\to+\infty} \lambda_k A_{\lambda_k}(x_k) = 0.$$

Now use $A_{\lambda_{k_n}}(x_{k_n}) \in A(J_{\lambda_{k_n}A}(x_{k_n}))$. Equivalently,

$$A_{\lambda_{k_n}}(x_{k_n}) \in A(x_{k_n} - \lambda_{k_n}A_{\lambda_{k_n}}(x_{k_n})). \tag{2.44}$$

According to the demi-closedness of the graph of A, passing to the limit in (2.44) gives $0 \in A(x^*)$. According to Opial's lemma, we finally obtain that (x_k) converges weakly to an element $\hat{x}$ in S.

Finally, by definition of y_k, we have

$$y_k - x_k = \alpha_k(x_k - x_{k-1}) - \beta(A_{\lambda_k}(x_k) - A_{\lambda_{k-1}}(x_{k-1})),$$

which, combined with (2.32), (2.42), and the fact that (α_k) is bounded, gives

$$\|y_k - x_k\| = \mathcal{O}\left(\frac{1}{k}\right), \text{ as } k \to +\infty.$$

Therefore, (y_k) also converges weakly toward the same element $\hat{x}$ in S.

2.4. Comparison with related algorithms.

2.4.1. Comparison with (RIPA). By taking $\beta = 0$ and $r = \frac{1}{\alpha-1}$, $q = -\frac{1}{\alpha-1}$ in (PRINAM), we obtain the algorithm (RIPA) considered by Attouch and Peypouquet in [14]. This algorithm and its convergence properties are recalled below: given $x_1, x_2 \in \mathcal{H}$, for $k \geq 2$ let

$$\text{(RIPA)} \qquad \begin{cases} y_k \;=\; x_k + \left(1 - \dfrac{\alpha}{k}\right)(x_k - x_{k-1}), \\ x_{k+1} \;=\; \dfrac{\lambda_k}{\lambda_k + s} y_k + \dfrac{s}{\lambda_k + s} J_{(\lambda_k+s)A}(y_k). \end{cases}$$

THEOREM (Attouch–Peypouquet [14]). *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator with $S = A^{-1}(0) \neq \emptyset$. Let (x_k) be a sequence generated by* (RIPA) *where $s > 0$, $\alpha > 2$ and for all $k \geq 2$*

$$\lambda_k = \lambda k^2 \text{ for some } \lambda > \frac{s}{\alpha(\alpha-2)}.$$

Then, the sequences (x_k) *and* (y_k) *converge weakly, as* $k \to +\infty$*, to some* $\hat{x} \in S$*. In addition,* $\|x_{k+1} - x_k\| = \mathcal{O}(\frac{1}{k})$ *as* $k \to +\infty$*, and* $\sum_{k\geq 2} k\|x_k - x_{k-1}\|^2 < +\infty$*.*

A natural question is to compare (PRINAM) to (RIPA) and show what the introduction of the correcting term in (PRINAM) ($\beta > 0$) brings. We emphasize that, for small β and $r = \frac{1}{\alpha-1}$, the lower bound for λ obtained in Theorem 2.1, namely $\lambda > \frac{(2\beta+s)^2 r^2}{s}$, is better than the lower bound obtained in the above result, namely $\lambda > \frac{s}{\alpha(\alpha-2)}$. Further, in (PRINAM) the more general condition $\alpha > 1$ is allowed. As a model example of a maximally monotone operator which is not the subdifferential of a convex function, consider $A : \mathbb{R}^2 \to \mathbb{R}^2$ given for any $x = (\xi, \eta) \in \mathbb{R}^2$ by

$$A(\xi, \eta) = (-\eta, \xi).$$

A is a skew-symmetric linear operator whose single zero is $x^* = (0, 0)$. An easy computation shows that A and A_λ can be identified respectively with the matrices

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad A_\lambda = \begin{pmatrix} \frac{\lambda}{1+\lambda^2} & \frac{-1}{1+\lambda^2} \\ \frac{1}{1+\lambda^2} & \frac{\lambda}{1+\lambda^2} \end{pmatrix}.$$

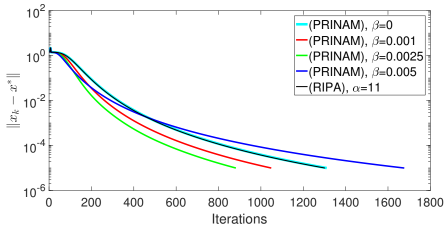
Let's compare (PRINAM) and (RIPA) by considering different instances of the parameters involved.

- Take $\alpha = 3$, then $\alpha = 11$ in (RIPA), which corresponds to respectively $t_k = 0.5k - 0.5$, $t_k = 0.1k - 0.1$ ($r = 0.5$, $r = 0.1$, $q = -0.5$, $q = -0.1$), in (PRINAM).
- Take $\lambda_k = \lambda k^2$ with λ chosen as follows: to satisfy the condition $\lambda > \frac{s}{\alpha(\alpha-2)}$ in (RIPA), we take $\lambda = 1.01\frac{s}{\alpha(\alpha-2)}$ in (RIPA).
  To satisfy the condition $\lambda > \frac{(2\beta+s)^2 r^2}{s}$ in (PRINAM), we take $\lambda = 1.01\frac{(2\beta+s)^2 r^2}{s}$.
- For the step size s, we consider the following instances: $s \in \{0.01, 0.1, 0.5, 1\}$.
  For (PRINAM) we consider the values $\beta \in \{0, 0.1s, 0.25s, 0.35s, 0.5s\}$.

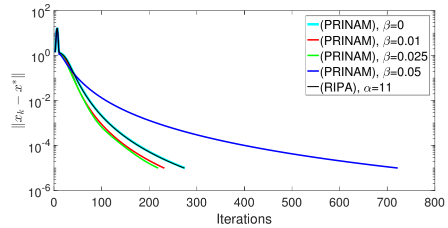
To start the algorithm we take $x_1 = (1, -1)$, $x_2 = (-1, 1)$. We run the algorithms until the iteration error $\|x_k - x^*\|$ reaches the value 10^{-5}.

The results are depicted at Figures 2.2(a)–2.2(d). The horizontal and vertical axes show respectively the number of iterations and the value of the error $\|x_k - x^*\|$. Despite the fact that it is difficult to draw general conclusions from a single numerical experiment, the above result shows the numerical interest of the introduction of the correcting term ($\beta > 0$) and also that the step size s must be taken not too large (therefore not remaining too far from the continuous dynamics). We only report here numerical examples where s is relatively small; for large values of s the convergence properties are less good. Further, r should be taken small (or α large) in order to obtain fast convergence.

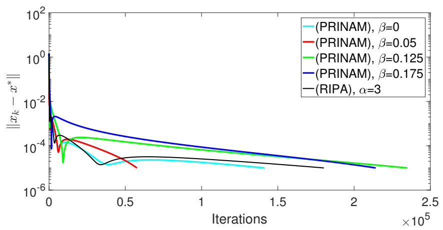
2.4.2. Comparison with (PPM). Let's compare (PRINAM) with the proximal point method (PPM), which is the basic algorithm from which it is derived; see [19, 37, 39] for a review of the properties of (PPM) in the case of monotone inclusions. The two algorithms make it possible to consider a general maximally monotone operator A acting on a Hilbert space. To apply these algorithms, it suffices to be able to calculate the resolvents of A. They provide the weak convergence of the iterates to a zero of A under the sole assumption that the set of solutions is nonempty. As a distinctive feature, (PRINAM) involves an inertial effect which is attached to the accelerated gradient method of Nesterov. The oscillations, which occur naturally with inertial methods, are significantly attenuated by the presence of the correcting term. In (PRINAM) the proximal parameter λ_k tends to infinity in a controlled way, namely



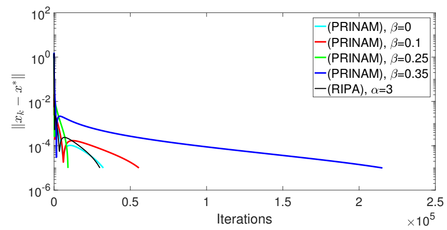
(a) $s = 0.01$, $r = 0.1$, $q = -0.1$



(b) $s = 0.1$, $r = 0.1$, $q = -0.1$



(c) $s = 0.5$, $r = 0.5$, $q = -0.5$



(d) $s = 1$, $r = 0.5$, $q = -0.5$

FIG. 2.2. *Iteration error* $\|x_k - x^*\|$ *for different instances of* (PRINAM) *and* (RIPA).

$\lambda_k = \lambda k^2$ with λ sufficiently large. This property balances the vanishing property of the damping coefficient. Indeed, the adjustment of the Yosida regularization parameter λ_k in (PRINAM) is a central question. In [14] it is proved that λ_k of order k^2 is critical for the convergence property.

The main difference between the two algorithms, and which makes (PRINAM) attractive, concerns the convergence rates. In (PRINAM), the rate of convergence of the velocities $\|x_{k+1} - x_k\|$ is of order $\mathcal{O}(1/k)$. Recall that, for (PPM), we have only the rate $\mathcal{O}(1/\sqrt{k})$; see, for example, Peypouquet and Sorin [37, Theorem 53], where one can find an extended presentation of the link between (PPM) and continuous dynamic systems. Indeed, Gu and Yang [29] showed that the convergence rate of the velocities $\mathcal{O}(1/\sqrt{k})$ is achieved by (PPM) in the case of the skew-symmetric rotation operator. In addition, for (PRINAM), $\|A_{\lambda_k}(x_k)\| = o(1/k^2)$ leads to $\|x_k - J_{\lambda_k A}(x_k)\| \to 0$ as $k \to +\infty$. This means that, since $x_k \rightharpoonup x^* \in A^{-1}(0)$ as $k \to +\infty$, we have $J_{\lambda_k A}(x_k) \rightharpoonup x^*$ as $k \to +\infty$. Most importantly, as we will see in the next section, when the operator $A = \partial f$ is the subdifferential of a lower semicontinuous, convex, and proper function f, (PRINAM) improves the accelerated gradient method of Nesterov by giving the convergence rate $o\left(\frac{1}{k^2}\right)$ of the values, and the rapid convergence of the gradients toward zero. In addition, $\|x_{k+1} - x_k\|$ is of order $o(1/k)$. Moreover, in this case $\|x_k - \mathrm{prox}_{\lambda_k f}(x_k)\| = o(\sqrt{\lambda_k}/k)$ for $\lambda_k = \lambda k^t$, $t > 1$.

2.4.3. Implicit versus explicit methods. Using implicit (backward) steps results in proximal algorithms that can be applied to general maximally monotone operators. The use of forward steps for monotone Lipschitz operators that are not cocoercive requires some caution. In this case, one can use Tseng's forward-backward-forward splitting algorithm, or some of its variants recently discovered by Malitsky and Tam [34] and Böhm et al. [21]. Continuous dynamics related to these schemes have been investigated by Csetnek, Malitsky, and Tam in [28]. Of course, in the composite case, it is natural to combine the two types of techniques, which is a subject largely to explore in the case of inertial methods; see Attouch and Cabot [6] for some first results. The machinery around proximal calculus has grown considerably in recent years, which makes the use of proximal algorithms (even for composite problems) more and more attractive; see Combettes and Glaudin [27].

3. The convex case. Let us specialize the previous results to the case of convex minimization and show the rapid convergence of values. Given a lower semi-continuous convex and proper function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ such that $\operatorname{argmin} f \neq \emptyset$, we consider the minimization problem

$$(\mathcal{P}) \quad \inf_{x \in \mathcal{H}} f(x).$$

Fermat's rule states that x is a global minimum of f if and only if $0 \in \partial f(x)$. Therefore, $(\mathcal{P})$ is equivalent to a monotone inclusion problem, and $\operatorname{argmin} f = (\partial f)^{-1}(0)$. The Yosida approximation of ∂f is equal to the gradient of the Moreau envelope of f: for any $\lambda > 0$

$$(\partial f)_\lambda = \nabla f_\lambda. \tag{3.1}$$

Recall that $f_\lambda : \mathcal{H} \to \mathbb{R}$ is a $\mathcal{C}^{1,1}$ function, which is defined by, for any $x \in \mathcal{H}$,

$$f_\lambda(x) = \inf_{\xi \in \mathcal{H}} \left\{ f(\xi) + \frac{1}{2\lambda} \|x - \xi\|^2 \right\}.$$

Let us specialize (PRINAM) in the case $A = \partial f$. According to (3.1), the formulation (2.7) gives the following:

(PRINAM)-convex

Take $x_1 \in \mathcal{H}$, $x_2 \in \mathcal{H}$, $s > 0$, $\beta \geq 0$.

$$\text{Step}\, k : \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) - \beta(\nabla f_{\lambda_k}(x_k) - \nabla f_{\lambda_{k-1}}(x_{k-1})), \\ \\ x_{k+1} = y_k - s\nabla f_{\lambda_{k+1}+s}(y_k). \end{cases}$$

3.1. Convergence results. The following result is a direct consequence of Theorem 2.1.

THEOREM 3.1. *Let (x_k), (y_k) be sequences generated by algorithm* (PRINAM)-convex. *Assume that $\alpha_k = \frac{t_k - 1}{t_{k+1}}$, $t_k = rk + q$, $r > 0, q \in \mathbb{R}$, and for all $k \geq 1$*

$$\lambda_k = \lambda k^2, \quad \text{with } \lambda > \frac{(2\beta + s)^2 r^2}{s}.$$

Then, the following properties are satisfied:

(i) *The speed tends to zero, and we have the estimates*

$$\begin{array}{ll} (\textit{pointwise}) & \|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right), \quad \|\nabla f_{\lambda_k}(x_k)\| = o\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty. \\ (\textit{summation}) & \sum_{k \geq 2} k\|x_k - x_{k-1}\|^2 < +\infty, \quad \sum_{k \geq 1} k^3 \|\nabla f_{\lambda_k}(x_k)\|^2 < +\infty. \end{array}$$

(ii) *The sequences $(x_k), (y_k)$ converge weakly, as $k \to +\infty$, to some $\hat{x} \in \operatorname{argmin} f$.*

(iii) *We have the convergence rates of the values: as $k \to +\infty$*

$$f_{\lambda_k}(x_k) - \min f = o\left(\frac{1}{k^2}\right) \quad \text{and} \quad f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f = o\left(\frac{1}{k^2}\right).$$

In addition, $\| \operatorname{prox}_{\lambda_k f}(x_k) - x_k\| \to 0$ as $k \to +\infty$.

Proof. (i) and (ii) follow directly from Theorem 2.1 applied to the operator ∂f and using (3.1).

(iii) Take $x^* \in \operatorname{argmin} f$. From the gradient inequality, and (x_k) is bounded, for all $k \geq 0$ we have

$$\begin{aligned} f_{\lambda_k}(x_k) - \min_{\mathcal{H}} f = f_{\lambda_k}(x_k) - f_{\lambda_k}(x^*) &\leq \langle \nabla f_{\lambda_k}(x_k), x_k - x^* \rangle \\ &\leq \|\nabla f_{\lambda_k}(x_k)\| \|x_k - x^*\| \leq M \|\nabla f_{\lambda_k}(x_k)\|. \end{aligned}$$

Combining the above relation with $\|\nabla f_{\lambda_k}(x_k)\| = o\left(\frac{1}{k^2}\right)$ as $k \to +\infty$ (see (i)), we obtain

$$f_{\lambda_k}(x_k) - \min_{\mathcal{H}} f = o\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty. \tag{3.2}$$

By the definition of f_{λ_k} and of the proximal mapping, we have

$$f_{\lambda_k}(x_k) - \min_{\mathcal{H}} f = f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min_{\mathcal{H}} f + \frac{1}{2\lambda_k}\|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\|^2. \tag{3.3}$$

Combining (3.2) with (3.3), we obtain

(3.4)
$$f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min_{\mathcal{H}} f = o\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty, \quad \lim_{k \to +\infty} k^2 \frac{1}{2\lambda_k}\|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\|^2 = 0.$$

The above relation leads to $\lim_{k \to +\infty} \|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\| = 0$, which completes the proof. ∎

Remark 2. When $A = \partial f$, f convex, we have additional tools, such as the gradient inequality. We will show in the following theorem that, in this case, some assumptions can be weakened. When $\beta = 0$, we will obtain fast convergence of the values for $\lambda_k = \lambda k^t$, $t \geq 0$, $\lambda > 0$, that is, under the mild assumption that the sequence (λ_k) is nondecreasing. Further, fast convergence can be obtained in the general case $\beta > 0$ provided that the sequence (λ_k) is constant or $\lambda_k = \lambda k^t$, $t > 1$, $\lambda > 0$.

THEOREM 3.2. *Let (x_k), (y_k) be sequences generated by algorithm* (PRINAM)-convex. *Assume that $\alpha_k = \frac{t_k - 1}{t_{k+1}}$ and $\lambda_k = \lambda k^t$, $\lambda > 0$, $t \geq 0$, for all $k \geq 1$, and further $t_k = rk + q$, $r \in (0, \frac{1}{2})$, $q \in \mathbb{R}$, that is, there exist $k_1 \geq 1$ and $m \in (0,1)$ such that*

$$t_k \geq 1, \; m t_{k+1} \geq t_{k+1}^2 - t_k^2 \text{ for all } k \geq k_1. \tag{3.5}$$

(i) *Suppose one of the following conditions is fulfilled:*
   (a) $\beta = 0$, $t \geq 0$.
   (b) $\beta > 0$, $t = 0$, $s > 2\beta$.
   (c) $\beta > 0$, $t > 1$ *or* $\beta > 0$, $t = 1$ *and* $\lambda > \frac{2(\beta+s)\beta r}{(1-m)s}$.

Then, the speed $(x_{k+1} - x_k)_{k \geq 1}$ tends to zero, and we have the following estimates as $k \to +\infty$:

$$\text{(pointwise)} \quad f_{\lambda_k}(x_k) - \min f = \mathcal{O}\left(\frac{1}{k^2}\right), \; f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f = \mathcal{O}\left(\frac{1}{k^2}\right),$$

$$\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right), \; \|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\| = \mathcal{O}\left(\frac{\sqrt{\lambda_k}}{k}\right),$$

$$\|\nabla f_{\lambda_k}(x_k)\| = \mathcal{O}\left(\frac{1}{k\sqrt{\lambda_k}}\right).$$

$$(summation) \quad \sum_{k\geq 2} k\|x_k - x_{k-1}\|^2 < +\infty, \quad \sum_{k\geq 1} k\lambda_k\|\nabla f_{\lambda_k}(x_k)\|^2 < +\infty,$$

$$\sum_{k\geq 1} k(f_{\lambda_k}(x_k) - \min f) < +\infty, \quad \sum_{k\geq 1} k^2\|\nabla f_{\lambda_k}(x_k)\|^2 < +\infty.$$

(ii) *For $\beta = 0$, $t \geq 0$ or $\beta > 0$ and $t > 1$ we have the following convergence rates of the values:*

$$f_{\lambda_k}(x_k) - \min f = o\left(\frac{1}{k^2}\right) \quad \textit{and} \quad f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f = o\left(\frac{1}{k^2}\right) \quad \textit{as } k \to +\infty.$$

In addition, $\|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right)$ *and* $\lim_{k\to+\infty} \frac{k}{\sqrt{\lambda_k}}\|\operatorname{prox}_{\lambda_k f}(x_k) - x_k\| = 0$.

(iii) *For $\beta = 0$, $t \in [0, 2]$ or $\beta > 0$, $t \in]1, 2]$, (x_k) and (y_k) converge weakly to some $\hat{x} \in \operatorname{argmin} f$.*

Proof. I. *The discrete energy functions.* Take $z \in \operatorname{argmin} f$. In each of the three cases (a)–(c), our Lyapunov analysis is based on a different energy function.

(a) *Case $\beta = 0$.* For each $k \geq 2$, consider the discrete energy function as in the proof of Theorem 2.1, with $a = 0$ (in accordance with (2.12)), and with $A_{\lambda_k} = \nabla f_{\lambda_k}$, that is,

$$\mathcal{E}^k_{0,b} = \frac{1}{2}\|b(x_{k-1} - z) + t_k(x_k - x_{k-1} + s\nabla f_{\lambda_k}(x_k))\|^2 + \frac{b(1-b)}{2}\|x_{k-1} - z\|^2. \tag{3.6}$$

Arguing as in the proof of Theorem 2.1, (2.18) in this particular instance becomes

$$\begin{aligned}\mathcal{E}^{k+1}_{0,b} - \mathcal{E}^k_{0,b} = &- bst_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle - \frac{1}{2}s^2t_k^2\|\nabla f_{\lambda_k}(x_k)\|^2 \\ &- st_k(t_k - b)\langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle + \frac{1}{2}(b-1)(2t_k - 1)\|x_k - x_{k-1}\|^2.\end{aligned} \tag{3.7}$$

By using successively the gradient inequality and the fact that the function $\lambda \mapsto f_\lambda$ is nonincreasing and the sequence (λ_k) is nondecreasing we get

$$\langle \nabla f_{\lambda_k}(x_k), x_{k-1} - x_k\rangle \leq f_{\lambda_k}(x_{k-1}) - f_{\lambda_k}(x_k) \leq f_{\lambda_{k-1}}(x_{k-1}) - f_{\lambda_k}(x_k).$$

Set $\epsilon := \frac{-1+\sqrt{9-8m}}{8}$ and $b := m + \epsilon$. Since $0 < m < 1$, one can easily verify that $\epsilon > 0$ and $0 < b < 1$. Since by assumption there exists $k_1 \geq 1$ such that $1 \leq t_k$ for all $k \geq k_1$, we obtain

$$\begin{aligned}&- st_k(t_k - b)\langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle = st_k(t_k - m - \epsilon)\langle \nabla f_{\lambda_k}(x_k), x_{k-1} - x_k\rangle \\ &\leq st_k(t_k - m)(f_{\lambda_{k-1}}(x_{k-1}) - f_{\lambda_k}(x_k)) + \epsilon st_k\langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \text{ for all } k > k_1.\end{aligned} \tag{3.8}$$

Moreover, according to the gradient inequality, we have that, for all $k \geq 1$,

$$\begin{aligned}-bst_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle &= -(m+\epsilon)st_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle \\ &\leq mst_k(f_{\lambda_k}(z) - f_{\lambda_k}(x_k)) - \epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle \\ &= mst_k(\min f - f_{\lambda_k}(x_k)) - \epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle.\end{aligned} \tag{3.9}$$

Now using the fact that $f_{\lambda_k}(z) - f_{\lambda_k}(x_k) = (f_{\lambda_k}(z) - \min f) - (f_{\lambda_k}(x_k) - \min f)$, and using (3.5), the last two relations give

(3.10)
$$
\begin{aligned}
&- bst_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle - st_k(t_k - b)\langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
&\quad \le st_k(t_k - m)(f_{\lambda_{k-1}}(x_{k-1}) - \min f) - st_k^2(f_{\lambda_k}(x_k) - \min f) \\
&\qquad - \epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle + \epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
&\quad \le st_{k-1}^2(f_{\lambda_{k-1}}(x_{k-1}) - \min f) - st_k^2(f_{\lambda_k}(x_k) - \min f) \\
&\qquad - \epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle + \epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \text{ for all } k \ge k_1 + 1.
\end{aligned}
$$

Combining (3.7) and (3.10), we obtain, for all $k \ge k_1 + 1$,

$$
\begin{aligned}
&\mathcal{E}_{0,b}^{k+1} - \mathcal{E}_{0,b}^{k} + st_k^2(f_{\lambda_k}(x_k) - \min f) - st_{k-1}^2(f_{\lambda_{k-1}}(x_{k-1}) - \min f) \\
&\quad \le -\epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle + \epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\
&\qquad - \frac{1}{2}s^2 t_k^2 \|\nabla f_{\lambda_k}(x_k)\|^2 - \frac{1}{2}(1-b)(2t_k - 1)\|x_k - x_{k-1}\|^2.
\end{aligned} \tag{3.11}
$$

Take $p > \frac{\epsilon s}{4(1-b)}$, and write the elementary algebraic inequality

$$
\epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \le p\epsilon st_k \|\nabla f_{\lambda_k}(x_k)\|^2 + \frac{\epsilon st_k}{4p}\|x_k - x_{k-1}\|^2.
$$

Since $t_k = rk + q, r > 0$, there exists $\epsilon_1 > 0$, $\epsilon_2 > 0$, and $k_2 \ge k_1 + 1$ such that for all $k \ge k_2$

$$
p\epsilon st_k - \frac{1}{2}s^2 t_k^2 < -\epsilon_1 t_k^2 \quad \text{and} \quad \frac{\epsilon st_k}{4p} - \frac{1}{2}(1-b)(2t_k - 1) < -\epsilon_2 t_k,
$$

where the last inequality above comes from the choice of p. Therefore,

(3.12)
$$
\begin{aligned}
&\epsilon st_k \langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle - \frac{1}{2}s^2 t_k^2 \|\nabla f_{\lambda_k}(x_k)\|^2 - \frac{1}{2}(1-b)(2t_k - 1)\|x_k - x_{k-1}\|^2 \\
&\quad \le -\epsilon_1 t_k^2 \|\nabla f_{\lambda_k}(x_k)\|^2 - \epsilon_2 t_k \|x_k - x_{k-1}\|^2 \text{ for all } k \ge k_2.
\end{aligned}
$$

According to the λ_k-cocoerciveness of ∇f_{λ_k}, $\nabla f_{\lambda_k}(z) = 0$, and the gradient inequality, we have

$$
\langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle \ge \frac{1}{2}(f_{\lambda_k}(x_k) - \min f) + \frac{\lambda_k}{2}\|\nabla f_{\lambda_k}(x_k)\|^2 \text{ for all } k \ge k_2. \tag{3.13}
$$

Consequently, (3.11) becomes, for all $k \ge k_2$,

$$
\begin{aligned}
&\mathcal{E}_{0,b}^{k+1} - \mathcal{E}_{0,b}^{k} + st_k^2(f_{\lambda_k}(x_k) - \min f) - st_{k-1}^2(f_{\lambda_{k-1}}(x_{k-1}) - \min f) \\
&\quad + \frac{\epsilon}{2}st_k(f_{\lambda_k}(x_k) - \min f) + \frac{\epsilon}{2}st_k\lambda_k \|\nabla f_{\lambda_k}(x_k)\|^2 + \epsilon_1 t_k^2 \|\nabla f_{\lambda_k}(x_k)\|^2 \\
&\quad + \epsilon_2 t_k \|x_k - x_{k-1}\|^2 \le 0.
\end{aligned} \tag{3.14}
$$

(b) *Case* $\beta > 0$ *and* $t = 0$. Then $\lambda_k = \lambda > 0$. For each $k \ge 2$, consider the discrete energy function as in the proof of Theorem 2.1, with $A_{\lambda_k} = \nabla f_\lambda$, that is,

$$
\begin{aligned}
\mathcal{E}_{a,b}^{k} &= at_{k-1} \langle \nabla f_\lambda(x_{k-1}), x_{k-1} - z\rangle + \frac{1}{2}\|b(x_{k-1} - z) + t_k(x_k - x_{k-1} + s\nabla f_\lambda(x_k))\|^2 \\
&\quad + \frac{b(1-b)}{2}\|x_{k-1} - z\|^2
\end{aligned}
$$

and for $d>0$ (which will be fixed later) set

$$W_{a,b,d}^k := \mathcal{E}_{a,b}^k + dk^2\|\nabla f_\lambda(x_{k-1})\|^2. \tag{3.15}$$

Arguing as in the proof of Theorem 2.1, (2.18) in this particular instance becomes

$$\begin{aligned}
&\mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^k \\
&\quad = (at_k - b(\beta t_{k+1} + st_k))\langle \nabla f_\lambda(x_k), x_k - z\rangle + (b\beta t_{k+1} - at_{k-1})\langle \nabla f_\lambda(x_{k-1}), x_{k-1} - z\rangle \\
&\qquad + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2)\|\nabla f_\lambda(x_k)\|^2 - \beta^2 t_{k+1}^2\langle \nabla f_\lambda(x_k), \nabla f_\lambda(x_{k-1})\rangle + \frac{1}{2}\beta^2 t_{k+1}^2\|\nabla f_\lambda(x_{k-1})\|^2 \\
&\qquad - (\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle \\
&\qquad + \beta t_{k+1}(t_k + b - 1)\langle \nabla f_\lambda(x_{k-1}), x_k - x_{k-1}\rangle + \frac{1}{2}(b-1)(2t_k - 1)\|x_k - x_{k-1}\|^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
&W_{a,b,d}^{k+1} - W_{a,b,d}^k \\
&\quad = (at_k - b(\beta t_{k+1} + st_k))\langle \nabla f_\lambda(x_k), x_k - z\rangle + (b\beta t_{k+1} - at_{k-1})\langle \nabla f_\lambda(x_{k-1}), x_{k-1} - z\rangle \\
&\qquad + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2 + 2d(k+1)^2)\|\nabla f_\lambda(x_k)\|^2 - \beta^2 t_{k+1}^2\langle \nabla f_\lambda(x_k), \nabla f_\lambda(x_{k-1})\rangle \\
&\qquad + \frac{1}{2}(\beta^2 t_{k+1}^2 - 2dk^2)\|\nabla f_\lambda(x_{k-1})\|^2 - (\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle \\
&\qquad + \beta t_{k+1}(t_k + b - 1)\langle \nabla f_\lambda(x_{k-1}), x_k - x_{k-1}\rangle + \frac{1}{2}(b-1)(2t_k - 1)\|x_k - x_{k-1}\|^2.
\end{aligned} \tag{3.16}$$

According to the monotonicity of ∇f_λ and the assumption $t_k + b - 1 > 0$ for all $k > k_1$, we have

$$\begin{aligned}
&-(\beta t_{k+1}(t_k - 1) + st_k(t_k - b))\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle + \beta t_{k+1}(t_k + b - 1)\langle \nabla f_\lambda(x_{k-1}), x_k - x_{k-1}\rangle \\
&\quad \le (\beta b t_{k+1} - st_k(t_k - b)))\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle \\
&\quad = (\beta b t_{k+1} + sbt_k))\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle + st_k^2\langle \nabla f_\lambda(x_k), x_{k-1} - x_k\rangle.
\end{aligned} \tag{3.17}$$

According to the gradient inequality we have, for all $k > k_1$,

$$\begin{aligned}
&st_k^2\langle \nabla f_\lambda(x_k), x_{k-1} - x_k\rangle \le st_k^2((f_\lambda(x_{k-1}) - \min f) - (f_\lambda(x_k) - \min f)) \\
&\quad = st_{k-1}^2(f_\lambda(x_{k-1}) - \min f) - st_k^2(f_\lambda(x_k) - \min f) + s(t_k^2 - t_{k-1}^2)(f_\lambda(x_{k-1}) - \min f).
\end{aligned} \tag{3.18}$$

Combining (3.17) and (3.18) with (3.16), we obtain

$$\begin{aligned}
&W_{a,b,d}^{k+1} - W_{a,b,d}^k + st_k^2(f_\lambda(x_k) - \min f) - st_{k-1}^2(f_\lambda(x_{k-1}) - \min f) \\
&\le (at_k - b(\beta t_{k+1} + st_k))\langle \nabla f_\lambda(x_k), x_k - z\rangle + (b\beta t_{k+1} - at_{k-1})\langle \nabla f_\lambda(x_{k-1}), x_{k-1} - z\rangle \\
&\quad + \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2 + 2d(k+1)^2)\|\nabla f_\lambda(x_k)\|^2 - \beta^2 t_{k+1}^2\langle \nabla f_\lambda(x_k), \nabla f_\lambda(x_{k-1})\rangle \\
&\quad + \frac{1}{2}(\beta^2 t_{k+1}^2 - 2dk^2)\|\nabla f_\lambda(x_{k-1})\|^2 + (\beta b t_{k+1} + sbt_k))\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle \\
&\quad + \frac{1}{2}(b-1)(2t_k - 1)\|x_k - x_{k-1}\|^2 + s(t_k^2 - t_{k-1}^2)(f_\lambda(x_{k-1}) - \min f) \text{ for all } k > k_1.
\end{aligned} \tag{3.19}$$

Take $\beta b < a < b(\beta + s)$ and $0 < b < 1$. Since $r < \frac{1}{2}$, we can choose a and b satisfying the previous inequalities and such that there exists $k_2 > k_1$ and $\epsilon_3 > 0$ such that

$$at_k - b(\beta t_{k+1} + st_k) + \epsilon_3 k \leq 0 \text{ and } b\beta t_{k+1} - at_{k-1} + s(t_k^2 - t_{k-1}^2) \leq 0$$

for all $k \geq k_2$ (take $a = b\beta + (1-\epsilon)bs$ with ϵ sufficiently small, so that $1 > b > \frac{2r}{1-\epsilon}$). By using the gradient inequality

$$\begin{aligned}
&\epsilon_3 k(f_\lambda(x_k) - \min f) \leq \epsilon_3 k\langle \nabla f_\lambda(x_k), x_k - z\rangle,\\
&(b\beta t_{k+1} - at_{k-1})\langle \nabla f_\lambda(x_{k-1}), x_{k-1} - z\rangle\\
&\quad \leq (b\beta t_{k+1} - at_{k-1})(f_\lambda(x_{k-1}) - \min f) \text{ for all } k \geq k_2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
(3.20)\qquad &(at_k - b(\beta t_{k+1} + st_k) + \epsilon_3 k)\langle \nabla f_\lambda(x_k), x_k - z\rangle\\
&\quad + (b\beta t_{k+1} - at_{k-1})\langle \nabla f_\lambda(x_{k-1}), x_{k-1} - z\rangle\\
&\quad + s(t_k^2 - t_{k-1}^2)(f_\lambda(x_{k-1}) - \min f) \leq 0 \text{ for all } k \geq k_2.
\end{aligned}$$

Combining (3.19) with (3.20), we get, for all $k \geq k_2$,

(3.21)

$$\begin{aligned}
&W_{a,b,d}^{k+1} - W_{a,b,d}^{k} + st_k^2(f_\lambda(x_k) - \min f) - st_{k-1}^2(f_\lambda(x_{k-1}) - \min f) + \epsilon_3 k(f_\lambda(x_k) - \min f)\\
&\quad \leq \frac{1}{2}(\beta^2 t_{k+1}^2 - s^2 t_k^2 + 2d(k+1)^2)\|\nabla f_\lambda(x_k)\|^2 - \beta^2 t_{k+1}^2\langle \nabla f_\lambda(x_k), \nabla f_\lambda(x_{k-1})\rangle\\
&\qquad + \frac{1}{2}(\beta^2 t_{k+1}^2 - 2dk^2)\|\nabla f_\lambda(x_{k-1})\|^2 + (\beta bt_{k+1} + sbt_k))\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle\\
&\qquad + \frac{1}{2}(b-1)(2t_k - 1)\|x_k - x_{k-1}\|^2.
\end{aligned}$$

We now use the following elementary algebraic inequalities:

$$\begin{aligned}
&-\beta^2 t_{k+1}^2\langle \nabla f_\lambda(x_k), \nabla f_\lambda(x_{k-1})\rangle \leq \frac{\beta^2 t_{k+1}^2}{2}\left(\|\nabla f_\lambda(x_k)\|^2 + \|\nabla f_\lambda(x_{k-1})\|^2\right)\\
&\quad (\beta bt_{k+1} + sbt_k)\langle \nabla f_\lambda(x_k), x_k - x_{k-1}\rangle\\
&\quad \leq (\beta bt_{k+1} + sbt_k)\left(\frac{\sqrt{k}}{2}\|\nabla f_\lambda(x_k)\|^2 + \frac{1}{2\sqrt{k}}\|x_k - x_{k-1}\|^2\right).
\end{aligned}$$

Taking into account that $s > 2\beta$, we choose d such that $\beta^2 r^2 < d < -\beta^2 r^2 + \frac{s^2}{2}r^2$. Then, there exists $k_3 \geq k_2$ and $\epsilon_4, \epsilon_5 > 0$ such that, for all $k \geq k_3$,

$$\begin{aligned}
&\frac{1}{2}\left(2\beta^2 t_{k+1}^2 - s^2 t_k^2 + 2d(k+1)^2 + (\beta bt_{k+1} + sbt_k)\sqrt{k}\right) + \epsilon_4 k^2 \leq 0;\\
&\beta^2 t_{k+1}^2 - dk^2 \leq 0 \text{ and } \frac{1}{2}\left((b-1)(2t_k - 1) + \frac{\beta bt_{k+1} + sbt_k}{\sqrt{k}}\right) + \epsilon_5 k \leq 0.
\end{aligned}$$

Consequently, (3.21) becomes, for all $k \geq k_3$,

(3.22)
$$W_{a,b,d}^{k+1}-W_{a,b,d}^{k}+st_k^2(f_\lambda(x_k)-\min f)-st_{k-1}^2(f_\lambda(x_{k-1})-\min f)+\epsilon_3 k(f_\lambda(x_k)-\min f)$$
$$+\epsilon_4 k^2\|\nabla f_\lambda(x_k)\|^2+\epsilon_5 k\|x_k-x_{k-1}\|^2$$
$$\leq \frac{1}{2}\left(2\beta^2 t_{k+1}^2-s^2t_k^2+2d(k+1)^2+(\beta b t_{k+1}+sbt_k)\sqrt{k}+2\epsilon_4 k^2\right)\|\nabla f_\lambda(x_k)\|^2$$
$$+(\beta^2 t_{k+1}^2-dk^2)\|\nabla f_\lambda(x_{k-1})\|^2+\frac{1}{2}\left((b-1)(2t_k-1)\right.$$
$$\left.+\frac{\beta b t_{k+1}+sbt_k}{\sqrt{k}}+2\epsilon_5 k\right)\|x_k-x_{k-1}\|^2\leq 0.$$

(c) *Case* $\beta>0$ *and* $t\geq 1$. For each $k\geq 2$, consider the discrete energy function as in the proof of Theorem 2.1, with $A_{\lambda_k}=\nabla f_{\lambda_k}$, that is,

$$\mathcal{E}_{a,b}^k=at_{k-1}\left\langle \nabla f_{\lambda_{k-1}}(x_{k-1}),x_{k-1}-z\right\rangle+\frac{1}{2}\|b(x_{k-1}-z)$$
$$+t_k(x_k-x_{k-1}+s\nabla f_{\lambda_k}(x_k))\|^2+\frac{b(1-b)}{2}\|x_{k-1}-z\|^2.$$

Arguing as in the proof of Theorem 2.1, (2.18) in this particular instance becomes

$$\begin{aligned}\mathcal{E}_{a,b}^{k+1}-\mathcal{E}_{a,b}^k=&\,(at_k-b(\beta t_{k+1}+st_k))\left\langle\nabla f_{\lambda_k}(x_k),x_k-z\right\rangle\\&+(b\beta t_{k+1}-at_{k-1})\left\langle\nabla f_{\lambda_{k-1}}(x_{k-1}),x_{k-1}-z\right\rangle\\&+\frac{1}{2}(\beta^2t_{k+1}^2-s^2t_k^2)\|\nabla f_{\lambda_k}(x_k)\|^2-\beta^2t_{k+1}^2\langle\nabla f_{\lambda_k}(x_k),\nabla f_{\lambda_{k-1}}(x_{k-1})\rangle\\&+\frac{1}{2}\beta^2t_{k+1}^2\|\nabla f_{\lambda_{k-1}}(x_{k-1})\|^2\\&-(\beta t_{k+1}(t_k-1)+st_k(t_k-b))\langle\nabla f_{\lambda_k}(x_k),x_k-x_{k-1}\rangle\\&+\beta t_{k+1}(t_k+b-1)\langle\nabla f_{\lambda_{k-1}}(x_{k-1}),x_k-x_{k-1}\rangle\\&+\frac{1}{2}(b-1)(2t_k-1)\|x_k-x_{k-1}\|^2\text{ for all }k\geq 2.\end{aligned}$$

From Lemma A.2, for all $k\geq 2$ we have

(3.23)
$$\|\lambda_k\nabla f_{\lambda_k}(x_k)-\lambda_{k-1}\nabla f_{\lambda_{k-1}}(x_{k-1})\|\leq 2\|x_k-x_{k-1}\|+|\lambda_k-\lambda_{k-1}|\|\nabla f_{\lambda_k}(x_k)\|.$$

Set $F_k:=\langle\nabla f_{\lambda_{k-1}}(x_{k-1}),x_k-x_{k-1}\rangle$. Hence, we have for all $k\geq 2$ that

(3.24)
$$\begin{aligned}F_k=&\,\frac{1}{\lambda_{k-1}}\langle\lambda_{k-1}\nabla f_{\lambda_{k-1}}(x_{k-1})-\lambda_k\nabla f_{\lambda_k}(x_k),x_k-x_{k-1}\rangle\\&+\frac{\lambda_k}{\lambda_{k-1}}\langle\nabla f_{\lambda_k}(x_k),x_k-x_{k-1}\rangle\\\leq&\,\frac{1}{\lambda_{k-1}}\|\lambda_{k-1}\nabla f_{\lambda_{k-1}}(x_{k-1})-\lambda_k\nabla f_{\lambda_k}(x_k)\|\|x_k-x_{k-1}\|+\frac{\lambda_k}{\lambda_{k-1}}\langle\nabla f_{\lambda_k}(x_k),x_k-x_{k-1}\rangle\\\leq&\,\frac{2}{\lambda_{k-1}}\|x_k-x_{k-1}\|^2+\frac{|\lambda_k-\lambda_{k-1}|}{\lambda_{k-1}}\|\nabla f_{\lambda_k}(x_k)\|\|x_k-x_{k-1}\|\\&+\frac{\lambda_k}{\lambda_{k-1}}\langle\nabla f_{\lambda_k}(x_k),x_k-x_{k-1}\rangle.\end{aligned}$$

Moreover, for every $p_1 > 0$ we have

$$\|\nabla f_{\lambda_k}(x_k)\|\|x_k - x_{k-1}\| \leq p_1\sqrt{k}\|\nabla f_{\lambda_k}(x_k)\|^2 + \frac{1}{4p_1\sqrt{k}}\|x_k - x_{k-1}\|^2.$$

Therefore, (3.24) becomes, for all $k \geq 2$,

$$\begin{aligned}\langle \nabla f_{\lambda_{k-1}}(x_{k-1}), x_k - x_{k-1}\rangle &\leq \frac{\lambda_k}{\lambda_{k-1}}\langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\ &+ \left(\frac{2}{\lambda_{k-1}} + \frac{|\lambda_k - \lambda_{k-1}|}{4p_1\sqrt{k}\lambda_{k-1}}\right)\|x_k - x_{k-1}\|^2 + \frac{p_1\sqrt{k}|\lambda_k - \lambda_{k-1}|}{\lambda_{k-1}}\|\nabla f_{\lambda_k}(x_k)\|^2.\end{aligned} \tag{3.25}$$

Combining the above results, we obtain (we write shortly $\Delta_k = \mathcal{E}_{a,b}^{k+1} - \mathcal{E}_{a,b}^{k}$), for all $k \geq 2$,

(3.26)

$$\begin{aligned}\Delta_k \leq{}& (at_k - b(\beta t_{k+1} + st_k))\left\langle \nabla f_{\lambda_k}(x_k), x_k - z\right\rangle + (b\beta t_{k+1} - at_{k-1}) \\ &\times \left\langle \nabla f_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\right\rangle \\ &+ \frac{1}{2}\left(\beta^2 t_{k+1}^2 - s^2 t_k^2 + 2\beta t_{k+1}(t_k + b - 1)\frac{p_1\sqrt{k}|\lambda_k - \lambda_{k-1}|}{\lambda_{k-1}}\right)\|\nabla f_{\lambda_k}(x_k)\|^2 \\ &- \beta^2 t_{k+1}^2\langle \nabla f_{\lambda_k}(x_k), \nabla f_{\lambda_{k-1}}(x_{k-1})\rangle + \frac{1}{2}\beta^2 t_{k+1}^2\|\nabla f_{\lambda_{k-1}}(x_{k-1})\|^2 \\ &+ \left(\beta t_{k+1}(t_k - 1)\frac{\lambda_k - \lambda_{k-1}}{\lambda_{k-1}} + \beta b t_{k+1}\frac{\lambda_k}{\lambda_{k-1}} - st_k(t_k - b)\right)\langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\ &+ \left(\beta t_{k+1}(t_k + b - 1)\left(\frac{2}{\lambda_{k-1}} + \frac{|\lambda_k - \lambda_{k-1}|}{4p_1\sqrt{k}\lambda_{k-1}}\right) + \frac{1}{2}(b-1)(2t_k - 1)\right)\|x_k - x_{k-1}\|^2.\end{aligned}$$

Further estimates give

(3.27)

$$-\beta^2 t_{k+1}^2\langle \nabla f_{\lambda_k}(x_k), \nabla f_{\lambda_{k-1}}(x_{k-1})\rangle \leq \frac{\beta^2 t_{k+1}^2}{2}\left(\|\nabla f_{\lambda_k}(x_k)\|^2 + \|\nabla f_{\lambda_{k-1}}(x_{k-1})\|^2\right);$$

(3.28)

$$\begin{aligned}&\left(\beta t_{k+1}(t_k - 1)\frac{\lambda_k - \lambda_{k-1}}{\lambda_{k-1}} + \beta b t_{k+1}\frac{\lambda_k}{\lambda_{k-1}} + sbt_k\right)\langle \nabla f_{\lambda_k}(x_k), x_k - x_{k-1}\rangle \\ &\leq \left(\beta t_{k+1}(t_k - 1)\frac{\lambda_k - \lambda_{k-1}}{\lambda_{k-1}} + \beta b t_{k+1}\frac{\lambda_k}{\lambda_{k-1}} + sbt_k\right) \\ &\quad \times \left(\sqrt{k}\|\nabla f_{\lambda_k}\|^2 + \frac{1}{4\sqrt{k}}\|x_k - x_{k-1}\|^2\right).\end{aligned}$$

To simplify the formulation of the formulas, let us denote

$$\begin{aligned}r_1(k) &= -\frac{1}{2}(b-1) + \frac{2\beta(b-1)t_{k+1}}{\lambda_{k-1}} + \beta t_{k+1}(t_k + b - 1)\frac{|\lambda_k - \lambda_{k-1}|}{4p_1\sqrt{k}\lambda_{k-1}}, \\ r_2(k) &= \left(\beta t_{k+1}(t_k - 1)\frac{\lambda_k - \lambda_{k-1}}{\lambda_{k-1}} + \beta b t_{k+1}\frac{\lambda_k}{\lambda_{k-1}} + sbt_k\right)\frac{1}{4\sqrt{k}} \\ r_3(k) &= \left(\beta t_{k+1}(t_k - 1)\frac{\lambda_k - \lambda_{k-1}}{\lambda_{k-1}} + \beta b t_{k+1}\frac{\lambda_k}{\lambda_{k-1}} + sbt_k\right)\sqrt{k} \\ &\quad + \beta t_{k+1}(t_k + b - 1)\frac{p_1\sqrt{k}|\lambda_k - \lambda_{k-1}|}{\lambda_{k-1}}.\end{aligned}$$

Since $\lim_{x\to+\infty}\frac{x^t-(x-1)^t}{(x-1)^{t-1}}=t$, we have $\frac{\lambda_k-\lambda_{k-1}}{\lambda_{k-1}}=\mathcal{O}(\frac{1}{k})$ as $k\to+\infty$. Hence,

$$r_1(k)=\mathcal{O}\left(k^{\frac{1}{2}}\right),\ r_2(k)=\mathcal{O}\left(k^{\frac{1}{2}}\right),\ r_3(k)=\mathcal{O}\left(k^{\frac{3}{2}}\right),\ k\to+\infty. \tag{3.29}$$

Consequently, (3.26), (3.27), and (3.28) yield, for all $k\geq 2$,

$$\begin{aligned}\Delta_k\leq{}&(at_k-b(\beta t_{k+1}+st_k))\left\langle\nabla f_{\lambda_k}(x_k),x_k-z\right\rangle+(b\beta t_{k+1}-at_{k-1})\\&\times\left\langle\nabla f_{\lambda_{k-1}}(x_{k-1}),x_{k-1}-z\right\rangle\\&+\frac{1}{2}(2\beta^2t_{k+1}^2-s^2t_k^2+2r_3(k))\|\nabla f_{\lambda_k}(x_k)\|^2+\beta^2t_{k+1}^2\|\nabla f_{\lambda_{k-1}}(x_{k-1})\|^2\\&-st_k^2\langle\nabla f_{\lambda_k}(x_k),x_k-x_{k-1}\rangle+\left((b-1)t_k+\frac{2\beta t_{k+1}t_k}{\lambda_{k-1}}+r_1(k)+r_2(k)\right)\|x_k-x_{k-1}\|^2.\end{aligned} \tag{3.30}$$

According to the gradient inequality and the fact that the function $\lambda\to f_\lambda$ is non-increasing and (λ_k) is non-decreasing, we have, for all $k\geq 2$

$$\begin{aligned}&st_k^2\langle\nabla f_{\lambda_k}(x_k),x_{k-1}-x_k\rangle\leq st_k^2((f_{\lambda_{k-1}}(x_{k-1})-\min f)-(f_{\lambda_k}(x_k)-\min f))\\&\quad=st_{k-1}^2(f_{\lambda_{k-1}}(x_{k-1})-\min f)-st_k^2(f_{\lambda_k}(x_k)-\min f)+s(t_k^2-t_{k-1}^2)(f_{\lambda_{k-1}}(x_{k-1})-\min f).\end{aligned} \tag{3.31}$$

Now, (3.30) becomes, for all $k\geq 2$

$$\begin{aligned}&\Delta_k+st_k^2(f_{\lambda_k}(x_k)-\min f)-st_{k-1}^2(f_{\lambda_{k-1}}(x_{k-1})-\min f)\\&\quad\leq(at_k-b(\beta t_{k+1}+st_k))\langle\nabla f_{\lambda_k}(x_k),x_k-z\rangle+(b\beta t_{k+1}-at_{k-1})\left\langle\nabla f_{\lambda_{k-1}}(x_{k-1}),x_{k-1}-z\right\rangle\\&\qquad+\frac{1}{2}(2\beta^2t_{k+1}^2-s^2t_k^2+2r_3(k))\|\nabla f_{\lambda_k}(x_k)\|^2+\beta^2t_{k+1}^2\|\nabla f_{\lambda_{k-1}}(x_{k-1})\|^2\\&\qquad+s(t_k^2-t_{k-1}^2)(f_{\lambda_{k-1}}(x_{k-1})-\min f)+\left((b-1)t_k+\frac{2\beta t_{k+1}t_k}{\lambda_{k-1}}+r_1(k)+r_2(k)\right)\|x_k-x_{k-1}\|^2.\end{aligned} \tag{3.32}$$

Further, by the gradient inequality and the fact that $t_k^2-t_{k-1}^2\leq mt_k$ for all $k\geq k_1+1$ we have

$$\begin{aligned}&(b\beta t_{k+1}-at_{k-1})\left\langle\nabla f_{\lambda_{k-1}}(x_{k-1}),x_{k-1}-z\right\rangle+s(t_k^2-t_{k-1}^2)(f_{\lambda_{k-1}}(x_{k-1})-\min f)\\&\qquad\leq(b\beta t_{k+1}-at_{k-1}+mst_k)\langle\nabla f_{\lambda_{k-1}}(x_{k-1}),x_{k-1}-z\rangle\text{ for all }k\geq k_1+1.\end{aligned}$$

Assume that $t=1$ and let $b=\frac{\beta+ms}{\beta+s}\in(m,1)$ and $a=\beta b+\frac{(m+b)s}{2}\in(\beta b+ms,\beta b+bs)$. Since by assumption we have $\lambda>\frac{2(\beta+s)\beta r}{(1-m)s}$ we conclude that there exist $\epsilon_6,\epsilon_7>0$ such that

$$\begin{aligned}&((a-\beta b-bs)r+\epsilon_6)\lambda+\frac{1}{2}(2\beta^2r^2-s^2r^2)<0,\\&(b-1)r+\frac{2\beta r^2}{\lambda}+\epsilon_7<0\ \text{ and }\ (\beta br-ar+msr)\lambda+\beta^2r^2<0.\end{aligned} \tag{3.33}$$

Assume that $t>1$ and fix $b\in(m,1)$, $a\in(\beta b+ms,\beta b+bs)$. Then take $\epsilon_6,\epsilon_7>0$ such that

$$(a-\beta b-bs)r+\epsilon_6<0\ \text{ and }\ (b-1)r+\epsilon_7<0. \tag{3.34}$$

From now on, we do not need to distinguish the cases $t = 1$ and $t > 1$. According to the λ_k cocoerciveness of ∇f_{λ_k}, (3.33), (3.34), (3.29), and $t \geq 1$, we deduce that

$$(at_k - b(\beta t_{k+1}+st_k)+\epsilon_6 k)\,\langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle + \frac{1}{2}(2\beta^2 t_{k+1}^2 - s^2 t_k^2 + 2r_3(k))\|\nabla f_{\lambda_k}(x_k)\|^2 \leq 0,$$
$$(b\beta t_{k+1} - at_{k-1} + mst_k)\langle \nabla f_{\lambda_{k-1}}(x_{k-1}), x_{k-1} - z\rangle + \beta^2 t_{k+1}^2 \|\nabla f_{\lambda_{k-1}}(x_{k-1})\|^2 \leq 0,$$
$$\left((b-1)t_k + \frac{2\beta t_{k+1} t_k}{\lambda_{k-1}} + r_1(k) + r_2(k) + \epsilon_7 k\right)\|x_k - x_{k-1}\|^2 \leq 0$$

holds for some for $k_2 \geq k_1 + 1$ and all $k \geq k_2$. Consequently, (3.32) leads to

$$\begin{aligned}\Delta_k + st_k^2(f_\lambda(x_k) - \min f) - st_{k-1}^2(f_\lambda(x_{k-1}) - \min f)\\ \leq -\epsilon_6 k\,\langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle - \epsilon_7 k\|x_k - x_{k-1}\|^2 \text{ for all } k \geq k_2.\end{aligned} \tag{3.35}$$

Then, we use (3.13) to get $-\epsilon_6 k\langle \nabla f_{\lambda_k}(x_k), x_k - z\rangle \leq -\frac{\epsilon_6}{2}k\lambda_k\|\nabla f_{\lambda_k}(x_k)\|^2 + \frac{\epsilon_6}{2}k(\min f - f_{\lambda_k}(x_k))$, and to finally obtain

$$\begin{aligned}&\Delta_k + st_k^2(f_\lambda(x_k) - \min f) - st_{k-1}^2(f_\lambda(x_{k-1}) - \min f)\\ &+ \frac{\epsilon_6}{2}k\lambda_k\|\nabla f_{\lambda_k}(x_k)\|^2 + \frac{\epsilon_6}{2}k(f_{\lambda_k}(x_k) - \min f) + \epsilon_7 k\|x_k - x_{k-1}\|^2 \leq 0, \text{ for all } k \geq k_2.\end{aligned} \tag{3.36}$$

II. Estimates According to (3.14), (3.22), (3.36) the sequences of nonnegative numbers

$$(\mathcal{E}_{0,b}^k + st_k^2(f_{\lambda_k}(x_k) - \min f))_{k\geq 2},\ (W_{a,b,d}^k + st_k^2(f_{\lambda_k}(x_k) - \min f))_{k\geq 2},$$
$$\text{and } (\mathcal{E}_{a,b}^k + st_k^2(f_{\lambda_k}(x_k) - \min f))_{k\geq 2}$$

that correspond to the cases (a), (b), and (c) are nonincreasing after an index $k_2 \geq 2$ and therefore converge. In particular, they are bounded. From this, and by adding the inequalities in (3.14), (3.22), and (3.36), we obtain

$$\sup_k k^2(f_{\lambda_k}(x_k) - \min f) < +\infty, \tag{3.37}$$

$$\sup_k \|b(x_k - z) + t_{k+1}(x_{k+1} - x_k + s\nabla f_{\lambda_{k+1}}(x_{k+1}))\|^2 < +\infty, \tag{3.38}$$

$$\sup_k \|x_k - z\|^2 < +\infty, \tag{3.39}$$

$$\sum_{k=1}^{+\infty} k(f_{\lambda_k}(x_k) - \min f) < +\infty, \tag{3.40}$$

$$\sum_{k=1}^{+\infty} k\lambda_k\|\nabla f_{\lambda_k}(x_k)\|^2 < +\infty, \tag{3.41}$$

$$\sum_{k=1}^{+\infty} k^2\|\nabla f_{\lambda_k}(x_k)\|^2 < +\infty, \tag{3.42}$$

$$\sum_{k=2}^{+\infty} k\|x_k - x_{k-1}\|^2 < +\infty. \tag{3.43}$$

Obviously, (3.37) guarantees that

$$f_{\lambda_k}(x_k) - \min f = \mathcal{O}\left(\frac{1}{k^2}\right) \text{ as } k \to +\infty. \tag{3.44}$$

Since the general term of a convergent series goes to zero, we deduce from (3.42) that

$$\|\nabla f_{\lambda_k}(x_k)\| = o\left(\frac{1}{k}\right) \text{ as } k \to +\infty. \tag{3.45}$$

The same argument applied to (3.41) yields

$$\|\nabla f_{\lambda_k}(x_k)\| = o\left(\frac{1}{\sqrt{k\lambda_k}}\right) \text{ as } k \to +\infty. \tag{3.46}$$

Further, (3.39) shows that $\|x_k - z\|$ is bounded. Consequently, the sequence (x_k) is bounded. Combining the above results with (3.38), we obtain

$$\|x_k - x_{k-1}\| = \mathcal{O}\left(\frac{1}{k}\right) \text{ as } k \to +\infty. \tag{3.47}$$

From $f_{\lambda_k}(x_k) - \min f = f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f + \frac{1}{2\lambda_k}\|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\|^2$, we deduce that

(3.48)

$$f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f = \mathcal{O}\left(\frac{1}{k^2}\right), \quad \|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\| = \mathcal{O}\left(\frac{\sqrt{\lambda_k}}{k}\right) \text{ as } k \to +\infty.$$

Further we have $\nabla f_{\lambda_k} = (\partial f)_{\lambda_k} = \frac{1}{\lambda_k}(I - \operatorname{prox}_{\lambda_k f})$, hence

$$\|\nabla f_{\lambda_k}(x_k)\| = \mathcal{O}\left(\frac{1}{k\sqrt{\lambda_k}}\right) \text{ as } k \to +\infty. \tag{3.49}$$

III. The limit. Using the Opial lemma, let us prove that the sequence (x_k) generated by the algorithm (PRINAM)-convex converges weakly toward an element of $\operatorname{argmin} f$. Take $z \in \operatorname{argmin} f$, and consider the anchor sequence (h_k) defined by $h_k = \frac{1}{2}\|x_k - z\|^2$ for $k \geq 1$.

Arguing as in the proof of Theorem 2.1 and rewriting (2.40) in our case, we get for every $k \geq 2$

(3.50)

$$\begin{aligned} h_{k+1} - h_k \leq{}& \alpha_k(h_k - h_{k-1}) + \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 \\ &- \beta\langle \nabla f_{\lambda_k}(x_k) - \nabla f_{\lambda_{k-1}}(x_{k-1}), x_k - z\rangle + s\|\nabla f_{\lambda_{k+1}}(x_{k+1})\|\|x_{k+1} - x_k\|. \end{aligned}$$

Since $\|x_k - z\|$ is bounded, we get the existence of a constant $M > 0$ such that, for all $k \geq 2$,

(3.51)

$$\begin{aligned} h_{k+1} - h_k \leq{}& \alpha_k(h_k - h_{k-1}) + \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 \\ &+ M\beta\|\nabla f_{\lambda_k}(x_k) - \nabla f_{\lambda_{k-1}}(x_{k-1})\| + s\|\nabla f_{\lambda_{k+1}}(x_{k+1})\|\|x_{k+1} - x_k\|. \end{aligned}$$

Further, when $\beta > 0$, from Lemma A.2, we get

$$\begin{aligned} &\|\nabla f_{\lambda_k}(x_k) - \nabla f_{\lambda_{k-1}}(x_{k-1})\| \\ &\quad \leq \frac{2}{\lambda_k}\|x_k - x_{k-1}\| + \frac{|\lambda_k - \lambda_{k-1}|}{\lambda_k}\left(\|\nabla f_{\lambda_k}(x_k)\| + \|\nabla f_{\lambda_{k-1}}(x_{k-1})\|\right). \end{aligned}$$

Recall that $\lambda_k = \lambda k^t$, $t > 1$, and $\frac{\lambda_k - \lambda_{k-1}}{\lambda_k} = \mathcal{O}\left(\frac{1}{k}\right)$ as $k \to +\infty$. Then, from (3.47) and (3.49) we obtain that there exists $\bar{k} \geq 2$ and $C > 0$ such that, for all $k \geq \bar{k}$,

$$\frac{2}{\lambda_k}\|x_k - x_{k-1}\| \leq \frac{C}{k^{1+t}} \quad \text{and} \quad \frac{|\lambda_k - \lambda_{k-1}|}{\lambda_k}\left(\|\nabla f_{\lambda_k}(x_k)\| + \|\nabla f_{\lambda_{k-1}}(x_{k-1})\|\right) \leq \frac{C}{k^{2+\frac{t}{2}}}.$$

Therefore, for all $k \geq \bar{k}$,

$$\|\nabla f_{\lambda_k}(x_k) - \nabla f_{\lambda_{k-1}}(x_{k-1})\| \leq \frac{C}{k^{1+t}} + \frac{C}{k^{2+\frac{t}{2}}}. \tag{3.52}$$

Consequently, (3.51) leads to

$$h_{k+1} - h_k \leq \alpha_k(h_k - h_{k-1}) + \omega_k \tag{3.53}$$

for all $k \geq \bar{k}$, where

$$\omega_k = \begin{cases} \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 + s\|\nabla f_{\lambda_{k+1}}(x_{k+1})\|\|x_{k+1} - x_k\| \text{ if } \beta = 0, \\[2ex] \frac{1}{2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k}{2}\|x_k - x_{k-1}\|^2 + \beta MC\left(\frac{1}{k^{1+t}} + \frac{1}{k^{2+\frac{t}{2}}}\right) \\ \quad + s\|\nabla f_{\lambda_{k+1}}(x_{k+1})\|\|x_{k+1} - x_k\| \\[2ex] \text{if } \beta > 0,\ t > 1. \end{cases}$$

As a direct consequence of the majorization

$$\|\nabla f_{\lambda_{k+1}}(x_{k+1})\|\|x_{k+1} - x_k\| \leq \frac{1}{2}\|\nabla f_{\lambda_{k+1}}(x_{k+1})\|^2 + \frac{1}{2}\|x_{k+1} - x_k\|^2,$$

of (3.43) and (3.42), and of the fact that $t > 1$ if $\beta > 0$, we have $\sum_{k\geq\bar{k}} t_{k+1}\omega_k < +\infty$. Therefore, by applying Lemma A.1 to the sequence $a_k = [h_k - h_{k-1}]_+$ we obtain $\sum_k [h_k - h_{k-1}]_+ < +\infty$. Since h_k is nonnegative, this property classically gives the existence of $\lim_{k\to+\infty} h_k$ and hence of the existence of $\lim_{k\to+\infty}\|x_k - z\|$. This shows item (i) of the Opial lemma.

Let us return to the fact that, according to (3.14) and (3.36), the sequences of nonnegative numbers

$$(\mathcal{E}^k_{0,b} + st_k^2(f_{\lambda_k}(x_k) - \min f))_{k\geq 2} \text{ and } (\mathcal{E}^k_{a,b} + st_k^2(f_{\lambda_k}(x_k) - \min f))_{k\geq 2}$$

are nonincreasing after an index $k_2 \geq 2$ and therefore converge. Since $\|x_k - z\|$ converges, and $t_k\|\nabla f_{\lambda_k}(x_k)\| \to 0$, as $k \to +\infty$ we obtain that the following limit exists:

$$\lim_{k\to+\infty}\left(t_k^2\|x_k - x_{k-1}\|^2 + t_k^2(f_{\lambda_k}(x_k) - \min f)\right). \tag{3.54}$$

On the other hand, according to (3.40) we have $\sum_{k=1}^{+\infty} t_k(f_{\lambda_k}(x_k) - \min f) < +\infty$, and according to (3.43) we have $\sum_{k=2}^{+\infty} t_k\|x_k - x_{k-1}\|^2 < +\infty$. Therefore,

$$\sum_{k=2}^{+\infty}\frac{1}{t_k}\left(t_k^2\|x_k - x_{k-1}\|^2 + t_k^2(f_{\lambda_k}(x_k) - \min f)\right) < +\infty. \tag{3.55}$$

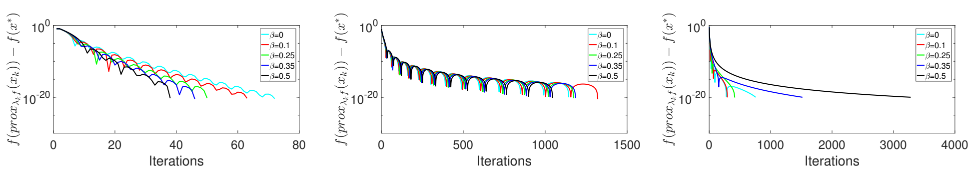
Combining (3.54) and (3.55) we obtain that $\lim_{k\to+\infty}(t_k^2\|x_k - x_{k-1}\|^2 + t_k^2(f_{\lambda_k}(x_k) - \min f)) = 0$, that is, $\|x_k - x_{k-1}\| = o(\frac{1}{k})$ and $f_{\lambda_k}(x_k) - \min f = o(\frac{1}{k^2})$ as $k \to +\infty$.

From $f_{\lambda_k}(x_k) - \min f = f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f + \frac{1}{2\lambda_k}\|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\|^2$, we deduce that

$$(3.56)\qquad f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f = o\left(\frac{1}{k^2}\right), \quad \lim_{k\to+\infty} \frac{k^2}{\lambda_k}\|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\|^2 = 0.$$

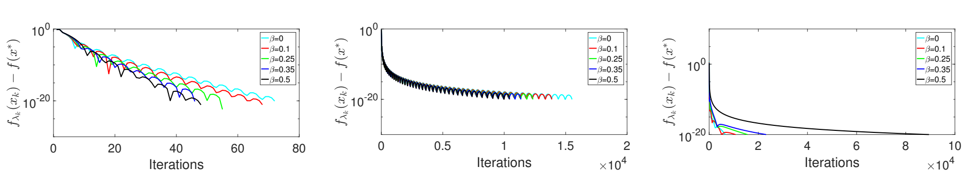
It remains to show that every weak cluster point x^* of the sequence (x_k) belongs to $\operatorname{argmin} f$. Let $x_{k_n} \rightharpoonup x^*$, $n \to +\infty$. If $t \leq 2$, then one has $\lim_{k\to+\infty}\|x_k - \operatorname{prox}_{\lambda_k f}(x_k)\| = 0$. Therefore, $\operatorname{prox}_{\lambda_{k_n} f}(x_{k_n}) \rightharpoonup x^*$, $n \to +\infty$. Since f is lower semicontinuous and convex, it is weakly lower semicontinuous. Combined with $\lim_{k\to+\infty} (f(\operatorname{prox}_{\lambda_k f}(x_k)) - \min f) = 0$, it yields $0 = \liminf_{n\to+\infty}(f(\operatorname{prox}_{\lambda_{k_n} f}(x_{k_n})) - \min f) \geq f(x^*) - \min f$. The latter relation shows that $x^* \in \operatorname{argmin} f$ and consequently, according to the Opial lemma, the sequence (x_k) converges weakly to an element $\hat{x} \in \operatorname{argmin} f$. Finally, since $\|x_k - x_{k-1}\| = o(1/k)$ and (α_k) is bounded, for $\beta = 0$ we get $\|x_k - y_k\| = o(1/k)$, as $k \to +\infty$. Therefore y_k converges weakly to the same element $\hat{x} \in \operatorname{argmin} f$. If $\beta > 0$ and $t > 1$, then (3.52) gives $\|\nabla f_{\lambda_k}(x_k) - \nabla f_{\lambda_{k-1}}(x_{k-1})\| = o\left(1/k^2\right)$, as $k \to +\infty$. Combining this latter relation with the fact that $\|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right)$ and (α_k) is bounded, we get $\|x_k - y_k\| = o(1/k)$. Therefore, (y_k) converges weakly to the same element $\hat{x} \in \operatorname{argmin} f$. ∎

3.2. Numerical experiments. The following numerical experiments show the good behavior of (PRINAM-convex). As a model of poorly conditioned quadratic function, take $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = \frac{1}{2}(x^2 + 1000y^2)$. Denote by $x^* = (0,0)$ the unique minimizer of f. Then, for $\lambda > 0$ $f_\lambda(x,y) = \frac{1}{2(\lambda+1)}x^2 + \frac{500}{1000\lambda+1}y^2$. We implement (PRINAM-convex) for $\lambda_k = \lambda k^t$, where in concordance to the hypotheses of Theorem 3.2 we take into account the following instances. For $t = 0$ we take $s = 1.1$ and $\lambda = 1.1$, for $t = 1$ we take $s = 0.5$, $\lambda = 1.1$, and for $t = 2$ we consider $s = 1$ and $\lambda = \frac{1.01(2\beta+s)^2 r^2}{s}$. Further, for different instances of β everywhere we consider $r = 0.3$, $q = 0.5$. The initial points x_1, x_2 are taken equal to $(1, -1)$. The convergence rates are shown in Figure 3.1(a)–(c) and Figure 3.2(a)–(c).



(a) $s = 1.1$, $\lambda = 1.1$, $t = 0$ (b) $s = 0.5$, $\lambda = 1.1$, $t = 1$ (c) $s = 1$, $\lambda = \frac{1.01(2\beta+s)^2)r^2}{s}$, $t = 2$

FIG. 3.1. *The behaviour of* (PRINAM-convex).



(a) $s = 1.1$, $\lambda = 1.1$, $t = 0$ (b) $s = 0.5$, $\lambda = 1.1$, $t = 1$ (c) $s = 1$, $\lambda = \frac{1.01(2\beta+s)^2)r^2}{s}$, $t = 2$

FIG. 3.2. *Moreau envelope error for* (PRINAM-convex).

4. Conclusion, perspectives. (PRINAM) is a proximal-based inertial algorithm which aims to solve general monotone inclusions in Hilbert spaces. It has several favorable features:

1. Under the sole assumption that the solution set is nonempty, each sequence generated by the algorithm converges weakly to a zero of the operator.

2. The algorithm involves a correcting term which is naturally linked to the Hessian driven damping in the case of convex minimization and to the Newton method for general monotone inclusions. There is numerical evidence that this correcting term attenuates the oscillations which naturally occur with the inertial methods. The algorithm falls into the category of adaptive algorithms where the damping takes into account the geometry of the problem, a class of algorithms which is the subject of current research; see [38].

3. When specializing the operator to the subdifferential of a convex lower semicontinuous proper function, the algorithm improves the accelerated gradient method of Nesterov by giving the convergence rate $o(\frac{1}{k^2})$ of the values and the fast convergence of the gradients toward zero.

The article presents the basic elements of the convergence theory for (PRINAM), many aspects of which have yet to be developed. We need to enlarge the framework by considering structured composite monotone inclusions and show how to use (PRINAM) as the basic block of splitting algorithms such as ADMM and Douglas–Rachford, to cite some of them. For numerical reasons, it is important to consider the introduction of perturbations and errors in the algorithms. Considering a Tikhonov regularization term with vanishing coefficient would allow to obtain strong convergence of the iterates towards the minimum norm solution, a desirable feature for the inverse problems.

Appendix A. Auxiliary results. In our analysis of (PRINAM) we need the following results. We omit the proofs, which follow from standard arguments.

LEMMA A.1. *Let (a_k) be a sequence of nonnegative real numbers which satisfies that there exists $k' \geq 0$ such that for all $k \geq k'$ $a_{k+1} \leq \alpha_k a_k + \omega_k$, where $\sum_k t_{k+1}\omega_k < +\infty$. Then $\sum_k a_k < +\infty$.*

LEMMA A.2. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, and let $\gamma, \nu > 0$ and $x, y \in \mathcal{H}$. Then, the following estimates hold:*

(a) $\|\gamma A_\gamma(x) - \nu A_\nu(y)\| \leq 2\|x - y\| + |\gamma - \nu|\|A_\gamma(x)\|$.

(b) $\|A_\gamma(x) - A_\nu(y)\| \leq \frac{2}{\gamma}\|x - y\| + \frac{|\gamma-\nu|}{\gamma}(\|A_\gamma(x)\| + \|A_\nu(y)\|)$.

REFERENCES

[1] B. Abbas, H. Attouch, and B. F. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, J. Optim. Theory Appl., 161 (2014), pp. 331–360.

[2] F. Álvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Anal., 9 (2001), pp. 3–11.

[3] F. Alvarez, H. Attouch, J. Bolte, and P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping*, J. Math. Pures Appl., 81 (2002), pp. 747–779.

[4] V. Apidopoulos, J.-F. Aujol, and Ch. Dossal, *Convergence rate of inertial forward-backward algorithm beyond Nesterov's rule*, Math. Program., 180 (2020), pp. 137–156, https://doi.org/10.1007/s10107-018-1350-9.

[5] H. Attouch and A. Cabot, *Convergence of a relaxed inertial proximal algorithm for maximally monotone operators*, Math. Program., 184 (2020), pp. 243–287, https://doi.org/10.1007/s10107-019-01412-0.

[6] H. Attouch and A. Cabot, *Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions*, Appl. Math. Optim., 80 (2019), pp. 547–598.
[7] H. Attouch and A. Cabot, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., 28 (2018), pp. 849–874.
[8] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, Math. Program. (2020), https://doi.org/10.1007/s10107-020-01591-1.
[9] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program. Ser. B, 168 (2018), pp. 123–175.
[10] H. Attouch, Z. Chbani, and H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case* $\alpha \leq 3$, ESAIM Control Optim. Calc. Var., 25 (2019), https://doi.org/10.1051/cocv/2017083.
[11] H. Attouch and P. E. Maingé, *Asymptotic behavior of second order dissipative evolution equations combining potential with non-potential effects*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 836–857.
[12] H. Attouch, P. E. Maingé, and P. Redont, *A second-order differential system with Hessian-driven damping; Application to non-elastic shock laws*, Differ. Equ. Appl., 4 (2012), pp. 27–65.
[13] H. Attouch, M. Marques Alves, and B. F. Svaiter, *A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert Spaces, with complexity* $\mathcal{O}(1/n^2)$, J. Convex Anal., 23 (2016), pp. 139–180.
[14] H. Attouch and J. Peypouquet, *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators*, Math. Program., 174 (2019), pp. 391–432.
[15] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than* $1/k^2$, SIAM J. Optim., 26 (2016), pp. 1824–1834.
[16] H. Attouch, J. Peypouquet, and P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations, 26, (2016), pp. 5734–5783.
[17] H. Attouch, P. Redont, and B. F. Svaiter, *Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces*, J. Optim. Theory Appl., 157 (2013), pp. 624–650.
[18] H. Attouch and B. F. Svaiter, *A continuous dynamical Newton-Like approach to solving monotone inclusions*, SIAM J. Control Optim., 49 (2011), pp. 574–598.
[19] H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Math., Springer, New York, 2011.
[20] C. Castera, J. Bolte, C. Févotte, and E. Pauwels, *An Inertial Newton Algorithm for Deep Learning*, HAL-02140748, 2019.
[21] A. Böhm, M. Sedlmayer, E. R. Csetnek, and R. I. Boţ, *Two Steps at a Time—Taking GAN Training in Stridewith Tseng's Method*, arXiv:2006.09033v1 [math.OC], 2020.
[22] R. I. Boţ and E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim., 54 (2016), pp. 1423–1443.
[23] R. I. Boţ, E. R. Csetnek, and S. C. László, *Tikhonov regularization of a second order dynamical system with Hessian damping*, Math. Program., to appear, https://doi.org/10.1007/s10107-020-01528-8.
[24] A. Chambolle and Ch. Dossal, *On the convergence of the iterates of the fast iterative shrinkage thresholding algorithm*, J. Optim. Theory Appl., 166 (2015), pp. 968–982.
[25] P. L. Combettes, *Monotone operator theory in convex optimization*, Math. Program. Ser. B, 170 (2018), pp. 177–206.
[26] P. L. Combettes and L. Glaudin, *Quasi-nonexpansive iterations on the affine hull of orbits: From Mann's mean value algorithm to inertial methods*, SIAM J. Optim., 27 (2017), pp. 2356–2380.
[27] P. L. Combettes and L. Glaudin, *Proximal activation of smooth functions in splitting algorithms for convex image recovery*, SIAM J. Imaging Sci., 12 (2019), pp. 1905–1935.
[28] E. R. Csetnek, Y. Malitsky, and M. K. Tam, *Shadow Douglas-Rachford splitting for monotone inclusions*, Appl. Math. Optim., 80 (2019), pp. 665–678.
[29] G. Gu and J. Yang, *Optimal Nonergodic Sublinear Convergence Rate of Proximal Point Algorithm for Maximal Monotone Inclusion Problems*, arXiv:1904.05495v2 [math.OC], 2019.
[30] F. Iutzeler and J. M. Hendrickx, *A generic online acceleration scheme for optimization algorithms via relaxation and inertia*, Optim. Methods Softw., 34 (2019), pp. 383–405.
[31] D. Kim, *Accelerated Proximal Point Method for Maximally Monotone Operators*, preprint, arXiv:1905.05149v3 [math.OC], 2020.

[32] T. Lin and M. I. Jordan, *A Control-Theoretic Perspective on Optimal High-Order Optimization*, arXiv:1912.07168v1 [math.OC], 2019.

[33] D. A. Lorenz and T. Pock, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vision, 51 (2015), pp. 311–325.

[34] Y. Malitsky and M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM J. Optim., 30 (2020), pp. 1451–1472.

[35] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish J. Math., 41 (2017), pp. 681–685.

[36] A. Moudafi and M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, J. Comput. Appl. Math., 155 (2003), pp. 447–454.

[37] J. Peypouquet and S. Sorin, *Evolution equations for maximal monotone operators: Asymptotic analysis in continuous and discrete time*, J. Convex Anal., 17 (2020), pp. 1113–1163.

[38] C. Poon and J. Liang, *Geometry of First-order Methods and Adaptive Acceleration*, arXiv:2003.03910v1, 2020.

[39] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[40] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su, *Understanding the Acceleration Phenomenon via High-Resolution Differential Equations*, arXiv:1810.08907, 2018.

[41] W. J. Su, S. Boyd, and E. J. Candès, *A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights*, Neural Information Processing Systems, 27 (2014), pp. 2510–2518.