Fast second-order dynamics with slow vanishing damping approaching the zeros of a monotone and continuous operator

Radu Ioan Boţ* David Alexander Hulett† Dang-Khoa Nguyen‡§

Abstract

In this work, we approach the problem of finding the zeros of a continuous and monotone operator through a second-order dynamical system with a damping term of the form $1/t^r$, where $r \in [0,1]$. The system features the time derivative of the operator evaluated along the trajectory, which is a Hessian-driven type damping term when the governing operator comes from a potential. Also entering the system is a time rescaling parameter $\beta(t)$ which satisfies a certain growth condition. We derive $o\left(\frac{1}{t^{2r}\beta(t)}\right)$ convergence rates for the norm of the operator evaluated along the generated trajectories as well as for a gap function which serves as a measure of optimality for the associated variational inequality. The parameter r enters the growth condition for $\beta(t)$: when $r < 1$, the damping $1/t^r$ approaches zero at a slower speed than Nesterov's $1/t$ damping; in this case, we are allowed to choose $\beta(t)$ to be an exponential function, thus having linear convergence rates for the involved quantities. We also show weak convergence of the trajectories towards zeros of the governing operator. Through a particular choice for the operator, we establish a connection with the problem of minimizing a smooth and convex function with linear constraints. The convergence rates we derived in the operator case are inherited by the objective function evaluated at the trajectories and for the feasibility gap. We also prove weak convergence of the trajectories towards primal-dual solutions of the problem.

A discretization of the dynamical system yields an implicit algorithm that exhibits analogous convergence properties to its continuous counterpart.

We complement our theoretical findings with two numerical experiments.

Key Words. monotone equation, variational inequality, slow asymptotically vanishing damping, sublinear convergence rates, linear convergence rates, convergence of trajectories

AMS subject classification. 47H05, 47J20, 65K10, 65K15

*Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: radu.bot@univie.ac.at. Research partially supported by FWF (Austrian Science Fund), project W 1260, and by a grant of the Romanian Ministry of Research, Innovation and Digitization, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2021-0138, within PNCDI III.

†Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: david.alexander.hulett@univie.ac.at. Research partially supported by the VGSCO (Vienna Graduate School on Computational Optimization), project W 1260.

‡Faculty of Mathematics and Computer Science, University of Science, Ho Chi Minh City, Vietnam, e-mail: ndkhoa@hcmus.edu.vn. Research funded by the Postdoctoral Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2024.STS.37

§Vietnam National University, Ho Chi Minh City, Vietnam

Contents

1 Introduction **2**
- 1.1 Problem statement and motivation . . . 2
- 1.2 Previous and related continuous time systems . . . 4
- 1.3 Our work . . . 5

2 Continuous time system **7**
- 2.1 Convergence rates and weak convergence of trajectories . . . 7
- 2.2 Linearly constrained convex minimization as a particular case . . . 19
- 2.3 Some numerical experiments . . . 21

3 An implicit discretization of the continuous system **23**
- 3.1 Discretization of an equivalent first-order reformulation . . . 23
- 3.2 Convergence rates and weak convergence of iterates . . . 24

A Auxiliary results **41**

B Proof of Lemma 3.5 **45**

1 Introduction

We dedicate this work to the memory of Hedy Attouch, a remarkable mathematician and a kind, generous soul whose absence is deeply felt.

1.1 Problem statement and motivation

In the setting of a real Hilbert space $\mathcal{H}$ and a monotone and continuous operator $V : \mathcal{H} \to \mathcal{H}$, we study the following problem:

$$\text{find } z_* \in \mathcal{H} \text{ such that } V(z_*) = 0. \tag{1}$$

It is simple to see that the continuity and monotonicity of V ensure that z_* satisfies (1) if and only if

$$\langle z - z_*, V(z) \rangle \geq 0 \quad \forall z \in \mathcal{H}. \tag{2}$$

One of the principal motivations to study the monotone inclusion (1) comes from minimax problems. Indeed, consider

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y), \tag{3}$$

where $\mathcal{X}, \mathcal{Y}$ are real Hilbert spaces and $\Phi \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is continuously differentiable, convex in the first variable and concave in the second one. Solutions to (3) are saddle points of Φ: that is, a pair $(x_*, y_*) \in \mathcal{X} \times \mathcal{Y}$ such that

$$\Phi(x_*, y) \leq \Phi(x_*, y_*) \leq \Phi(x, y_*) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

This is equivalent to

$$\begin{cases} \nabla_x \Phi(x_*, y_*) & = 0, \\ -\nabla_y \Phi(x_*, y_*) & = 0, \end{cases}$$

and this is nothing else than a monotone inclusion problem involving the monotone and continuous operator $V : \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y}$ given by

$$V(x,y) := \Big(\nabla_x \Phi(x,y), -\nabla_y \Phi(x,y)\Big).$$

Formulations (1) and (3) underlie numerous problems in different fields, such as optimization, economics, game theory, and partial differential equations, and are of particular interest to the machine learning community: for example, they play a fundamental role in areas such as multi-agent reinforcement learning [17], robust adversarial learning [16] and generative adversarial networks (GANs) [9, 13]. An interesting example of (3), upon which we will elaborate further later, is linearly constrained convex minimization. Precisely, let us consider

$$\begin{array}{rl} \min & f(x), \\ \text{subject to} & Ax = b, \end{array} \tag{4}$$

where $f : \mathcal{X} \to \mathbb{R}$ is convex and continuously differentiable, $b \in \mathcal{Y}$ and $A : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator. Associated to (4) is the Lagrangian $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ given by

$$\mathcal{L}(x,\lambda) := f(x) + \langle \lambda, Ax - b\rangle.$$

Primal-dual solutions to (4), that is, pairs $(x_*, \lambda_*) \in \mathcal{X} \times \mathcal{Y}$ which satisfy

$$\begin{cases} \nabla f(x_*) + A^* \lambda_* & = 0, \\ b - Ax_* & = 0 \end{cases}$$

are precisely the saddle points of $\mathcal{L}$, i.e., the zeros of the monotone and continuous operator

$$V(x,\lambda) := \Big(\nabla_x \mathcal{L}(x,\lambda), -\nabla_\lambda \mathcal{L}(x,\lambda)\Big) = \Big(\nabla f(x) + A^*\lambda, b - Ax\Big). \tag{5}$$

Attached to (1), we will investigate the asymptotic properties of the trajectories generated by a certain second-order dynamical system. The system features a vanishing damping of the form $\frac{\alpha}{t^r}$, where $r \in [0,1]$, which approaches zero more slowly than Nesterov's classical $\frac{\alpha}{t}$ when $r < 1$. This is combined with the time derivative of the operator evaluated along the trajectory, which is also known as a Hessian-driven damping term when V is the gradient of a continuously differentiable function. The convergence behaviour of the solutions is greatly affected by the presence of a time rescaling parameter $\beta(t)$ which is positive, nondecreasing, and continuously differentiable, and needs to satisfy a certain growth condition. For the generated trajectory $z(t)$, we will derive rates of convergence of $o\left(\frac{1}{t^{2r}\beta(t)}\right)$ for $\|V(z(t))\|$ and for the restricted gap function associated to (2) as $t \to +\infty$. When $r < 1$, we are allowed to choose $\beta(t)$ such that $\beta(t)$ grows exponentially with time; such a choice is not possible when $r = 1$, i.e., when we take Nesterov's damping. We will also show weak convergence of $z(t)$ towards a zero of V as $t \to +\infty$.

When V is chosen to be (5), we recover a primal-dual system formulated in the spirit of He et al. [15]. However, it differs in the sense that our primal-dual system includes a Hessian-driven damping term. For the generated primal-dual trajectory $(x(t), \lambda(t))$, we will show convergence rates of $o\left(\frac{1}{t^{2r}\beta(t)}\right)$ for the functional values along the primal trajectory, the primal-dual gap formulated in terms of the Lagrangian $\mathcal{L}$, and for the feasibility gap which gauges how far the generated primal trajectory is from the constraints. Additionally, we will furnish the weak convergence of $(x(t), \lambda(t))$ towards a primal-dual solution of the minimization problem as $t \to +\infty$.

A temporal discretization of our system will produce an implicit algorithm that mimics the convergence properties of the continuous time system. If $(\beta_k)_{k\geq 1}$ is a positive, nondecreasing sequence which satisfies an analogous growth condition to the continuous one, and $(z^k)_{k\geq 1}$ are the iterates generated by the algorithm, we will produce convergence rates of $o\left(\frac{1}{k^{2r}\beta_k}\right)$ for $\left\|V(z^k)\right\|$ and for the discrete gap function associated to (2) as $k \to +\infty$. Similar to the continuous case, $r < 1$ allows us to choose β_k such that β_k grows exponentially as $k \to +\infty$. Furthermore, we will show that the iterates converge weakly to a zero of V.

1.2 Previous and related continuous time systems

In the last years, there have been many advances in the study of continuous time systems attached to monotone inclusion problems. We briefly visit them in the following paragraphs.

Extending the Heavy ball with friction dynamics introduced by Álvarez in [1] for unconstrained minimization, Álvarez and Attouch [2] and Attouch and Maingé [5] studied the dynamics

$$\ddot{z}(t) + \mu\dot{z}(t) + A(z(t)) = 0, \tag{6}$$

where $A\colon \mathcal{H} \to \mathcal{H}$ is a λ-cocoercive operator and $\mu > 0$. The authors showed that under the assumption $\lambda\mu^2 > 1$, a solution to (6) weakly converges to a zero of A. Recall that for a maximally monotone (but not necessarily single-valued) operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, its Yosida approximation of index $\lambda > 0$ is given by

$$A_\lambda := \frac{1}{\lambda}(\mathrm{Id} - J_{\lambda A}),$$

where $J_{\lambda A} := \left(\mathrm{Id} + \lambda A\right)^{-1}$ is the resolvent of λA. The operator A_λ is single-valued, λ-cocoercive, and shares the same set of zeros of A. Therefore, if $\lambda\mu^2 > 1$, then any solution to

$$\ddot{z}(t) + \mu\dot{z}(t) + A_\lambda(z(t)) = 0$$

weakly converges to a zero of A.

Related to (6), Boţ and Csetnek [10] studied the system

$$\ddot{z}(t) + \gamma(t)\dot{z}(t) + \nu(t)A(z(t)) = 0, \tag{7}$$

where $A\colon \mathcal{H} \to \mathcal{H}$ is again λ-cocoercive. Under the assumption that γ and ν are locally absolutely continuous, $\dot{\gamma}(t) \leq 0 \leq \dot{\nu}(t)$ for almost every $t \in [0, +\infty)$ and $\inf_{t\geq 0} \frac{\gamma^2(t)}{\nu(t)} > \frac{1}{\lambda}$, the solutions to this system converge weakly to zeros of A.

Linked with the Newton and Levenberg-Marquardt methods, in [8] Attouch and Svaiter studied the following first-order differential inclusion:

$$\begin{cases} v(t) \in A(z(t)), \\ \dot{z}(t) + \beta(t)\dot{v}(t) + \beta(t)v(t) = 0, \end{cases} \tag{8}$$

where $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone and β is positive and locally absolutely continuous. Under the assumption $\sup_{t\geq 0} \frac{\dot{\beta}(t)}{\beta(t)} \leq 1$, the authors proved that the solutions to this system converge weakly to zeros of A.

Ever since the landmark work by Su et al. in [19], where the authors provided a continuous time counterpart of Nesterov's accelerated gradient algorithm, including a damping term $\frac{\alpha}{t}$ attached to the velocity has been widely successful in having an accelerating behaviour for convex minimization problems in the unconstrained (see, for example, Attouch et al. [7]) and linearly constrained settings (see He et al. [14, 15], Zeng et al. [20], Attouch et al. [3], Boţ and Nguyen [12]). Interestingly, the effects of this vanishing damping go beyond the optimization setting: for $A : \mathcal{H} \to 2^{\mathcal{H}}$ maximally monotone and $t \geq t_0 > 0$, in [6] Attouch and Peypouquet addressed the Nesterov-like system

$$\ddot{z}(t) + \frac{\alpha}{t}\dot{z}(t) + A_{\lambda(t)}(z(t)) = 0,$$

where $\alpha > 1$ and $\lambda(t) = \mathcal{O}(t^2)$ as $t \to +\infty$. Solutions to this system converge weakly to zeros of A. Furthermore, a solution $t \mapsto z(t)$ to this system is shown to satisfy the convergence rates

$$\|\dot{z}(t)\| = \mathcal{O}\left(\frac{1}{t}\right), \quad \|\ddot{z}(t)\| = \mathcal{O}\left(\frac{1}{t^2}\right) \quad \text{and} \quad \left\|A_{\lambda(t)}(z(t))\right\| = o\left(\frac{1}{t^2}\right) \tag{9}$$

as $t \to +\infty$.

Later, Attouch and László [4] introduced a Hessian-driven damping term to the previous system

$$\ddot{z}(t) + \frac{\alpha}{t}\dot{z}(t) + \xi\frac{d}{dt}\left(A_{\lambda(t)}(z(t))\right) + A_{\lambda(t)}(z(t)) = 0, \tag{10}$$

where $\xi \geq 0$. Through a more involved Lyapunov analysis, they proved that the rates (9) are preserved. Moreover, they showed that for $\xi > 0$

$$\|\dot{z}(t)\| = o\left(\frac{1}{t}\right), \quad \text{and} \quad \left\|\frac{d}{dt}\left(A_{\lambda(t)}(z(t))\right)\right\| = \mathcal{O}\left(\frac{1}{t^3}\right)$$

as $t \to +\infty$. Numerical experiments ran in [4] also show that for $\xi = 0$ the trajectories $z(t)$ show an oscillatory behaviour, while for $\xi > 0$ these oscillations are attenuated.

For a continuous and monotone operator $V : \mathcal{H} \to \mathcal{H}$ and $t \geq t_0 > 0$, Boţ et al. [11] considered the second order dynamics

$$\ddot{z}(t) + \frac{\alpha}{t}\dot{z}(t) + \beta(t)\frac{d}{dt}V(z(t)) + \frac{1}{2}\left(\dot{\beta}(t) + \frac{\alpha}{t}\beta(t)\right)V(z(t)) = 0, \tag{11}$$

where $\alpha > 0$ and β is positive, nondecreasing and continuously differentiable. Assuming $\alpha > 2$ and the growth condition

$$\sup_{t \geq t_0} \frac{t\dot{\beta}(t)}{\beta(t)} < \alpha - 2, \tag{12}$$

for z_* a zero of V and $t \mapsto z(t)$ a solution to the authors produce the convergence rates

$$\|\dot{z}(t)\| = o\left(\frac{1}{t}\right), \quad \|V(z(t))\| = o\left(\frac{1}{t\beta(t)}\right) \quad \text{and} \quad \langle z(t) - z_*, V(z(t))\rangle = o\left(\frac{1}{t\beta(t)}\right)$$

as $t \to +\infty$. Furthermore, the trajectories converge weakly to zeros of V.

1.3 Our work

In the context of linearly constrained convex minimization, for a primal-dual system like that in Zeng et al. [20], Attouch et al. [3] considered a vanishing damping term of the form $\frac{\alpha}{t^r} + \frac{r}{t}$, while He et al. [15] measured the effects of a term $\frac{\alpha}{t^r}$. In both cases, the choice $r < 1$ allowed the authors to derive exponential rates of convergence for the primal-dual gap. We wanted to see if this influence carries over to a more general setting of a continuous and monotone operator $V : \mathcal{H} \to \mathcal{H}$. In the following, we state the system and the hypotheses on its parameters, which will be assumed throughout Section 2.

Continuous time scheme: for $t \geq t_0 > 0$, we will study the asymptotic properties of the solutions to

$$\ddot{z}(t) + \frac{\alpha}{t^r}\dot{z}(t) + \theta t^r \beta(t)\frac{d}{dt}V(z(t)) + \beta(t)V(z(t)) = 0. \tag{13}$$

Here, $r \in [0,1]$, and $\alpha, \theta > 0$ are taken as follows:

- If $r \in [0,1)$, then $\frac{2}{\alpha} < \theta$;
- If $r = 1$, then $\frac{2}{\alpha+1} \leq \theta < \frac{1}{2}$.

$\beta : [t_0, +\infty) \to (0, +\infty)$ is a continuously differentiable and nondecreasing function which satisfies the following growth condition:

$$\sup_{t \geq t_0} t^r \left(\frac{\dot{\beta}(t)}{\beta(t)} + \frac{2r}{t} \right) < \frac{1}{\theta}. \tag{14}$$

For z_* an arbitrary zero of V and $t \mapsto z(t)$ a solution to (13), we will show that

$$\|V(z(t))\| = o\left(\frac{1}{t^{\rho+r}\beta(t)}\right), \quad \langle z(t) - z_*, V(z(t))\rangle = o\left(\frac{1}{t^{2\rho}\beta(t)}\right) \quad \text{and} \quad \|\dot{z}(t)\| = o\left(\frac{1}{t^\rho}\right)$$

as $t \to +\infty$, where $\rho = 0$ or $\rho = 1$ if $r = 0$ or $r = 1$ respectively, and $\rho \in (0, r)$ when $r \in (0,1)$. This allows us to get as close to the state-of-the-art rates as possible: if we assume that $z(t)$ is bounded when $r \in (0,1)$, we are able to plug $\rho = r$. The weak convergence of $z(t)$ towards a zero of V can be guaranteed in the cases $r = 0$ and $r = 1$, and in the case $r \in (0,1)$ provided $z(t)$ is assumed bounded.

Remark 1.1. When $r < 1$, the growth condition (14) holds if for some $0 < \delta < \frac{1}{\theta}$ the following differential equation is fulfilled by β:

$$t^r \left(\frac{\dot{\beta}(t)}{\beta(t)} + \frac{2r}{t} \right) = \frac{1}{\theta} - \delta \quad \forall t \geq t_0.$$

A solution to this equation is

$$\beta(t) = \frac{1}{t^{2r}} \cdot \exp\left(\left(\frac{1}{\theta} - \delta \right) \frac{t^{1-r}}{1-r} \right),$$

thus convergence rates of $o\left(\exp\left(-\left(\frac{1}{\theta} - \delta\right)\frac{t^{1-r}}{1-r}\right)\right)$ are attained for $\|V(z(t))\|$ and $\langle z(t) - z_*, V(z(t))\rangle$ as $t \to +\infty$. To our knowledge, this is the first time such fast rates are achieved in the general context of equations governed by monotone and continuous operators.

A discretization of (13) yields an implicit algorithm, which has analogous convergence properties to its continuous counterpart. We will only consider $r \in (0,1]$.

Discrete time scheme: let z^0, z^1 be initial points in $\mathcal{H}$. We iteratively define for all $k \geq 1$

$$z^{k+1} := z^k + \frac{k^r}{\alpha - rk^{r-1} + (k+1)^r}(z^k - z^{k-1}) - \frac{\theta k^{2r}\beta_{k-1}}{\alpha - rk^{r-1} + (k+1)^r}\Big[V(z^{k+1}) - V(z^k)\Big] - \frac{\theta\Big\{\Big[(k+1)^{2r} - k^{2r}\Big] - 2rk^{2r-1}\Big\} + k^r}{\alpha - rk^{r-1} + (k+1)^r}\beta_k V(z^{k+1}). \tag{15}$$

Here, $r \in (0, 1]$, $\alpha, \theta > 0$ are taken as follows:

- If $r \in (0, 1)$, then $\frac{2}{\alpha} < \theta$;
- If $r = 1$, then $\frac{2}{\alpha+1} \leq \theta < \frac{1}{4}$.

The sequence $(\beta_k)_{k\geq 0}$ is positive, nondecreasing, and it satisfies

$$\sup_{k\geq k_0} k^r \left(\frac{\beta_k - \beta_{k-1}}{\beta_k} + \frac{2r}{k}\right) < \frac{1}{2\theta} \tag{16}$$

for some positive integer k_0.

For z_* an arbitrary zero of V and $(z^k)_{k\geq 0}$ given by (15), we will prove that

$$\|V(z^k)\| = o\left(\frac{1}{k^{\rho+r}\beta_k}\right), \quad \langle z^k - z_*, V(z^k)\rangle = o\left(\frac{1}{k^{2\rho}\beta_k}\right) \quad \text{and} \quad \|z^k - z^{k-1}\| = o\left(\frac{1}{k^{\rho}}\right)$$

as $k \to +\infty$, where $\rho = 1$ if $r = 1$, and $\rho \in (0, r)$ when $r \in (0, 1)$. Like in the continuous case, this allows us to get as close to the state-of-the-art rates as possible: if we assume that $(z^k)_{k\geq 0}$ is bounded when $r \in (0, 1)$, we are able to plug $\rho = r$. The weak convergence of $(z^k)_{k\geq 0}$ towards a zero of V can be guaranteed in the cases $r = 0$ and $r = 1$, and in the case $r \in (0, 1)$ provided $(z^k)_{k\geq 0}$ is assumed bounded.

Remark 1.2. Similar to the continuous case, when $r \in (0, 1)$, we will later show that for some $0 < \delta < \frac{1}{2\theta}$ the sequence

$$\beta_k = \frac{1}{k^{2r}} \cdot e^{\left(\frac{1}{2\theta} - \delta\right)\frac{k^{1-r}}{1-r}} \ \forall k \geq k_0$$

fulfills the discrete growth condition (16). Hence, a convergence rate of $o\left(e^{-\left(\frac{1}{2\theta} - \delta\right)\frac{k^{1-r}}{1-r}}\right)$ is exhibited by $\|V(z^k)\|$ and $\langle z^k - z_*, V(z^k)\rangle$ as $k \to +\infty$. To our knowledge, this is the first time such fast rates have been achieved in the general context of implicit algorithms governed by monotone and continuous operators.

2 Continuous time system

2.1 Convergence rates and weak convergence of trajectories

Energy function: to analyze the asymptotic properties of the solutions to (13), we will make use of the following energy function:

$$\mathcal{E}_\lambda(t) = \frac{1}{2}\Big\|2\lambda t^{\rho-r}(z(t) - z_*) + 2t^\rho \dot{z}(t) + \theta t^{\rho+r}\beta(t)V(z(t))\Big]\Big\|^2 \tag{17}$$

$$+ 2\lambda t^{2(\rho-r)}(\alpha - (2\rho - r)t^{r-1} - \lambda)\|z(t) - z_*\|^2 \tag{18}$$
$$+ 2\lambda\theta t^{2\rho}\beta(t)\langle z(t) - z_*, V(z(t))\rangle \tag{19}$$
$$+ \frac{\theta^2}{2} t^{2(\rho+r)}\beta^2(t)\|V(z(t))\|^2. \tag{20}$$

Here, ρ and λ are taken as follows:

- If $r = 0$, then $0 < \lambda < \alpha$ and $\rho = r = 0$;
- If $r \in (0, 1)$, then $0 < \lambda < \alpha$ and $0 < \rho < r$;
- If $r = 1$, then $0 < \lambda < \alpha - 1$ and $\rho = r = 1$.

Theorem 2.1. *Suppose that $\alpha, \theta > 0$, $r \in [0, 1]$ and that $\beta : [t_0, +\infty) \to (0, +\infty)$ satisfy the assumptions laid down in subsection (1.3). Let $t \mapsto z(t)$ be a solution to (13) and let z_* be a zero of V. Consider the following convergence rates as $t \to +\infty$:*

$$\|V(z(t))\| = o\left(\frac{1}{t^{\rho+r}\beta(t)}\right), \quad \langle z(t) - z_*, V(z(t))\rangle = o\left(\frac{1}{t^{2\rho}\beta(t)}\right), \quad \|\dot{z}(t)\| = o\left(\frac{1}{t^{\rho}}\right). \tag{21}$$

The following statements are true:

(i) *If $r \in (0, 1)$, the above rates hold for $\rho \in (0, r)$. Furthermore, if $t \mapsto z(t)$ is bounded, then these rates hold for $\rho = r$.*

(ii) *If $r = 0$ or $r = 1$, then $t \mapsto z(t)$ is bounded; moreover, the above rates hold for $\rho = 0$ (if $r = 0$) and for $\rho = 1$ (if $r = 1$).*

(iii) *If $r = 0$ or $r = 1$ or $\Big(r \in (0, 1)$ and $t \mapsto z(t)$ is bounded$\Big)$, then $z(t)$ converges weakly to a zero of V as $t \to +\infty$.*

Proof. Let us first briefly describe the idea of our proof to help the readers follow it more easily.

Sketch of the proof. We start with the energy function $\mathcal{E}_\lambda(t)$. By carefully examining its derivative, we can show that for some situations $\mathcal{E}_\lambda(t)$ is eventually nonincreasing and thus bounded. Notice that the energy function can be decomposed into four smaller parts (lines (17)-(20)) which will be respectively denoted by $\mathcal{E}^i_\lambda$, with $i = 1, 2, 3, 4$. We will compute their time derivatives separately, and after some preliminary estimates, we see that $\frac{d}{dt}\mathcal{E}_\lambda(t)$ is bounded from above by either nonpositive terms or terms that can be combined to become nonpositive after completing the square.

From here, we can deduce the convergence rates as in (21) by considering different scenarios. For the critical cases $r = 0$ and $r = 1$, we can further show that the trajectories are bounded.

Finally, by utilizing the Opial's lemma A.5, we can show that the trajectories $z(t)$ converge weakly to a zero of V as $t \to +\infty$, provided that it is bounded. In order to do that, we need to produce in addition some integrability results and show that $\mathcal{E}_\lambda(t)$ converges for more than one value of λ.

Deriving the convergence rates (21). First of all, notice that $\mathcal{E}_\lambda(t)$ is eventually nonnegative. Indeed, go back to line (18). If $r \in [0,1)$, then $0 < \lambda < \alpha$ together with the fact that $t^{r-1} \to 0$ as $t \to +\infty$ gives $\alpha - (2\rho - r)t^{r-1} - \lambda \geq 0$ for large enough t. If $r = 1$, then $\alpha - (2\rho - r)t^{r-1} - \lambda \equiv \alpha - 1 - \lambda > 0$ by assumption. To carry out the proof, we compute the time derivative of $\mathcal{E}_\lambda$ at a $t \geq t_0$.

$$
\begin{aligned}
\frac{d}{dt}\mathcal{E}^1_\lambda(t) &= \Big\langle 2\lambda t^{\rho-r}(z(t)-z_*) + 2t^\rho \dot{z}(t) + \theta t^{\rho+r}\beta(t)V(z(t)),\\
&\qquad 2\lambda(\rho-r)t^{\rho-r-1}(z(t)-z_*) + (2\lambda t^{\rho-r} + 2\rho t^{\rho-1} - 2t^{\rho-r}\alpha)\dot{z}(t)\\
&\qquad + \Big[(\theta(\rho+r)t^{\rho+r-1} - 2t^\rho)\beta(t) + \theta t^{\rho+r}\dot{\beta}(t)\Big]V(z(t)) - \theta t^{\rho+r}\beta(t)\frac{d}{dt}V(z(t))\Big\rangle\\
&= 4\lambda^2(\rho-r)t^{2(\rho-r)-1}\|z(t)-z_*\|^2\\
&\quad + \Big[2\lambda t^{\rho-r}(2\lambda t^{\rho-r} + 2\rho t^{\rho-1} - 2t^{\rho-r}\alpha) + 4\lambda(\rho-r)t^{2\rho-r-1}\Big]\langle z(t)-z_*, \dot{z}(t)\rangle\\
&\quad + \Big\{2\lambda t^{\rho-r}\Big[(\theta(\rho+r)t^{\rho+r-1} - 2t^\rho)\beta(t) + \theta t^{\rho+r}\dot{\beta}(t)\Big] + 2\lambda\theta(\rho-r)t^{2\rho-1}\beta(t)\Big\}\langle z(t)-z_*, V(z(t))\rangle\\
&\quad - 2\lambda\theta t^{2\rho}\beta(t)\left\langle z(t)-z_*, \frac{d}{dt}V(z(t))\right\rangle + 2t^\rho(2\lambda t^{\rho-r} + 2\rho t^{\rho-1} - 2t^{\rho-r}\alpha)\|\dot{z}(t)\|^2\\
&\quad + \Big\{2t^\rho\Big[(\theta(\rho+r)t^{\rho+r-1} - 2t^\rho)\beta(t) + \theta t^{\rho+r}\dot{\beta}(t)\Big]\\
&\qquad + \theta t^{\rho+r}\beta(t)(2\lambda t^{\rho-r} + 2\rho t^{\rho-1} - 2t^{\rho-r}\alpha)\Big\}\langle \dot{z}(t), V(z(t))\rangle\\
&\quad - 2\theta t^{2\rho+r}\beta(t)\left\langle \dot{z}(t), \frac{d}{dt}V(z(t))\right\rangle - \theta^2 t^{2(\rho+r)}\beta^2(t)\left\langle V(z(t)), \frac{d}{dt}V(z(t))\right\rangle\\
&\quad + \theta t^{\rho+r}\beta(t)\Big[(\theta(\rho+r)t^{\rho+r-1} - 2t^\rho)\beta(t) + \theta t^{\rho+r}\dot{\beta}(t)\Big]\|V(z(t))\|^2,\\
\frac{d}{dt}\mathcal{E}^2_\lambda(t) &= 2\lambda\Big[2\alpha(\rho-r)t^{2(\rho-r)-1} - (2\rho-r)(2\rho-r-1)t^{2\rho-r-2} - 2(\rho-r)\lambda t^{2(\rho-r)-1}\Big]\|z(t)-z_*\|^2\\
&\quad + 4\lambda t^{2(\rho-r)}(\alpha - (2\rho-r)t^{r-1} - \lambda)\langle z(t)-z_*, \dot{z}(t)\rangle,\\
\frac{d}{dt}\mathcal{E}^3_\lambda(t) &= \theta\Big[4\lambda\rho t^{2\rho-1}\beta(t) + 2\lambda t^{2\rho}\dot{\beta}(t)\Big]\langle z(t)-z_*, V(z(t))\rangle + 2\lambda\theta t^{2\rho}\beta(t)\langle \dot{z}(t), V(z(t))\rangle\\
&\quad + 2\lambda\theta t^{2\rho}\beta(t)\left\langle z(t)-z_*, \frac{d}{dt}V(z(t))\right\rangle,\\
\frac{d}{dt}\mathcal{E}^4_\lambda(t) &= \frac{\theta^2}{2}\Big[2(\rho+r)t^{2(\rho+r)-1}\beta^2(t) + 2t^{2(\rho+r)}\beta(t)\dot{\beta}(t)\Big]\|V(z(t))\|^2\\
&\quad + \theta^2 t^{2(\rho+r)}\beta^2(t)\left\langle V(z(t)), \frac{d}{dt}V(z(t))\right\rangle.
\end{aligned}
$$

Putting everything together yields

$$
\begin{aligned}
\frac{d}{dt}\mathcal{E}_\lambda(t) = \Big\{&4\lambda^2(\rho-r)t^{2(\rho-r)-1} + 2\lambda\Big[2\alpha(\rho-r)t^{2(\rho-r)-1}\\
&- (2\rho-r)(2\rho-r-1)t^{2\rho-r-2} - 2(\rho-r)\lambda t^{2(\rho-r)-1}\Big]\Big\}\|z(t)-z_*\|^2
\end{aligned} \tag{22}
$$

$$
\begin{aligned}
&+ \Big[2\lambda t^{\rho-r}\big(2\lambda t^{\rho-r} + 2\rho t^{\rho-1} - 2t^{\rho-r}\alpha\big) + 4\lambda(\rho-r)t^{2\rho-r-1} \\
&\qquad + 4\lambda t^{2(\rho-r)}\big(\alpha - (2\rho-r)t^{r-1} - \lambda\big)\Big]\langle z(t)-z_*, \dot z(t)\rangle \qquad (23)\\
&+ \Big\{2\lambda t^{\rho-r}\Big[\big(\theta(\rho+r)t^{\rho+r-1} - 2t^{\rho}\big)\beta(t) + \theta t^{\rho+r}\dot\beta(t)\Big] + 2\lambda\theta(\rho-r)t^{2\rho-1}\beta(t) \\
&\qquad + \theta\Big[4\lambda\rho t^{2\rho-1}\beta(t) + 2\lambda t^{2\rho}\dot\beta(t)\Big]\Big\}\langle z(t)-z_*, V(z(t))\rangle \qquad (24)\\
&+ 2t^{\rho}\big(2\lambda t^{\rho-r} + 2\rho t^{\rho-1} - 2t^{\rho-r}\alpha\big)\|\dot z(t)\|^2 \qquad (25)\\
&+ \Big\{2t^{\rho}\Big[\big(\theta(\rho+r)t^{\rho+r-1} - 2t^{\rho}\big)\beta(t) + \theta t^{\rho+r}\dot\beta(t)\Big] \\
&\qquad + \theta t^{\rho+r}\beta(t)\big(2\lambda t^{\rho-r} + 2\rho t^{p-1} - 2t^{\rho-r}\alpha\big) + 2\lambda\theta t^{2\rho}\beta(t)\Big\}\langle \dot z(t), V(z(t))\rangle \qquad (26)\\
&- 2\theta t^{2\rho+r}\beta(t)\left\langle \dot z(t), \frac{d}{dt}V(z(t))\right\rangle \qquad (27)\\
&+ \Big\{\theta t^{\rho+r}\beta(t)\Big[\big(\theta(\rho+r)t^{\rho+r-1} - 2t^{\rho}\big)\beta(t) + \theta t^{\rho+r}\dot\beta(t)\Big] \\
&\qquad + \theta^2\Big[(\rho+r)t^{2(\rho+r)-1}\beta^2(t) + t^{2(\rho+r)}\beta(t)\dot\beta(t)\Big]\Big\}\|V(z(t))\|^2. \qquad (28)
\end{aligned}
$$

We will check that lines (22), (24), (25), (27) and (28) are nonpositive for large enough t, and that line (23) vanishes.

(22): we have

$$
\begin{aligned}
&4\lambda^2(\rho-r)t^{2(\rho-r)-1} + 2\lambda\Big[2\alpha(\rho-r)t^{2(\rho-r)-1} - (2\rho-r)(2\rho-r-1)t^{2\rho-r-2} - 2\lambda(\rho-r)t^{2(\rho-r)-1}\Big] \\
&= 4\lambda^2(\rho-r)t^{2(\rho-r)-1} + 2\lambda\Big[2\alpha(\rho-r)t^{2(\rho-r)-1} - (2\rho-r)(2\rho-r-1)t^{2\rho-r-2}\Big] - 4\lambda^2(\rho-r)t^{2(\rho-r)-1} \\
&= 2\lambda t^{2(\rho-r)-1}\Big[2\alpha(\rho-r) - (2\rho-r)(2\rho-r-1)t^{r-1}\Big].
\end{aligned}
$$

If $r = 0$, then $\rho = 0$ and the line vanishes. If $r \in (0,1)$, then $\rho < r$, which gives $2\alpha(\rho-r) < 0$. Since $t^{r-1} \to 0$ as $t \to +\infty$, for large enough t it holds $2\alpha(\rho-r) - (2\rho-r)(2\rho-r-1)t^{r-1} < 0$. If $r = 1$, then $\rho = 1$ and this line again vanishes.

(23): we have

$$
\begin{aligned}
&2\lambda t^{\rho-r}\big(2\lambda t^{\rho-r} + 2\rho t^{\rho-1} - 2\rho t^{\rho-r}\alpha\big) + 4\lambda(\rho-r)t^{2\rho-r-1} + 4\lambda t^{2(\rho-r)}\big(\alpha - (2\rho-r)t^{r-1} - \lambda\big) \\
&= 4\lambda t^{2(\rho-r)}\big(\lambda + (2\rho-r)t^{r-1} - \alpha\big) + 4\lambda t^{2(\rho-r)}\big(\alpha - (2\rho-r)t^{r-1} - \lambda\big),
\end{aligned}
$$

which means that this line is identically zero.

(24): the monotonicity of V ensures $\langle z(t)-z_*, V(z(t))\rangle \geq 0$ for all $t \geq t_0$. We have

$$
\begin{aligned}
&2\lambda t^{\rho-r}\Big[\big(\theta(\rho+r)t^{\rho+r-1} - 2t^{\rho}\big)\beta(t) + \theta t^{\rho+r}\dot\beta(t)\Big] + 2\lambda(\rho-r)\theta t^{2\rho-1}\beta(t) \\
&+ \theta\Big[4\lambda\rho t^{2\rho-1}\beta(t) + 2\lambda t^{2\rho}\dot\beta(t)\Big]
\end{aligned}
$$

$$
\begin{aligned}
&= 2\lambda t^{2\rho-r}\Big\{\Big[\theta(\rho+r+2\rho)t^{r-1}-2+(\rho-r)\theta t^{r-1}\Big]\beta(t)+2\theta t^r\dot\beta(t)\Big\}\\
&= 4\lambda t^{2\rho-r}\Big[(2\rho\theta t^{r-1}-1)\beta(t)+\theta t^r\dot\beta(t)\Big]\\
&\le 4\lambda t^{2\rho-r}\Big[(2r\theta t^{r-1}-1)\beta(t)+\theta t^r\dot\beta(t)\Big]\\
&\le 0,
\end{aligned}
$$

where the last equality comes from $\rho+r+2\rho+\rho-r=4\rho$, and the last inequality is a consequence of the growth condition on β.

(25): arguing as we did at the beginning of the proof, we have

$$
2t^\rho(2\lambda t^{\rho-r}+2\rho t^{\rho-1}-2t^{\rho-r}\alpha)=4t^{2\rho-r}(\lambda+\rho t^{r-1}-\alpha)\le 0
$$

for large enough t.

(27): since the monotonicity of V ensures $\left\langle \dot z(t), \frac{d}{dt}V(z(t))\right\rangle \ge 0$ for every $t\ge t_0$, this line is evidently nonpositive.

(28): we have

$$
\begin{aligned}
&\theta t^{2\rho+r}\beta(t)\Big[(\theta(\rho+r)t^{r-1}-2)\beta(t)+\theta t^r\dot\beta(t)\Big]+\theta t^{2\rho+r}\beta(t)\Big[\theta(\rho+r)t^{r-1}\beta(t)+\theta t^r\dot\beta(t)\Big]\\
&= 2\theta t^{2\rho+r}\beta(t)\Big[(\theta(\rho+r)t^{r-1}-1)\beta(t)+\theta t^r\dot\beta(t)\Big]\\
&\le 2\theta t^{2\rho+r}\beta(t)\Big[(2r\theta t^{r-1}-1)\beta(t)+\theta t^r\dot\beta(t)\Big]\\
&\le 0,
\end{aligned}
$$

where we used $\rho+r\le 2r$, and the growth condition on β.

Additionally, we need the term accompanying $\langle \dot z(t), V(z(t))\rangle$, which according to (26) reads

$$
\begin{aligned}
&2t^\rho\Big[(\theta(\rho+r)t^{\rho+r-1}-2t^\rho)\beta(t)+\theta t^{\rho+r}\dot\beta(t)\Big]+\theta t^{\rho+r}\beta(t)(2\lambda t^{\rho-r}+2\rho t^{\rho-1}-2t^{\rho-r}\alpha)+2\lambda\theta t^{2\rho}\beta(t)\\
&= 2t^{2\rho}\Big\{\Big[(\theta(\rho+r)t^{r-1}-2)+\theta(\lambda+\rho t^{r-1}-\alpha)+\lambda\theta\Big]\beta(t)+\theta t^r\dot\beta(t)\Big\}\\
&= 2t^{2\rho}\left\{\left[2\theta\left(\lambda+\frac{\rho+r}{2}t^{r-1}-\alpha\right)+(\theta\alpha+\theta\rho t^{r-1}-2)\right]\beta(t)+\theta t^r\dot\beta(t)\right\}.
\end{aligned}
$$

According to the computations we have now made for (22)-(28), and taking into account that $\rho\le\frac{\rho+r}{2}$, for large enough t we have

$$
\begin{aligned}
\frac{d}{dt}\mathcal{E}_\lambda(t) \le\;& 2\lambda t^{2(\rho-r)-1}\Big[2\alpha(\rho-r)-(2\rho-r)(2\rho-r-1)t^{r-1}\Big]\|z(t)-z_*\|^2 &(29)\\
&+4\lambda t^{2\rho-r}\Big[(2r\theta t^{r-1}-1)\beta(t)+\theta t^r\dot\beta(t)\Big]\langle z(t)-z_*, V(z(t))\rangle &(30)\\
&+4t^{2\rho-r}\left(\lambda+\frac{\rho+r}{2}t^{r-1}-\alpha\right)\|\dot z(t)\|^2 &(31)\\
&+2t^{2\rho}\left\{\left[2\theta\left(\lambda+\frac{\rho+r}{2}t^{r-1}-\alpha\right)+(\theta\alpha+\theta\rho t^{r-1}-2)\right]\beta(t)+\theta t^r\dot\beta(t)\right\}\langle \dot z(t), V(z(t))\rangle &(32)\\
&+2\theta t^{2\rho+r}\beta(t)\Big[(2r\theta t^{r-1}-1)\beta(t)+\theta t^r\dot\beta(t)\Big]\|V(z(t))\|^2. &(33)
\end{aligned}
$$

We will work with lines (31)-(33). Define

$$\varepsilon(t) := \alpha - \frac{\rho + r}{2} t^{r-1} - \lambda, \quad c(t) := \theta\alpha + \theta\rho t^{r-1} - 2.$$

Using our assumptions on α, θ and λ, we have that $\varepsilon(t) > 0$ and $c(t) \geq 0$ for large enough t. We will show that for certain λ, for t sufficiently large it holds

$$\begin{aligned} 0 \geq & -3\varepsilon(t)t^{2\rho-r}\|\dot{z}(t)\|^2 + 2t^{2\rho}\Big[(-2\theta\varepsilon(t) + c(t))\beta(t) + \theta t^r\dot{\beta}(t)\Big]\langle \dot{z}(t), V(z(t))\rangle \\ & + \frac{4}{3}\theta t^{2\rho+r}\beta(t)\Big[(2r\theta t^{r-1} - 1)\beta(t) + \theta t^r\dot{\beta}(t)\Big]\|V(z(t))\|^2. \end{aligned} \tag{34}$$

We will use Lemma A.2 with $X = \dot{z}(t)$ and $Y = V(z(t))$, and A, B, C chosen as follows:

$$B = t^{2\rho}\Big[(-2\theta\varepsilon(t) + c(t))\beta(t) + \theta t^r\dot{\beta}(t)\Big], \quad A = -3\varepsilon(t)t^{2\rho-r},$$
$$C = \frac{4}{3}\theta t^{2\rho+r}\beta(t)\Big[(2r\theta t^{r-1} - 1)\beta(t) + \theta t^r\dot{\beta}(t)\Big].$$

According to this lemma, it is sufficient to show that $B^2 - AC \leq 0$. We write

$$\begin{aligned} B^2 - AC &= t^{4\rho}\Big[(-2\theta\varepsilon(t) + c(t))\beta(t) + \theta t^r\dot{\beta}(t)\Big]^2 + 3\cdot\frac{4}{3}\varepsilon(t)\theta t^{2\rho-r}\cdot t^{2\rho+r}\beta(t)\Big[(2r\theta t^{r-1} - 1)\beta(t) + \theta t^r\dot{\beta}(t)\Big] \\ &= t^{4p}\Big[(-2\theta\varepsilon(t) + c(t))^2\beta^2(t) + 2\theta t^r(-2\theta\varepsilon(t) + c(t))\beta(t)\dot{\beta}(t) + \theta^2 t^{2r}(\dot{\beta}(t))^2\Big] \\ &\quad + 4\theta(2r\theta t^{r-1} - 1)\varepsilon(t)t^{4\rho}\beta^2(t) + 4\theta^2\varepsilon(t)t^{4\rho+r}\beta(t)\dot{\beta}(t) \\ &= t^{4\rho}\Big[(-2\theta\varepsilon(t) + c(t))^2\beta^2(t) + 2\theta t^r c(t)\beta(t)\dot{\beta}(t) + \theta^2 t^{2r}(\dot{\beta}(t))^2 + 4\theta(2r\theta t^{r-1} - 1)\varepsilon(t)\beta^2(t)\Big]. \end{aligned}$$

Now, according to the growth condition (14), there exists $\delta > 0$ such that

$$\frac{t^r\dot{\beta}(t)}{\beta(t)} \leq \frac{1}{\theta} - 2rt^{r-1} - \delta \quad \forall t \geq t_0. \tag{35}$$

This entails

$$\theta t^r\beta(t)\dot{\beta}(t) \leq \Big((1 - 2r\theta t^{r-1}) - \delta\theta\Big)\beta^2(t) \quad \text{and} \quad \theta^2 t^{2r}(\dot{\beta}(t))^2 \leq \Big((1 - 2r\theta t^{r-1}) - \delta\theta\Big)^2\beta^2(t). \tag{36}$$

Since $c(t) \geq 0$, we now arrive at

$$\begin{aligned} B^2 - AC &\leq t^{4\rho}\beta^2(t)\Big[(-2\theta\varepsilon(t) + c(t))^2 + 2c(t)\Big((1 - 2r\theta t^{r-1}) - \delta\theta\Big) \\ &\qquad + \Big((1 - 2r\theta t^{r-1}) - \delta\theta\Big)^2 + 4\theta(2r\theta t^{r-1} - 1)\varepsilon(t)\Big] \\ &= t^{4\rho}\beta^2(t)\left\{4\theta^2\varepsilon^2(t) - 4\theta\left(1 - 2r\theta t^{r-1} + c(t)\right)\varepsilon(t) + \Big[c(t) + \Big(\Big(1 - 2r\theta t^{r-1}\Big) - \delta\theta\Big)\Big]^2\right\} \\ &= t^{4\rho}\beta^2(t)\pi(t), \end{aligned} \tag{37}$$

where $\pi(t)$ is the term between curly brackets. We will analyze $\pi(t)$ separating the cases $r \in [0, 1)$ and $r = 1$.

Case $r \in [0, 1)$: define

$$c_r := \theta\alpha - 2, \quad \varepsilon_r := \alpha - \lambda, \quad \mu(t) := 4\theta^2 t^2 - 4\theta\left(c_r + 1\right)t + \left(c_r + 1 - \delta\theta\right)^2.$$

Notice that $c(t) \to c_r$, $\varepsilon(t) \to \varepsilon_r$ and, therefore, $\pi(t) \to \mu(\varepsilon_r)$ as $t \to +\infty$. The discriminant of the quadratic equation $\mu(t) = 0$ reads

$$\Delta = 16\theta^2 \left[(c_r + 1)^2 - (c_r + 1 - \delta\theta)^2 \right].$$

Since $0 < 1 - \delta\theta$, we have $c_r + 1 > c_r + 1 - \delta\theta > 0$ and thus $\Delta > 0$. This means that μ has two distinct roots: by setting $\tilde{\Delta} := \frac{1}{8\theta^2}\sqrt{\Delta}$, these are given by

$$\underline{\varepsilon} := \frac{c_r + 1}{2\theta} - \tilde{\Delta}, \quad \overline{\varepsilon} := \frac{c_r + 1}{2\theta} + \tilde{\Delta}.$$

The midpoint between the two roots is given by $\varepsilon_{r,1} := \frac{c_r+1}{2\theta}$, which fulfills

$$0 < \varepsilon_{r,1} = \frac{c_r + 1}{2\theta} = \frac{\theta\alpha - 1}{2\theta} = \frac{\alpha}{2} - \frac{1}{2\theta} < \alpha.$$

Choose $\varepsilon_{r,2} > 0$ such that $\varepsilon_{r,1} - \tilde{\Delta} < \varepsilon_{r,2} < \varepsilon_{r,1}$. Set $\lambda_{r,i} := \alpha - \varepsilon_{r,i}$, $i = 1, 2$. As we remarked previously, we have $\pi(t) \to \mu(\varepsilon_{r,i})$ as $t \to +\infty$. Therefore, for large enough t, we have

$$\pi(t) < \frac{\mu(\varepsilon_{r,i})}{2} = \frac{\mu(\alpha - \lambda_{r,i})}{2} < 0 \quad \text{for } i = 1, 2.$$

Going back to (37), this means that

$$B^2 - AC \le t^{4\rho}\beta^2(t)\frac{\mu(\alpha - \lambda_{r,i})}{2} < 0 \quad \text{for large enough } t \text{ and } i = 1, 2.$$

Case $r = 1$: here, we have

$$\varepsilon(t) \equiv \varepsilon := \alpha - 1 - \lambda, \quad \text{and} \quad c(t) \equiv c := \theta\alpha + \theta - 2.$$

In this case, we define

$$\mu(t) := 4\theta^2 t^2 - 4\theta\,(1 - 2\theta + c)\,t + (c + 1 - 2\theta - \delta\theta)^2.$$

Notice that with this definition we have $\pi(t) \equiv \mu(\varepsilon)$. The discriminant of the quadratic equation $p(t) = 0$ reads

$$\Delta = 16\theta^2 \left[(c + 1 - 2\theta))^2 - (c + 1 - 2\theta - \delta\theta))^2 \right].$$

Since $0 < 1 - 2\theta - \delta\theta$, we have $c + 1 - 2\theta > c + 1 - 2\theta - \delta\theta > 0$ and thus $\Delta > 0$. Again, define $\tilde{\Delta}$ as before. The roots of μ are given by

$$\underline{\varepsilon} := \frac{c + 1 - 2\theta}{2\theta} - \tilde{\Delta}, \quad \overline{\varepsilon} := \frac{c + 1 - 2\theta}{2\theta} + \tilde{\Delta}.$$

The midpoint between the roots is $\varepsilon_{1,1} := \frac{c+1-2\theta}{2\theta}$, which satisfies

$$0 < \varepsilon_{1,1} = \frac{c + 1 - 2\theta}{2\theta} = \frac{\theta\alpha - \theta - 1}{2\theta} = \frac{\alpha - 1}{2} - \frac{1}{2\theta} < \alpha - 1.$$

Similar as before, choose $\varepsilon_{1,2} > 0$ such that $\varepsilon_{1,1} - \tilde{\Delta} < \varepsilon_{1,2} < \varepsilon_{1,1}$. Set $\lambda_{1,i} := \alpha - 1 - \varepsilon_{1,i}$, $i = 1, 2$. Since $\mu(\alpha - 1 - \lambda_{1,i}) = \mu(\varepsilon_{1,i}) < 0$ for $i = 1, 2$, going back to (37) gives

$$B^2 - AC \le t^{4\rho}\beta^2(t)\mu(\alpha - 1 - \lambda_{1,i}) < 0 \quad \text{for large enough } t \text{ and } i = 1, 2.$$

Summarizing, for the choices $\lambda = \lambda_{r,i}$, $r \in [0,1]$, $i = 1,2$ it holds, recalling lines (29)-(33), together with inequality (34),

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_{\lambda_{r,i}}(t) = {} & 2\lambda_{r,i}t^{2(\rho-r)-1}\Big[2\alpha(\rho-r) - (2\rho-r)(2\rho-r-1)t^{r-1}\Big]\|z(t)-z_*\|^2 \\ & + 4\lambda_{r,i}t^{2\rho-r}\Big[(2r\theta t^{r-1}-1)\beta(t) + \theta t^r\dot{\beta}(t)\Big]\langle z(t)-z_*, V(z(t))\rangle \\ & + t^{2\rho-r}\left(\lambda_{r,i} + \frac{\rho+r}{2}t^{r-1} - \alpha\right)\|\dot{z}(t)\|^2 \\ & + \frac{2}{3}\theta t^{2\rho+r}\beta(t)\Big[(2r\theta t^{r-1}-1)\beta(t) + \theta t^r\dot{\beta}(t)\Big]\|V(z(t))\|^2 \\ \leq {} & 0, \end{aligned} \tag{38}$$

where t is taken large enough, say $t \geq T \geq t_0$. This means that $\mathcal{E}_{\lambda_{r,i}}$ monotonically decreases on $[T,+\infty)$, and thus

$$0 \leq \mathcal{E}_{\lambda_{r,i}}(t) \leq \mathcal{E}_{\lambda_{r,i}}(T) \text{ for } t \geq T,\ r \in [0,1] \text{ and } i = 1,2. \tag{39}$$

Deducing convergence rates from the estimates. We will now proceed by distinguishing between the cases $r \in (0,1)$, $r = 0$ and $r = 1$.

Case $r \in (0,1)$: according to the definition of the energy function (17)-(20) and (39), for $t \geq T$ we have

$$\langle z(t)-z_*, V(z(t))\rangle \leq \frac{\mathcal{E}_{\lambda_{r,1}}(T)}{2\lambda_{r,1}\theta}\cdot\frac{1}{t^{2\rho}\beta(t)} \quad \text{and} \quad \|V(z(t))\| \leq \frac{\sqrt{2\mathcal{E}_{\lambda_{r,1}}(T)}}{\theta}\cdot\frac{1}{t^{\rho+r}\beta(t)}. \tag{40}$$

Furthemore, say that for $t \geq T$ we have $(2\rho - r)t^{r-1} < \xi < \alpha - \lambda_{r,1}$. Going back to (18), this means that for $t \geq T$ it holds

$$2\lambda_{r,1}t^{2(\rho-r)}(\alpha-\xi-\lambda_{r,1})\|z(t)-z_*\|^2 \leq 2\lambda_{r,1}t^{2(\rho-r)}(\alpha-(2\rho-r)t^{r-1})-\lambda_{r,1})\|z(t)-z_*\|^2 \leq \mathcal{E}_{\lambda_{r,1}}(T)$$

and thus

$$2\lambda_{r,1}t^{\rho-r}\|z(t)-z_*\| \leq \sqrt{\frac{2\mathcal{E}_{\lambda_{r,1}}(T)}{(\alpha-\xi-\lambda_{r,1})}}.$$

Combining this together with (17), (39) and (40) yields

$$\begin{aligned} 2t^\rho\|\dot{z}(t)\| & \leq 2\lambda_{r,1}t^{\rho-r}\|z(t)-z_*\| + \theta t^{\rho+r}\beta(t)\|V(z(t))\| \\ & \quad + \Big\|2\lambda_{r,1}t^{\rho-r}(z(t)-z_*) + 2t^\rho\dot{z}(t) + \theta t^{\rho+r}\beta(t)V(z(t))\Big\| \\ & \leq \sqrt{\frac{2\mathcal{E}_{\lambda_{r,1}}(T)}{(\alpha-\xi-\lambda_{r,1})}} + 2\sqrt{2\mathcal{E}_{\lambda_{r,1}}(T)}. \end{aligned}$$

The previous inequality, together with (40) and the fact that $\rho \in (0,r)$ was arbitrary, tells us that so far we have, as $t \to +\infty$,

$$\|V(z(t))\| = o\left(\frac{1}{t^{\rho+r}\beta(t)}\right), \quad \langle z(t)-z_*, V(z(t))\rangle = o\left(\frac{1}{t^{2\rho}\beta(t)}\right), \quad \|\dot{z}(t)\| = o\left(\frac{1}{t^\rho}\right). \tag{41}$$

Case $r = 0$ or $r = 1$: going back to (18), we have, for $t \geq T$,

$$\|z(t)-z_*\| \leq \sqrt{\frac{\mathcal{E}_{\lambda_{0,1}}(T)}{2\lambda_{0,1}(\alpha-\lambda_{0,1})}} \text{ (when } r=0) \text{ or } \|z(t)-z_*\| \leq \sqrt{\frac{\mathcal{E}_{\lambda_{1,1}}(T)}{2\lambda_{1,1}(\alpha-1-\lambda_{1,1})}} \text{ (when } r=1),$$

which gives the boundedness of $t \mapsto z(t)$. Using the boundedness of $\mathcal{E}_{\lambda_{0,1}}$ (when $r = 0$) and that of $\mathcal{E}_{\lambda_{1,1}}$ (when $r = 1$) and arguing exactly as in the case $r \in (0, 1)$, we obtain

$$\|V(z(t))\| = \mathcal{O}\left(\frac{1}{t^{2r}\beta(t)}\right), \quad \langle z(t) - z_*, V(z(t))\rangle = \mathcal{O}\left(\frac{1}{t^{2r}\beta(t)}\right), \quad \|\dot{z}(t)\| = \mathcal{O}\left(\frac{1}{t^{r}}\right). \tag{42}$$

as $t \to +\infty$ for $r = 0$ and $r = 1$.

Weak convergence of the trajectories. In the following steps, we will assume that $t \mapsto z(t)$ is bounded when $r \in (0, 1)$. This will allow us to set $\rho = r$ in (41) and to show weak convergence of the trajectories towards a zero of V. Additionally, we will improve $\mathcal{O}$ to o in (42) and we will again show weak convergence of the trajectories towards a zero of V.

Taking $\rho \in (0, r)$ when $r \in (0, 1)$ is done only in order to obtain that the coefficient accompanying $\|z(t) - z_*\|^2$ eventually becomes nonpositive, therefore producing $\frac{d}{dt}\mathcal{E}_\lambda(t) \leq 0$ (recall line (22)). Setting $\rho = r$ creates no other issues otherwise, but changes the bound for $\frac{d}{dt}\mathcal{E}_\lambda(t)$: indeed, going back to (38) and plugging $\rho = r$ gives, for $t \geq T$ and $i = 1, 2$,

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_{\lambda_{r,i}}(t) &= 2\lambda_{r,i} r(1-r)t^{r-2}\|z(t) - z_*\|^2 \\
&\quad + 4\lambda_{r,i} t^r \left[(2r\theta t^{r-1} - 1)\beta(t) + \theta t^r \dot{\beta}(t)\right]\langle z(t) - z_*, V(z(t))\rangle \\
&\quad + t^r\left(\lambda_{r,i} + rt^{r-1} - \alpha\right)\|\dot{z}(t)\|^2 \\
&\quad + \frac{2}{3}\theta t^{3r}\beta(t)\left[(2r\theta t^{r-1} - 1)\beta(t) + \theta t^r \dot{\beta}(t)\right]\|V(z(t))\|^2 \\
&\leq 2\lambda_{r,i} r(1-r)t^{r-2}\|z(t) - z_*\|^2.
\end{aligned} \tag{43}$$

We have already established the boundedness of $t \mapsto z(t)$ when $r = 0$ or $r = 1$, and we will assume it when $r \in (0, 1)$. Say that $\|z(t) - z_*\|^2 \leq M$ for all $t \geq T$ for some $M \geq 0$. According to (43), for $t \geq T$ we now have

$$\frac{d}{dt}\mathcal{E}_{\lambda_{r,i}}(t) \leq 2\lambda_{r,i} r(1-r)Mt^{r-2}$$

and therefore

$$\frac{d}{dt}\left[\mathcal{E}_{\lambda_{r,i}}(t) + 2\lambda_{r,i} rMt^{r-1}\right] = \frac{d}{dt}\mathcal{E}_{\lambda_{r,i}}(t) + 2\lambda_{r,i} r(r-1)Mt^{r-2} \leq 0.$$

As a consequence of this, $t \mapsto \mathcal{E}_{\lambda_{r,i}}(t) + 2\lambda_{r,i} rMt^{r-1}$ is nonnegative and monotonically decreasing on $[T, +\infty)$, which means it has a limit as $t \to +\infty$. Since $2\lambda_{r,i} r(1-r)Mt^{r-1}$ also has a limit as $t \to +\infty$, we come to the following: for $\rho = r$, we have

$$0 \leq \mathcal{E}_{\lambda_{r,i}}(t) \leq \mathcal{E}_{\lambda_{r,i}}(T) + 2\lambda_{r,i} r(1-r)MT^{r-1} \text{ for } t \geq T,\ r \in [0, 1] \text{ and } i = 1, 2; \tag{44}$$

$$\lim_{t\to+\infty} \mathcal{E}_{\lambda_{r,i}}(t) \text{ exists for } r \in [0, 1] \text{ and } i = 1, 2. \tag{45}$$

By going back to lines (17)-(20), setting $\rho = r$ and reasoning exactly as we did before, we arrive at

$$\|V(z(t))\| = \mathcal{O}\left(\frac{1}{t^{2r}\beta(t)}\right), \quad \langle z(t) - z_*, V(z(t))\rangle = \mathcal{O}\left(\frac{1}{t^{2r}\beta(t)}\right), \quad \|\dot{z}(t)\| = \mathcal{O}\left(\frac{1}{t^{r}}\right). \tag{46}$$

as $t \to +\infty$, where these rates now hold for all $r \in [0,1]$.

For $r \in [0,1]$, we proceed to show three integrability results that we will need later. Recall (35): for $t \geq T$, we have

$$\frac{t^r \dot{\beta(t)}}{\beta(t)} \leq \frac{1}{\theta} - 2rt^{r-1} - \delta \;\Rightarrow\; \theta t^r \dot{\beta}(t) \leq (1 - 2r\theta t^{r-1})\beta(t) - \delta\theta\beta(t)$$
$$\Rightarrow\; \delta\theta\beta(t) \leq (1 - 2r\theta t^{r-1})\beta(t) - \theta t^r \dot{\beta}(t).$$

Going back to (43), we choose $i = 1$ and integrate the inequality from T to $t \geq T$:

$$\begin{aligned}
&\int_T^t s^r(\alpha - rs^{r-1} - \lambda_{r,1})\left\|\dot{z}(s)\right\|^2 ds + 4\lambda_{r,1}\delta\theta \int_T^t s^r \beta(s)\langle z(s) - z_*, V(z(s))\rangle ds \\
&+ \frac{2}{3}\delta\theta^2 \int_{t_5}^t s^{3r}\beta^2(s)\|V(z(s))\|^2 ds \\
\leq &\int_T^t s^r(\alpha - rs^{r-1} - \lambda_{r,1})\left\|\dot{z}(s)\right\|^2 ds \\
&+ 4\lambda_{r,1}\int_T^t s^r\Big[(1 - 2r\theta s^{r-1})\beta(s) - \theta s^r \dot{\beta}(s)\Big]\langle z(s) - z_*, V(z(s))\rangle ds \\
&+ \frac{2}{3}\theta\int_T^t s^{3r}\beta(s)\Big[(1 - 2r\theta s^{r-1})\beta(s) - \theta s^r\dot{\beta}(s)\Big]\|V(z(s))\|^2 ds \\
\leq &-\int_T^t \frac{d}{ds}\mathcal{E}_{\lambda_{r,1}}(s)ds + 2\lambda_{r,1}r(1-r)\int_{t_5}^t s^{r-2}\|z(s) - z_*\|^2 ds \\
\leq &\,\mathcal{E}_{\lambda_{r,1}}(T) + 2\lambda_{r,1}r(1-r)\int_T^t s^{r-2}\|z(s) - z_*\|^2 ds,
\end{aligned}$$

where in the last inequality we drop the nonpositive term $-\mathcal{E}_{\lambda_{r,1}}(t)$. The second summand in the last line vanishes in the cases $r = 0$ and $r = 1$, and in the case $r \in (0,1)$ we are assuming the boundedness of the trajectory in order for this integral to be finite as $t \to +\infty$. This produces the following statements for $r \in [0,1]$:

$$\int_T^{+\infty} t^r\left\|\dot{z}(t)\right\|^2 dt < +\infty, \tag{47}$$

$$\int_T^{+\infty} t^r\beta(t)\langle z(t) - z_*, V(z(t))\rangle dt < +\infty, \tag{48}$$

$$\int_T^{+\infty} t^{3r}\beta^2(t)\|V(z(t))\|^2 dt < +\infty. \tag{49}$$

Now, we proceed to show the o rates. According to the formula for the energy function, for $\rho = r$, $r \in [0,1]$ and $t \geq T$ we have

$$\begin{aligned}
\mathcal{E}_{\lambda_{r,2}}(t) - \mathcal{E}_{\lambda_{r,1}}(t) = \frac{1}{2}\Bigg[&\left\|2\lambda_{r,2}(z(t) - z_*) + t^r\Big[2\dot{z}(t) + \theta t^r\beta(t)V(z(t))\Big]\right\|^2 \\
&- \left\|2\lambda_{r,1}(z(t) - z_*) + t^r\Big[2\dot{z}(t) + \theta t^r\beta(t)V(z(t))\Big]\right\|^2\Bigg] \\
&+ 2\Big[\lambda_{r,2}(\alpha - rt^{r-1} - \lambda_{r,2}) - \lambda_{r,1}(\alpha - rt^{r-1} - \lambda_{r,1})\Big]\|z(t) - z_*\|^2 \\
&+ 2(\lambda_{r,2} - \lambda_{r,1})\theta t^{2r}\beta(t)\langle z(t) - z_*, V(z(t))\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\left[4(\lambda_{r,2}^2 - \lambda_{r,1}^2)\|z(t)-z_*\|^2 + 4(\lambda_{r,2}-\lambda_{r,1})\langle z(t)-z_*, 2\dot{z}(t) + \theta t^r \beta(t) v(z(t))\rangle\right] \\
&\quad + \left[2(\lambda_{r,2}-\lambda_{r,1})(\alpha - rt^{r-1}) - 2(\lambda_{r,2}^2 - \lambda_{r,1}^2)\right]\|z(t)-z_*\|^2 \\
&\quad + 2(\lambda_{r,2}-\lambda_{r,1})\theta t^{2r}\beta(t)\langle z(t)-z_*, V(z(t))\rangle \\
&= 4(\lambda_{r,2}-\lambda_{r,1})\left[\frac{1}{2}(\alpha - rt^{r-1})\|z(t)-z_*\|^2 + t^r\langle z(t)-z_*, \dot{z}(t) + \theta t^r \beta(t) V(z(t))\rangle\right].
\end{aligned} \tag{50}
$$

Define, for $t \geq T$,

$$
p(t) := \frac{1}{2}(\alpha - rt^{r-1})\|z(t)-z_*\|^2 + t^r\langle z(t)-z_*, \dot{z}(t) + \theta t^r \beta(t) V(z(t))\rangle. \tag{51}
$$

Since we already established that $\lim_{t\to+\infty} \mathcal{E}_{\lambda_{r,i}}$ exists for $i = 1, 2$, and $\lambda_{r,2} - \lambda_{r,1} \neq 0$, we obtain the existence of $\lim_{t\to+\infty} p(t)$. We may express $\mathcal{E}_{\lambda_{r,1}}(t)$ as

$$
\begin{aligned}
\mathcal{E}_{\lambda_{r,1}}(t) &= \frac{1}{2}\left\|2\lambda_{r,1}(z(t)-z_*) + t^r\left[2\dot{z}(t) + \theta t^r\beta(t)V(z(t))\right]\right\|^2 \\
&\quad + 2\lambda_{r,1}(\alpha - rt^{r-1} - \lambda_{r,1})\|z(t)-z_*\|^2 \\
&\quad + 2\lambda_{r,1}\theta t^{2r}\beta(t)\langle z(t)-z_*, V(z(t))\rangle + \frac{\theta^2}{2}t^{4r}\beta^2(t)\|V(z(t))\|^2 \\
&= \frac{1}{2}\Big[4\lambda_{r,1}^2\|z(t)-z_*\|^2 + 2\lambda_{r,1}t^r\langle z(t)-z_*, 2\dot{z}(t) + \theta t^r\beta(t)V(z(t))\rangle \\
&\qquad + t^{2r}\left\|2\dot{z}(t) + \theta t^r\beta(t)V(z(t))\right\|^2\Big] \\
&\quad + 2\lambda_{r,1}(\alpha - rt^{r-1})\|z(t)-z_*\|^2 - 2\lambda_{r,1}^2\|z(t)-z_*\|^2 + 2\lambda_{r,1}\theta t^{2r}\beta(t)\langle z(t)-z_*, V(z(t))\rangle \\
&\quad + \frac{\theta^2}{2}t^{4r}\beta^2(t)\|V(z(t))\|^2 \\
&= 4\lambda_{r,1}p(t) + \frac{t^{2r}}{2}\left\|2\dot{z}(t) + \theta t^r\beta(t)V(z(t))\right\|^2 + \frac{\theta^2}{2}t^{4r}\beta^2(t)\|V(z(t))\|^2 \\
&= 4\lambda_{r,1}p(t) + t^{2r}\left\|\dot{z}(t) + \theta t^r\beta(t)V(z(t))\right\|^2 + t^{2r}\|\dot{z}(t)\|^2.
\end{aligned}
$$

Define $h(t) := t^{2r}\|\dot{z}(t) + \theta t^r\beta(t)V(z(t))\|^2 + t^{2r}\|\dot{z}(t)\|^2$. Since both $\lim_{t\to+\infty} \mathcal{E}_{\lambda_{r,1}}(t)$ and $\lim_{t\to+\infty} p(t)$ exists, so does $\lim_{t\to+\infty} h(t)$. Observe that

$$
\int_T^t \frac{1}{s^r}h(s)ds \leq 3\int_T^t s^r\|\dot{z}(s)\|^2 ds + 2\theta^2\int_T^t s^{3r}\beta^2(s)\|V(z(s))\|^2 ds,
$$

and, according to (47) and (49), the integrals on the right-hand side become finite as $t \to +\infty$. Therefore, $\int_T^{+\infty} \frac{1}{t^r}h(t)dt < +\infty$, which gives

$$
\lim_{t\to+\infty} h(t) = 0,
$$

which in particular yields

$$
\lim_{t\to+\infty} t^{2r}\|\dot{z}(t)\|^2 = 0, \text{ equivalently } \|\dot{z}(t)\| = o\left(\frac{1}{t^r}\right) \text{ as } t \to +\infty.
$$

and plugging this into $\lim_{t\to+\infty} t^{2r}\left\|\dot{z}(t)+\theta t^r\beta(t)V(z(t))\right\|^2=0$ produces

$$\lim_{t\to+\infty} t^{4r}\beta^2(t)\|V(z(t))\|^2=0,\ \text{equivalently}\ \|V(z(t))\|=o\left(\frac{1}{t^{2r}\beta(t)}\right)\ \text{as}\ t\to+\infty.$$

Using the previous result together with the boundedness of $\|z(t)-z_*\|$ and the Cauchy-Schwarz inequality yields

$$\langle z(t)-z_*, V(z(t))\rangle = o\left(\frac{1}{t^{2r}\beta(t)}\right)\ \text{as}\ t\to+\infty.$$

as $t\to+\infty$. We have now shown statements (i) and (ii) of the theorem.

We now focus on the weak convergence of the trajectories. For this end, we will make use of Opial's lemma (see Lemma A.5). Define, for $t\geq T$,

$$q(t):=\frac{1}{2}\|z(t)-z_*\|^2+\theta\int_T^t s^r\beta(s)\langle z(s)-z_*, V(z(s))\rangle ds.$$

We distinguish between the cases $r\in[0,1)$ and $r=1$.

Case $r\in[0,1)$: recalling formula (51), notice that

$$\begin{aligned}
&\alpha q(t)+t^r\dot{q}(t)\\
&=\frac{\alpha}{2}\|z(t)-z_*\|^2+t^r\langle z(t)-z_*, \dot{z}(t)+\theta t^r\beta(t)V(z(t))\rangle+\theta\alpha\int_T^t s^r\beta(s)\langle z(s)-z_*, V(z(s))\rangle ds\\
&=\frac{r}{2}t^{r-1}\|z(t)-z_*\|^2+p(t)+\theta\alpha\int_T^t s^r\beta(s)\langle z(s)-z_*, V(z(s))\rangle ds.
\end{aligned}$$

Because of (48), the integral in the previous sum converges as $t\to+\infty$. We have already established the existence of $\lim_{t\to+\infty}p(t)$, and using that $t^{r-1}\to0$ as $t\to+\infty$ and the boundedness of $\|z(t)-z_*\|$ allows us to deduce the existence of $\lim_{t\to+\infty}\alpha q(t)+t^r\dot{q}(t)$. Using Lemma A.3, we obtain the existence of $\lim_{t\to+\infty}q(t)$, and again using that $\int_T^t s^r\beta(s)\langle z(s)-z_*, V(z(s))\rangle ds$ has a limit as $t\to+\infty$, we come to the existence of

$$\lim_{t\to+\infty}\|z(t)-z_*\|.$$

Case $r=1$: we repeat the reasoning of the previous case. We obtain the existence of $\lim_{t\to+\infty}(\alpha-1)q(t)+t\dot{q}(t)$, and once again we use Lemma A.3 to produce the existence of $\lim_{t\to+\infty}\|z(t)-z_*\|$.

In both cases, we have verified the first condition of Opial's Lemma. For the second condition, let $\overline{z}$ be a sequential cluster point of the trajectory $z(t)$ as $t\to+\infty$, which means there exists a sequence $(t_n)_{n\in\mathbb{N}}\subseteq[t_0,+\infty)$ such that $t_n\to+\infty$ as $n\to+\infty$ and

$$z(t_n)\rightharpoonup\overline{z}\quad\text{as}\quad n\to+\infty,$$

where $\rightharpoonup$ denotes weak convergence in $\mathcal{H}$. The convergence rate attained for $\|V(z(t))\|$ ensures $V(z(t_n))\to0$ as $n\to+\infty$. Since the graph of V is sequentially closed in $\mathcal{H}^{\text{weak}}\times\mathcal{H}^{\text{strong}}$, this finally allows us to conclude that

$$V(\overline{z})=0,$$

thereby fulfilling the second condition of Opial's Lemma. We have thus concluded the proof of this theorem. □

2.2 Linearly constrained convex minimization as a particular case

Recall the problem stated in the introduction

$$\begin{array}{rl} \min & f(x), \\ \text{subject to} & Ax = b, \end{array} \tag{52}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces, $b \in \mathcal{Y}$, $A : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator and $f : \mathcal{X} \to \mathbb{R}$ is a convex and continuously differentiable function. Formulating (13) in terms of the operator V given by

$$V(x, \lambda) = \Big(\nabla f(x) + A^* \lambda, b - Ax\Big) \tag{53}$$

yields the primal-dual system

$$\begin{cases} \ddot{x}(t) + \dfrac{\alpha}{t^r}\dot{x}(t) + \theta t^r \beta(t) \dfrac{d}{dt}\nabla f(x(t)) + \beta(t) A^*(\lambda(t) + \theta t^r \dot{\lambda}(t)) + \beta(t)\nabla f(x(t)) & = 0, \\ \ddot{\lambda}(t) + \dfrac{\alpha}{t^r}\dot{\lambda}(t) - \beta(t)\Big[A(\dot{x}(t) + \theta t^r \dot{x}(t)) - b\Big] & = 0. \end{cases} \tag{54}$$

Let $(x_*, \lambda_*) \in \mathcal{X} \times \mathcal{Y}$ be a primal-dual solution to (52). Equivalently, (x_*, λ_*) is zero of V, or a saddle point of the Lagrangian $\mathcal{L}(x, \lambda) = f(x) + \langle \lambda, Ax - b\rangle$. Let $t \mapsto (x(t), \lambda(t))$ be a solution to (54). Using the gradient inequality for convex functions, we have, for every $t \geq t_0$,

$$\begin{aligned} 0 &\leq \mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) \\ &= f(x(t)) - f(x_*) + \langle \lambda_*, Ax(t) - b\rangle \\ &\leq \langle x(t) - x_*, \nabla f(x(t))\rangle + \langle \lambda_*, Ax(t) - b\rangle \\ &= \langle x(t) - x_*, \nabla f(x(t))\rangle + \langle x(t) - x_*, A^*\lambda(t)\rangle + \langle \lambda(t) - \lambda_*, b - Ax(t)\rangle \\ &= \Big\langle (x(t), \lambda(t)) - (x_*, \lambda_*), V(x(t), \lambda(t))\Big\rangle. \end{aligned} \tag{55}$$

By exploiting the results in Theorem 2.1, and by taking into consideration additional potential information in this particular case, we come to obtain the following result.

Theorem 2.2. *Suppose that $\alpha, \theta > 0$, $r \in [0, 1]$ and that $\beta : [t_0, +\infty) \to (0, +\infty)$ satisfy the same assumptions as in Theorem 2.1. Let $(x_*, \lambda_*) \in \mathcal{X} \times \mathcal{Y}$ be a primal-dual solution to (52) and $t \mapsto (x(t), \lambda(t))$ a solution to (54). Consider the following convergence rates as $t \to +\infty$:*

$$\mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) = o\left(\frac{1}{t^{2\rho}\beta(t)}\right), \quad |f(x(t)) - f(x_*)| = o\left(\frac{1}{t^{2\rho}\beta(t)}\right),$$
$$\|Ax(t) - b\| = o\left(\frac{1}{t^{\rho+r}\beta(t)}\right), \quad \|\dot{x}(t)\| = o\left(\frac{1}{t^{\rho}}\right), \quad \|\dot{\lambda}(t)\| = o\left(\frac{1}{t^{\rho}}\right).$$

Additionally, if ∇f is L-Lipschitz continuous for some $L > 0$, consider the following rates as $t \to +\infty$:

$$\|\nabla f(x(t)) - \nabla f(x_*)\| = o\left(\frac{1}{t^{\rho}\sqrt{\beta(t)}}\right), \quad \|A^*(\lambda(t) - \lambda_*)\| = o\left(\frac{1}{t^{\rho}\sqrt{\beta(t)}}\right).$$

The following statements are true:

(i) *If $r \in (0, 1)$, the above rates hold for $\rho \in (0, r)$. Furthermore, if $t \mapsto (x(t), \lambda(t))$ is bounded, then these rates hold for $\rho = r$.*

(ii) *if* $r = 0$ *or* $r = 1$*, then* $t \mapsto (x(t), \lambda(t))$ *is bounded; moreover, the above rates hold for* $\rho = 0$ *(if* $r = 0$*) and for* $\rho = 1$ *(if* $r = 1$*).*

(iii) *If* $r = 0$ *or* $r = 1$ *or* $\Big(r \in (0, 1)$ *and* $t \mapsto (x(t), \lambda(t))$ *is bounded*$\Big)$*, then* $(x(t), \lambda(t))$ *converges weakly to a primal-dual solution to* (52) *as* $t \to +\infty$*.*

Proof. Parts (i) and (ii) are a direct consequence of Theorem 2.1 applied to the operator V defined in (53). Indeed, endow $\mathcal{X} \times \mathcal{Y}$ with the norm $\|(x, \lambda)\| := \|x\| + \|\lambda\|$. First, from (21) we obtain

$$\|\dot{x}(t)\| + \|\dot{\lambda}(t)\| = \|(\dot{x}(t), \dot{\lambda}(t))\| = o\left(\frac{1}{t^\rho}\right) \quad \text{as } t \to +\infty.$$

From (55) and (21) we get

$$0 \leq \mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) \leq \Big\langle (x(t), \lambda(t)) - (x_*, \lambda_*), V(x(t), \lambda(t)) \Big\rangle = o\left(\frac{1}{t^{2\rho}\beta(t)}\right) \quad \text{as } t \to +\infty. \tag{56}$$

Again using (21), we produce

$$\|\nabla f(x(t)) + A^*\lambda(t)\| + \|Ax(t) - b\| = \|V(x(t), \lambda(t))\| = o\left(\frac{1}{t^{\rho+r}\beta(t)}\right) \tag{57}$$

and thus

$$\|Ax(t) - b\| = o\left(\frac{1}{t^{\rho+r}\beta(t)}\right) \quad \text{as } t \to +\infty. \tag{58}$$

Combining (56) and (58) we come to

$$\begin{aligned} |f(x(t)) - f(x_*)| &\leq |f(x(t)) - f(x_*) + \langle \lambda_*, Ax(t) - b\rangle| + |\langle \lambda_*, Ax(t) - b\rangle| \\ &\leq \mathcal{L}(x(t), \lambda_*) - \mathcal{L}(x_*, \lambda(t)) + \|\lambda_*\|\|Ax(t) - b\| \\ &= o\left(\frac{1}{t^{2\rho}\beta(t)}\right) \quad \text{as } t \to +\infty, \end{aligned} \tag{59}$$

since we always have $2\rho \leq \rho + r$. Assume now that ∇f is L-Lipschitz continuous. We have now a stronger gradient inequality

$$0 \leq \frac{1}{2L}\|\nabla f(x(t)) - \nabla f(x_*)\|^2 \leq \langle x(t) - x_*, \nabla f(x(t))\rangle - (f(x(t)) - f(x_*)). \tag{60}$$

We will show that $\langle x(t) - x_*, \nabla f(x(t))\rangle = o\left(\frac{1}{t^{2\rho}\beta(t)}\right)$ as $t \to +\infty$. Indeed, we have

$$\begin{aligned} &\langle x(t) - x_*, \nabla f(x(t))\rangle + \langle \lambda_*, Ax(t) - b\rangle \\ &= \langle x(t) - x_*, \nabla f(x(t))\rangle + \langle x(t) - x_*, A^*\lambda(t)\rangle + \langle \lambda(t) - \lambda_*, b - Ax(t)\rangle \\ &= \Big\langle (x(t), \lambda(t)) - (x_*, \lambda_*), V(x(t), \lambda(t)) \Big\rangle \\ &= o\left(\frac{1}{t^{2\rho}\beta(t)}\right) \quad \text{as } t \to +\infty, \end{aligned}$$

so using this together with (58) yields $\langle x(t) - x_*, \nabla f(x(t))\rangle = o\left(\frac{1}{t^{2\rho}\beta(t)}\right)$ as $t \to +\infty$ (again, recall that $2\rho \leq \rho + r$). Plugging this rate into (60) yields

$$\|\nabla f(x(t)) - \nabla f(x_*)\| = o\left(\frac{1}{t^\rho\sqrt{\beta(t)}}\right) \quad \text{as } t \to +\infty. \tag{61}$$

Now,

$$\begin{aligned}\|A^*(\lambda(t)-\lambda_*)\| &\le \|\nabla f(x(t))-\nabla f(x_*)+A^*\lambda(t)-A^*\lambda_*\|+\|\nabla f(x(t))-\nabla f(x_*)\|\\ &= \|\nabla f(x(t))+A^*\lambda(t)\|+\|\nabla f(x(t))-\nabla f(x_*)\|.\end{aligned}$$

According to (57), the left summand is of order $o\left(\frac{1}{t^{\rho+r}\beta(t)}\right)$ as $t\to+\infty$; according to (61), the right summand is of order $o\left(\frac{1}{t^{\rho}\sqrt{\beta(t)}}\right)$ as $t\to+\infty$. Since β is nondecreasing on $[t_0,+\infty)$, we have $\sqrt{\beta(t)}=\frac{\beta(t)}{\sqrt{\beta(t)}}\le\frac{\beta(t)}{\sqrt{\beta(t_0)}}$. Since $t^{\rho}\le t^{\rho+r}$ for large t, this allows us to write the inequality

$$\frac{\sqrt{\beta(t_0)}}{t^{\rho+r}\beta(t)}\le\frac{1}{t^{\rho}\sqrt{\beta(t)}}\quad\text{for large enough } t,$$

from which we deduce that

$$\|A^*(\lambda(t)-\lambda_*)\|=o\left(\frac{1}{t^{\rho}\sqrt{\beta(t)}}\right)\quad\text{as } t\to+\infty.$$

We have thus shown parts (i) and (ii) of this theorem. Part (iii), i.e., the weak convergence of the trajectories to a primal-dual solution to (52), is again a direct corollary of part (iii) of Theorem 2.1. □

Remark 2.3. The system (54) resembles the one studied by He et al. in [15], for the case where $r=s$. There, the authors consider an extrapolation parameter of the form θt^s, where $s\in[r,1]$. The only difference between our system and theirs lies in the inclusion of the Hessian-driven damping term attached to the velocity of the primal trajectory $x(t)$. The rates we obtained are identical to those in [15], but our system further allows us to show the weak convergence of the trajectories towards primal-dual solutions of the minimization problem.

Remark 2.4. When $r=1$ and $\beta(t)\equiv 1$, we obtain a system similar to the one addressed by Boţ and Nguyen in [12]. Again, our system features an extra Hessian-driven damping term, which does not appear in [12]. Our rates coincide with those in this work, but, as an interesting point, our system does not require to assume that ∇f is L-Lipschitz continuous to show the weak convergence of the generated trajectories.

2.3 Some numerical experiments

In this subsection, we will complement the theoretical results with two numerical examples. The first one is the minimization of a strongly convex function under linear constraints, and the second one is finding the saddle points of a certain convex-concave function.

Example 2.5. Consider the minimization problem

$$\begin{aligned}\min\quad & f(x_1,x_2,x_3,x_4):=(x_1-1)^2+(x_2-1)^2+x_3^2+x_4^2\\ \text{subject to}\quad & x_1-x_2-x_3=0\\ & x_2-x_4=0.\end{aligned}$$

The optimality conditions can be calculated and lead to the primal-dual solution pair

$$x_*=\begin{bmatrix}0.8\\0.6\\0.2\\0.6\end{bmatrix}\qquad\text{and}\qquad\lambda_*=\begin{bmatrix}0.4\\1.2\end{bmatrix}.$$

For $t \geq t_0 = 1$, we plot the functional values as well as the feasibility gap along the trajectories generated by (54). In Figure 1 we have parameters $\alpha = 8$, $\theta = \frac{1}{4}$ and $\beta(t) \equiv 1$. The predicted convergence rates in this case are of $o\left(\frac{1}{t^{2r}}\right)$ as $t \to +\infty$, so as expected, we can see faster convergence behaviour as r goes from 0.2 to 1.

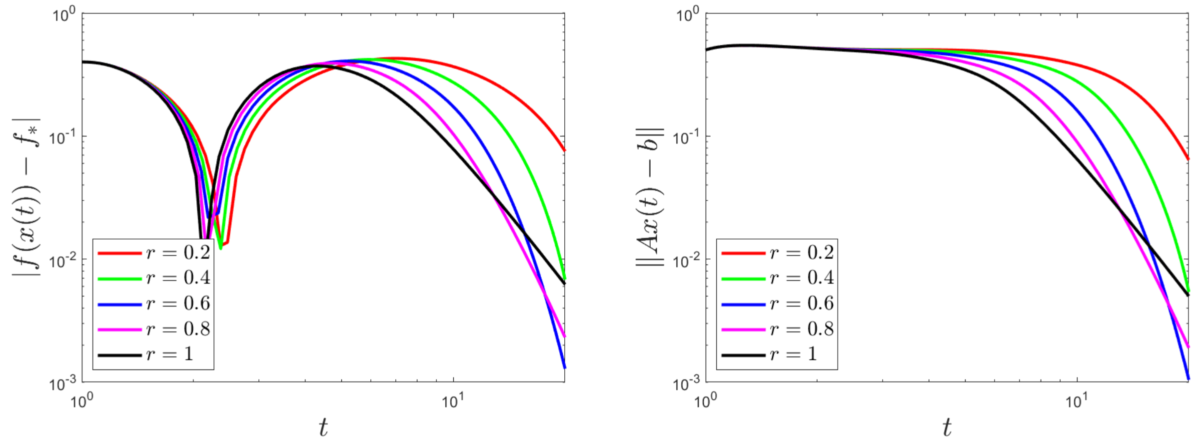

Figure 1: Influence of r on the functional values and feasibility gap along the trajectories, $\beta(t) \equiv 1$

In Figure 2 we plot the same quantities, but this time, we choose $\alpha = 8$, $\theta = \frac{1}{3}$, $\delta = 2$ and $\beta(t) = \frac{1}{t^{2r}} \exp\left[\left(\frac{1}{\theta} - \delta\right)\frac{t^{1-r}}{1-r}\right]$. Here, the predicted convergence rates are of $o\left(\exp\left[-\left(\frac{1}{\theta} - \delta\right)\frac{t^{1-r}}{1-r}\right]\right)$ as $t \to +\infty$. As expected, since t^{1-r} grows to infinity slower as r approaches 1, the convergence behaviour speeds up as r goes from near 1 to 0.

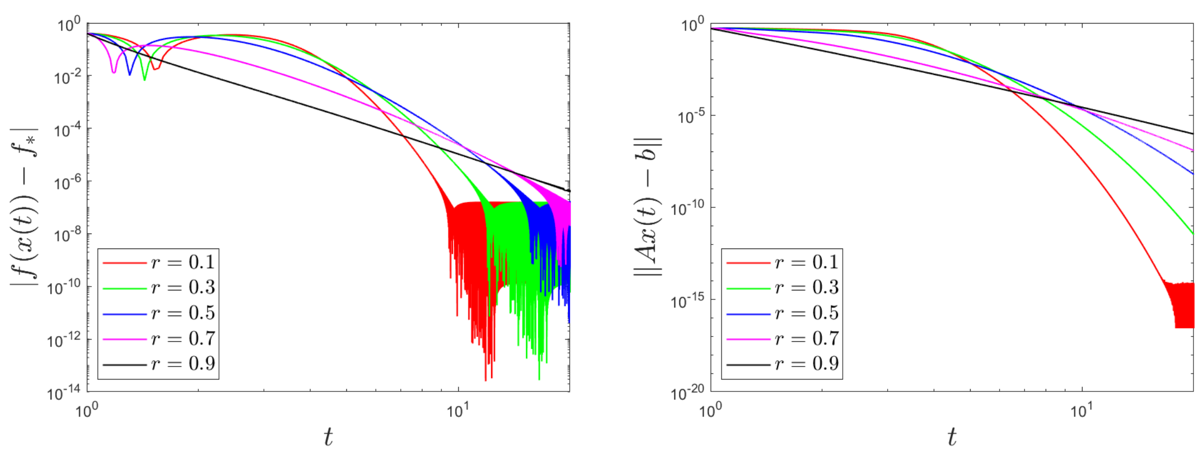

Figure 2: Influence of r on the functional values and feasibility gap along the trajectories, $\beta(t) = \frac{1}{t^{2r}} \exp\left[\left(\frac{1}{\theta} - \delta\right)\frac{t^{1-r}}{1-r}\right]$

Example 2.6. We address the saddle point problem

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^n} \Phi(x, y) := \frac{1}{2}\langle x, Hx\rangle - \langle x, h\rangle - \langle y, Ax - b\rangle,$$

where

$$A := \frac{1}{4}\begin{bmatrix} & & & -1 & 1 \\ & & \iddots & \iddots & \\ & -1 & 1 & & \\ -1 & 1 & & & \\ 1 & & & & \end{bmatrix} \in \mathbb{R}^{n\times n}, \quad H := 2A^*A, \quad b := \frac{1}{4}\begin{bmatrix}1\\1\\\vdots\\1\\1\end{bmatrix} \in \mathbb{R}^n, \quad h := \frac{1}{4}\begin{bmatrix}0\\0\\\vdots\\0\\1\end{bmatrix} \in \mathbb{R}^n.$$

The corresponding continuous and monotone operator whose zeros we wish to find is given by

$$V(x,y) := \begin{bmatrix} \nabla_x \Phi(x,y) \\ -\nabla_y \Phi(x,y) \end{bmatrix} = \begin{bmatrix} Hx - h - A^*y \\ Ax - b \end{bmatrix}.$$

In Figure 3, for $t \geq t_0 = 1$ we plot the norm of the operator V along the trajectories generated by (13). We choose $\alpha = 8$, $\theta = \frac{1}{4}$ and $\beta(t) \equiv 1$ and $\beta(t) = t$ for the first and second plots respectively, and $\delta = 3$ and $\beta(t) = \frac{1}{t^{2r}} \exp\left[\left(\frac{1}{\theta} - \delta\right)\frac{t^{1-r}}{1-r}\right]$ for the third plot. The predicted rates for the first two plots are $o\left(\frac{1}{t^{2r}}\right)$ and $o\left(\frac{1}{t^{2r+1}}\right)$ as $t \to +\infty$ respectively, so as expected we see an overall faster convergence behaviour in the second plot, and in both plots the quantities approach zero faster r moves from near 0 to 1. As we argued in the first example, in the third plot we have rates of $o\left(\exp\left[-\left(\frac{1}{\theta} - \delta\right)\frac{t^{1-r}}{1-r}\right]\right)$ as $t \to +\infty$, so we see a trend of faster convergence the smaller r is.

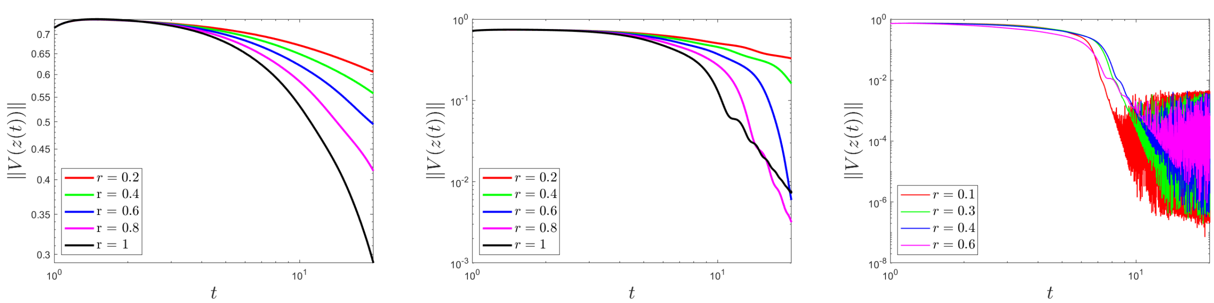

Figure 3: Influence of r on the norm of the operator along the trajectories, with parameter β chosen respectively as $\beta(t) \equiv 1$, $\beta(t) = t$ and $\beta(t) = \frac{1}{t^{2r}} \exp\left[\left(\frac{1}{\theta} - \delta\right)\frac{t^{1-r}}{1-r}\right]$

3 An implicit discretization of the continuous system

3.1 Discretization of an equivalent first-order reformulation

Given the second order system

$$\ddot{z}(t) + \frac{\alpha}{t^r}\dot{z}(t) + \theta t^r \beta(t) \frac{d}{dt}V(z(t)) + \beta(t)V(z(t)) = 0, \tag{62}$$

we readily see it is equivalent to the following first-order reformulation:

$$\begin{cases} \dot{u}(t) &= 2t^r\left[(2r\theta t^{r-1} - 1)\beta(t) + \theta t^r \dot{\beta}(t)\right]V(z(t)) + 2r(1-r)t^{r-2}z(t) \\ u(t) &= 2(\alpha - rt^{r-1})z(t) + 2t^r\dot{z}(t) + 2\theta t^{2r}\beta(t)V(z(t)). \end{cases} \tag{63}$$

A seemingly useful observation to obtain a discretization which works is to write

$$2r(1-r)t^{r-2}z(t) = 2r\frac{d}{dt}(-t^{r-1})z(t). \tag{64}$$

Taking (64) into account, we consider the following discretization of (63):

$$\begin{cases} u^{k+1} - u^k & = 2k^r\Big[(2r\theta k^{r-1}-1)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]V(z^{k+1}) + 2r\Big[k^{r-1} - (k+1)^{r-1}\Big]z^{k+1} \\ u^{k+1} & = 2\big[\alpha - r(k+1)^{r-1}\big]z^{k+1} + 2(k+1)^r(z^{k+1} - z^k) + 2\theta(k+1)^{2r}\beta_k V(z^{k+1}). \end{cases} \tag{65}$$

Now, from the second line of (65) we deduce

$$\begin{aligned} u^{k+1} - u^k &= 2\big[\alpha - r(k+1)^{r-1}\big]z^{k+1} + 2(k+1)^r(z^{k+1} - z^k) + 2\theta(k+1)^{2r}\beta_k V(z^{k+1}) \\ &\quad - 2\big[\alpha - rk^{r-1}\big]z^k - 2k^r(z^k - z^{k-1}) - 2\theta k^{2r}\beta_{k-1}V(z^k) \\ &= 2\alpha(z^{k+1} - z^k) - 2r\Big[(k+1)^{r-1}z^{k+1} - k^{r-1}z^k\Big] + 2(k+1)^r(z^{k+1} - z^k) \\ &\quad - 2k^r(z^k - z^{k-1}) + 2\theta\Big[(k+1)^{2r}\beta_k V(z^{k+1}) - k^{2r}\beta_{k-1}V(z^k)\Big] \\ &= 2\alpha(z^{k+1} - z^k) - 2r\Big[((k+1)^{r-1} - k^{r-1})z^{k+1} + k^{r-1}(z^{k+1} - z^k)\Big] \\ &\quad + 2(k+1)^r(z^{k+1} - z^k) - 2k^r(z^k - z^{k-1}) + 2\theta\Big[(k+1)^{2r}\beta_k V(z^{k+1}) - k^{2r}\beta_{k-1}V(z^k)\Big] \\ &= 2r\Big[k^{r-1} - (k+1)^{r-1}\Big]z^{k+1} + 2\Big[\alpha - rk^{r-1} + (k+1)^r\Big](z^{k+1} - z^k) - 2k^r(z^k - z^{k-1}) \\ &\quad + 2\theta\Big[(k+1)^{2r}\beta_k - k^{2r}\beta_{k-1}\Big]V(z^{k+1}) + 2\theta k^{2r}\beta_{k-1}\Big[V(z^{k+1}) - V(z^k)\Big] \\ &= 2k^r\Big[(2r\theta k^{r-1} - 1)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]V(z^{k+1}) + 2r\Big[k^{r-1} - (k+1)^{r-1}\Big]z^{k+1}, \end{aligned}$$

where the last equality comes from the first line of (65). Cancelling the second summand in the last line, we come to

$$\begin{aligned} &2\Big[\alpha - rk^{r-1} + (k+1)^r\Big](z^{k+1} - z^k) - 2k^r(z^k - z^{k-1}) + 2\theta\Big[(k+1)^{2r}\beta_k - k^{2r}\beta_{k-1}\Big]V(z^{k+1}) \\ &+ 2\theta k^{2r}\beta_{k-1}\Big[V(z^{k+1}) - V(z^k)\Big] \\ =\ &2k^r\Big[(2r\theta k^{r-1} - 1)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]V(z^{k+1}). \end{aligned} \tag{66}$$

Rearranging the terms of the previous equality leads to the formulation (15) presented in the introduction:

$$\begin{aligned} z^{k+1} &= z^k + \frac{k^r}{\alpha - rk^{r-1} + (k+1)^r}(z^k - z^{k-1}) - \frac{\theta k^{2r}\beta_{k-1}}{\alpha - rk^{r-1} + (k+1)^r}\Big[V(z^{k+1}) - V(z^k)\Big] \\ &\quad - \frac{\theta\Big\{\Big[(k+1)^{2r} - k^{2r}\Big] - 2rk^{2r-1}\Big\} + k^r}{\alpha - rk^{r-1} + (k+1)^r}\beta_k V(z^{k+1}). \end{aligned}$$

3.2 Convergence rates and weak convergence of iterates

Before we start with the analysis, we need to remark some inequalities involving the sequence $(\beta_k)_{k\geq 1}$.

Remark 3.1. In (16), the growth conditions reads $\sup_{k\geq k_0} k^r\left(\frac{\beta_k-\beta_{k-1}}{\beta_k}+\frac{2r}{k}\right)<\frac{1}{2\theta}<\frac{1}{\theta}$. Throughout the majority of the proofs, we will use a couple of inequalities involving the sequence $(\beta_k)_{k\geq 0}$ which are entailed by assuming the supremum is strictly less than $\frac{1}{\theta}$. More precisely, there exists $\delta>0$ such that $\delta<\frac{1}{\theta}$ and

$$\sup_{k\geq k_0} k^r\left[\frac{\beta_k-\beta_{k-1}}{\beta_k}+\frac{2r}{k}\right]<\frac{1}{\theta}-\delta. \tag{67}$$

It follows that $\frac{k^r\left(\beta_k-\beta_{k-1}\right)}{\beta_k}+2rk^{r-1}<\frac{1}{\theta}-\delta$ and therefore

$$(2r\theta k^{r-1}-1)\beta_k+\theta k^r(\beta_k-\beta_{k-1})\leq -\delta\theta\beta_k<0 \qquad \forall k\geq k_0. \tag{68}$$

The previous inequality also produces

$$\theta k^r(\beta_k-\beta_{k-1})\leq (1-2r\theta k^{r-1}-\delta\theta)\beta_k \qquad \forall k\geq k_0. \tag{69}$$

Regrouping the terms with β_k and β_{k-1} separately yields $\left(2rk^{r-1}-\frac{1}{\theta}+\delta+k^r\right)\beta_k\leq k^r\beta_{k-1}$ and thus

$$\beta_k\leq \frac{k^r}{2rk^{r-1}-\frac{1}{\theta}+\delta+k^r}\beta_{k-1}\leq M_\beta\beta_{k-1}, \tag{70}$$

where $M_\beta>1$ but may be taken arbitrarily close to 1 for k large enough.

Remark 3.2. As we stated in the introduction, when $r\in(0,1)$ then it is possible to choose β_k such that $k^{2r}\beta_k$ grows exponentially. We claim that if $0<\delta<\frac{1}{2\theta}$, then

$$\beta_k=\frac{1}{k^{2r}}\cdot e^{\left(\frac{1}{2\theta}-\delta\right)\frac{k^{1-r}}{1-r}}$$

satisfies growth condition (16). Indeed, first we notice that $\beta_k=\beta(k)$, where $\beta(t)$ is defined in continuous-time as

$$\beta(t)=\frac{1}{t^{2r}}\cdot e^{\left(\frac{1}{2\theta}-\delta\right)\frac{t^{1-r}}{1-r}}.$$

By the mean value theorem, there exists $\xi_k\in(k-1,k)$ such that

$$\beta_k-\beta_{k-1}=\beta(k)-\beta(k-1)=\dot\beta(\xi_k)=\exp\left[\left(\frac{1}{2\theta}-\delta\right)\frac{\xi_k^{1-r}}{1-r}\right]\left[-\frac{2r}{\xi_k^{2r+1}}+\left(\frac{1}{2\theta}-\delta\right)\frac{1}{\xi_k^{3r}}\right].$$

Therefore, for every k we have

$$k^r\left(\frac{\beta_k-\beta_{k-1}}{\beta_k}+\frac{2r}{k}\right)=\exp\left[\left(\frac{1}{2\theta}-\delta\right)\frac{\xi_k^{1-r}-k^{1-r}}{1-r}\right]\left(\frac{1}{2\theta}-\delta\right)\frac{k^{3r}}{\xi_k^{3r}} \tag{71}$$

$$-\exp\left[\left(\frac{1}{2\theta}-\delta\right)\frac{\xi_k^{1-r}-k^{1-r}}{1-r}\right]\frac{2rk^{3r}}{\xi_k^{2r+1}}+\frac{2r}{k^{1-r}}. \tag{72}$$

First, we look at line (71). Again by the mean value theorem, we have $k^{1-r}-\xi_k^{1-r}=(1-r)\tau_k^{-r}(k-\xi_k)$ for some $\tau_k\in(\xi_k,k)\subseteq(k-1,k)$. It follows that $\frac{1}{k^r}\leq\frac{1}{\tau_k^r}\leq\frac{1}{(k-1)^r}$ and thus $\xi_k^{1-r}-k^{1-r}\to 0$ as $k\to+\infty$. Since $1\leq\frac{k^{3r}}{\xi_k^{3r}}\leq\left(\frac{k}{k-1}\right)^{3r}$, we have $\frac{k^{3r}}{\xi_k^{3r}}\to 1$ as $k\to+\infty$. All in all, line (71) approaches $\frac{1}{2\theta}-\delta$ as $k\to+\infty$. Regarding line (72), as we just saw, the exponential term approaches 1 and $\frac{2r}{k^{1-r}}\to 0$ as $k\to+\infty$. Since $\frac{1}{k^{1-r}}\leq\frac{k^{3r}}{\xi_k^{2r+1}}\leq\frac{k^{3r}}{k^{2r+1}}$, we have $\frac{2rk^{3r}}{k^{2r+1}}\to 0$ as $k\to+\infty$. Summarizing, we have $k^r\left(\frac{\beta_k-\beta_{k-1}}{\beta_k}+\frac{2r}{k}\right)\to\frac{1}{2\theta}-\delta$ as $k\to+\infty$, which means that for large enough k_0 the supremum condition (67) is fulfilled.

Energy function: For analizing the convergence properties of the implicit algorithm (66), we make use of the following discrete energy function:

$$\mathcal{E}_\lambda^k := \frac{1}{2}\left\|k^{\rho-r}2\lambda(z^k - z_*) + 2k^\rho(z^k - z^{k-1}) + \theta k^{\rho+r}\beta_{k-1}V(z^k)\right\|^2 \tag{73}$$

$$+ 2\lambda k^{2(\rho-r)}(\alpha - (2\rho - r)k^{r-1} - \lambda)\|z^k - z_*\|^2 \tag{74}$$

$$+ 2\lambda\theta k^{2\rho}\beta_{k-1}\langle z^k - z_*, V(z^k)\rangle \tag{75}$$

$$+ \frac{\theta^2}{2}(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\|V(z^k)\|^2. \tag{76}$$

Here, ρ and λ are taken as follows:

- If $r \in (0,1)$, then $0 < \lambda < \alpha$ and $0 < \rho < r$;
- If $r = 1$, then $0 < \lambda < \alpha - 1$ and $\rho = r = 1$.

The following theorem shows that the $\mathcal{O}$ convergence rates, obtained as an intermediate step for the continuous dynamical system, carry over to the discrete setting.

Theorem 3.3. *Suppose that $\alpha, \theta > 0$, $r \in (0,1]$ and $(\beta_k)_{k\geq 1} \subseteq (0,+\infty)$ satisfy the assumptions laid down in subsection (1.3). Let z^0, z^1 be initial points in $\mathcal{H}$, let $(z^k)_{k\geq 2}$ be the sequence generated by the implicit algorithm (66), equivalently, by (15), and let z_* be a zero of V. Consider the following convergence rates as $k \to +\infty$:*

$$\|V(z^k)\| = \mathcal{O}\left(\frac{1}{k^{\rho+r}\beta_k}\right) \quad \langle z^{k+1} - z_*, V(z^{k+1})\rangle = \mathcal{O}\left(\frac{1}{k^{2\rho}\beta_k}\right), \quad \|z^k - z^{k-1}\| = \mathcal{O}\left(\frac{1}{k^\rho}\right).$$

The following statements are true:

(i) *If $r \in (0,1)$, the above rates hold for $\rho \in (0,r)$. Furthermore, if $(z^k)_{k\geq 2}$ is bounded, then these rates hold for $\rho = r$.*

(ii) *If $r = 1$, then $(z^k)_{k\geq 2}$ is bounded; moreover, the above rates hold for $\rho = 1$.*

To produce the proof for this theorem, we first exhibit two intermediate results.

Lemma 3.4. *Under the same hypotheses of Theorem 3.3, for large enough k it holds*

$$\begin{aligned}
&\mathcal{E}_\lambda^{k+1} - \mathcal{E}_\lambda^k \\
&\leq \Bigg\{2\lambda^2\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big] + 2\lambda(2\rho - r)\Big[k^{r-1} - (k+1)^{r-1}\Big]k^{2(\rho-r)} \\
&\qquad + 2\lambda(\alpha - (2\rho - r)(k+1)^{r-1} - \lambda)\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big]\Bigg\}\|z^{k+1} - z_*\|^2 \\
&\quad + \Bigg\{8\lambda(r-\rho)k^{2(\rho-r)} + 4\lambda\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big](k+1)^r\Bigg\}\langle z^{k+1} - z_*, z^{k+1} - z^k\rangle \\
&\quad + 4\lambda k^{2\rho-r}\Big[(2r\theta k^{r-1} - 1)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]\langle z^{k+1} - z_*, V(z^{k+1})\rangle \\
&\quad + 4k^{2\rho-r}(\lambda + rk^{r-1} - \alpha)\|z^{k+1} - z^k\|^2
\end{aligned}$$

$$+\Big\{k^{2(\rho-r)}\Big[2\eta_k(k+1)^r+2\theta(\lambda+rk^{r-1}-\alpha)(k+1)^{2r}\beta_k-2(\lambda+rk^{r-1}-\alpha)\eta_k\Big]$$

$$+2\theta\Big[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\Big](k+1)^{3r}\beta_k+2\lambda\theta k^{2\rho}\beta_{k-1}\Big\}\langle z^{k+1}-z^k,V(z^{k+1})\rangle$$

$$+\Big\{k^{2(\rho-r)}\left[\theta(k+1)^{2r}\eta_k\beta_k-\frac{1}{2}\eta_k^2\right]+\frac{\theta^2}{2}\Big[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\Big](k+1)^{4r}\beta_k^2$$

$$+\frac{\theta^2}{2}\Big[(k+2)^{2r}(k+1)^{2\rho}\beta_{k+1}\beta_k-(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\Big]\Big\}\|V(z^{k+1})\|^2$$

$$+\theta k^{2\rho}\beta_{k-1}\eta_k\langle V(z^{k+1}),V(z^{k+1})-V(z^k)\rangle$$

$$-k^{2(\rho-r)}\frac{\theta^2}{2}\Big[k^{4r}\beta_{k-1}^2+(k+1)^{2r}k^{2r}\beta_k\beta_{k-1}\Big]\|V(z^{k+1})-V(z^k)\|^2,$$

where

$$\eta_k:=2k^r\Big[(2r\theta k^{r-1}-1)\beta_k+\theta k^r(\beta_k-\beta_{k-1})\Big]-\theta\Big[(k+1)^{2r}\beta_k-k^{2r}\beta_{k-1}\Big].$$

Proof. First of all, just as in the continuous case, notice that $\mathcal{E}_\lambda^k$ is eventually nonnegative. Indeed, go back to line (74). If $r\in(0,1)$, then $0<\lambda<\alpha$ together with the fact that $k^{r-1}\to 0$ as $k\to+\infty$ gives $\alpha-(2\rho-r)k^{r-1}-\lambda\geq 0$ for large enough k. If $r=1$, then $\alpha-(2\rho-r)k^{r-1}-\lambda\equiv\alpha-1-\lambda>0$ by assumption. We will compute the difference $\mathcal{E}_\lambda^{k+1}-\mathcal{E}_\lambda^k$ line by line.

Line (73): Define, for each k,

$$u_\lambda^k:=2\lambda(z^k-z_*)+2k^r(z^k-z^{k-1})+\theta k^{2r}\beta_{k-1}V(z^k).$$

Notice that (73) can be written as

$$\frac{1}{2}\|k^{\rho-r}u_\lambda^k\|^2.$$

We have

$$\begin{aligned}&\frac{1}{2}\|(k+1)^{\rho-r}u_\lambda^{k+1}\|^2-\frac{1}{2}\|k^{\rho-r}u_\lambda^k\|^2\\&=k^{2(\rho-r)}\left[\frac{1}{2}\|u_\lambda^{k+1}\|^2-\frac{1}{2}\|u_\lambda^k\|^2\right]+\frac{1}{2}\Big[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\Big]\|u_\lambda^{k+1}\|^2\\&=k^{2(\rho-r)}\left[\langle u_\lambda^{k+1},u_\lambda^{k+1}-u_\lambda^k\rangle-\frac{1}{2}\|u_\lambda^{k+1}-u_\lambda^k\|^2\right]+\frac{1}{2}\Big[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\Big]\|u_\lambda^{k+1}\|^2,\end{aligned}\tag{77}$$

so we first compute $u_\lambda^{k+1}-u_\lambda^k$:

$$\begin{aligned}u_\lambda^{k+1}-u_\lambda^k&=2\lambda(z^{k+1}-z_*)+2(k+1)^r(z^{k+1}-z^k)+\theta(k+1)^{2r}\beta_kV(z^{k+1})\\&\quad-2\lambda(z^k-z_*)-2k^r(z^k-z^{k-1})-\theta k^{2r}\beta_{k-1}V(z^k)\\&=2\lambda(z^{k+1}-z^k)+2\Big[(k+1)^r(z^{k+1}-z^k)-k^r(z^k-z^{k-1})\Big]\\&\quad+\theta\Big[(k+1)^{2r}\beta_kV(z^{k+1})-k^{2r}\beta_{k-1}V(z^k)\Big]\\&=2(\lambda+rk^{r-1}-\alpha)(z^{k+1}-z^k)+2\Big[\alpha-rk^{r-1}+(k+1)^r\Big](z^{k+1}-z^k)-2k^r(z^k-z^{k-1})\\&\quad+\theta\Big[(k+1)^{2r}\beta_k-k^{2r}\beta_{k-1}\Big]V(z^{k+1})+\theta k^{2r}\beta_{k-1}\Big[V(z^{k+1})-V(z^k)\Big]\end{aligned}$$

$$\begin{aligned}
\text{use (66)} \quad &= 2\big(\lambda + rk^{r-1} - \alpha\big)\big(z^{k+1} - z^k\big) + 2k^r\Big[\big(2r\theta k^{r-1} - 1\big)\beta_k + \theta k^r\big(\beta_k - \beta_{k-1}\big)\Big]V(z^{k+1}) \\
&\quad - \theta\Big[(k+1)^{2r}\beta_k - k^{2r}\beta_{k-1}\Big]V(z^{k+1}) - \theta k^{2r}\beta_{k-1}\Big[V(z^{k+1}) - V(z^k)\Big] \\
&= 2\big(\lambda + rk^{r-1} - \alpha\big)\big(z^{k+1} - z^k\big) - \theta k^{2r}\beta_{k-1}\Big[V(z^{k+1}) - V(z^k)\Big] \\
&\quad + \underbrace{\Big\{2k^r\Big[\big(2r\theta k^{r-1} - 1\big)\beta_k + \theta k^r\big(\beta_k - \beta_{k-1}\big)\Big] - \theta\Big[(k+1)^{2r} - k^{2r}\Big]\Big\}}_{=\eta_k} V(z^{k+1}).
\end{aligned}$$

With this at hand, we compute $\langle u_\lambda^{k+1}, u_\lambda^{k+1} - u_\lambda^k\rangle$:

$$\begin{aligned}
&\langle u_\lambda^{k+1}, u_\lambda^{k+1} - u_\lambda^k\rangle \\
&= \Big\langle 2\lambda(z^{k+1} - z_*) + 2(k+1)^r\big(z^{k+1} - z^k\big) + \theta(k+1)^{2r}\beta_k V(z^{k+1}), \\
&\qquad 2\big(\lambda + rk^{r-1} - \alpha\big)\big(z^{k+1} - z^k\big) + \eta_k V(z^{k+1}) - \theta k^{2r}\beta_{k-1}\Big[V(z^{k+1}) - V(z^k)\Big]\Big\rangle \\
&= 4\lambda\big(\lambda + rk^{r-1} - \alpha\big)\langle z^{k+1} - z_*, z^{k+1} - z^k\rangle + 2\lambda\eta_k\langle z^{k+1} - z_*, V(z^{k+1})\rangle \\
&\quad - 2\lambda\theta k^{2r}\beta_{k-1}\langle z^{k+1} - z_*, V(z^{k+1}) - V(z^k)\rangle + 4\big(\lambda + rk^{r-1} - \alpha\big)(k+1)^r\big\|z^{k+1} - z^k\big\|^2 \\
&\quad + \Big[2\eta_k(k+1)^r + 2\theta\big(\lambda + rk^{r-1} - \alpha\big)(k+1)^{2r}\beta_k\Big]\langle z^{k+1} - z^k, V(z^{k+1})\rangle \\
&\quad - 2\theta(k+1)^r k^{2r}\beta_{k-1}\langle z^{k+1} - z^k, V(z^{k+1}) - V(z^k)\rangle + \theta(k+1)^{2r}\eta_k\beta_k\|V(z^{k+1})\|^2 \\
&\quad - \theta^2(k+1)^{2r}k^{2r}\beta_k\beta_{k-1}\langle V(z^{k+1}), V(z^{k+1}) - V(z^k)\rangle.
\end{aligned}$$

Now, $\|u_\lambda^{k+1} - u_\lambda^k\|^2$:

$$\begin{aligned}
\|u_\lambda^{k+1} - u_\lambda^k\|^2 &= \Big\|2\big(\lambda + rk^{r-1} - \alpha\big)\big(z^{k+1} - z^k\big) + \eta_k V(z^{k+1}) - \theta k^{2r}\beta_{k-1}\Big[V(z^{k+1}) - V(z^k)\Big]\Big\|^2 \\
&= 4\big(\lambda + rk^{r-1} - \alpha\big)^2\big\|z^{k+1} - z^k\big\|^2 + \eta_k^2\|V(z^{k+1})\|^2 \\
&\quad + \theta^2 k^{4r}\beta_{k-1}^2\big\|V(z^{k+1}) - V(z^k)\big\|^2 + 4\big(\lambda + rk^{r-1} - \alpha\big)\eta_k\langle z^{k+1} - z^k, V(z^{k+1})\rangle \\
&\quad - 4\theta\big(\lambda + rk^{r-1} - \alpha\big)k^{2r}\beta_{k-1}\langle z^{k+1} - z^k, V(z^{k+1}) - V(z^k)\rangle \\
&\quad - 2\theta\eta_k k^{2r}\beta_{k-1}\langle V(z^{k+1}), V(z^{k+1}) - V(z^k)\rangle.
\end{aligned}$$

Now, recalling (77), we put everything together:

$$\begin{aligned}
&\frac{1}{2}\big\|(k+1)^{\rho-r}u_\lambda^{k+1}\big\|^2 - \frac{1}{2}\big\|k^{\rho-r}u_\lambda^k\big\|^2 \\
&= k^{2(\rho-r)}\Big\{4\lambda\big(\lambda + rk^{r-1} - \alpha\big)\langle z^{k+1} - z_*, z^{k+1} - z^k\rangle + 2\lambda\eta_k\langle z^{k+1} - z_*, V(z^{k+1})\rangle \\
&\quad - 2\lambda\theta k^{2r}\beta_{k-1}\langle z^{k+1} - z_*, V(z^{k+1}) - V(z^k)\rangle \\
&\quad + \Big[4\big(\lambda + rk^{r-1} - \alpha\big)(k+1)^r - 2\big(\lambda + rk^{r-1} - \alpha\big)^2\Big]\big\|z^{k+1} - z^k\big\|^2 \\
&\quad + \Big[2\eta_k(k+1)^r + 2\theta\big(\lambda + rk^{r-1} - \alpha\big)(k+1)^{2r}\beta_k - 2\big(\lambda + rk^{r-1} - \alpha\big)\eta_k\Big]\langle z^{k+1} - z^k, V(z^{k+1})\rangle \\
&\quad + \Big[-2\theta(k+1)^r k^{2r}\beta_{k-1} + 2\theta\big(\lambda + rk^{r-1} - \alpha\big)k^{2r}\beta_{k-1}\Big]\langle z^{k+1} - z^k, V(z^{k+1}) - V(z^k)\rangle \\
&\quad + \Big[\theta(k+1)^{2r}\eta_k\beta_k - \frac{1}{2}\eta_k^2\Big]\|V(z^{k+1})\|^2 - \frac{\theta^2}{2}k^{4r}\beta_{k-1}^2\big\|V(z^{k+1}) - V(z^k)\big\|^2
\end{aligned}$$

$$+\Big[-\theta^2(k+1)^{2r}k^{2r}\beta_k\beta_{k-1}+\theta k^{2r}\eta_k\beta_{k-1}\Big]\langle V(z^{k+1}),V(z^{k+1})-V(z^k)\rangle\Big\}$$

$$+\frac{1}{2}\Big[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\Big]\Big\{4\lambda^2\|z^{k+1}-z_*\|^2+4(k+1)^r\big\|z^{k+1}-z^k\big\|^2+\theta^2(k+1)^{4r}\beta_k^2\|V(z^{k+1})\|^2$$

$$+8\lambda(k+1)^r\langle z^{k+1}-z_*,z^{k+1}-z^k\rangle+4\lambda\theta(k+1)^{2r}\beta_k\langle z^{k+1}-z_*,V(z^{k+1})\rangle$$

$$+4\theta(k+1)^{3r}\beta_k\langle z^{k+1}-z_k,V(z^{k+1})\rangle\Big\}.$$

Line (74): we have

$$2\lambda(k+1)^{2(\rho-r)}(\alpha-(2\rho-r)(k+1)^{r-1}-\lambda)\|z^{k+1}-z_*\|^2-2\lambda k^{2(\rho-r)}(\alpha-(2\rho-r)k^{r-1}-\lambda)\|z^k-z_*\|^2$$

$$=k^{2(\rho-r)}\Big\{2\lambda(2\rho-r)\Big[k^{r-1}-r(k+1)^{r-1}\Big]\|z^{k+1}-z_*\|^2-2\lambda(\alpha-(2\rho-r)k^{r-1}-\lambda)\big\|z^{k+1}-z^k\big\|^2$$

$$+4\lambda(\alpha-(2\rho-r)k^{r-1}-\lambda)\langle z^{k+1}-z_*,z^{k+1}-z^{k+1}-z^k\rangle\Big\}$$

$$+2\lambda\Big[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\Big]\Big(\alpha-(2\rho-r)(k+1)^{r-1}-\lambda\Big)\Big\|z^{k+1}-z_*\Big\|^2.$$

Line (75): we can write

$$2\lambda\theta(k+1)^{2\rho}\beta_k\langle z^{k+1}-z_*,V(z^{k+1})\rangle-2\lambda\theta k^{2\rho}\beta_{k-1}\langle z^k-z_*,V(z^k)\rangle$$

$$=2\lambda\theta\Big[(k+1)^{2\rho}\beta_k-k^{2\rho}\beta_{k-1}\Big]\langle z^{k+1}-z_*,V(z^{k+1})\rangle+2\theta k^{2\rho}\beta_{k-1}\Big[\langle z^{k+1}-z_*,V(z^{k+1})\rangle-\langle z^k-z_*,V(z^k)\rangle\Big]$$

$$=2\lambda\theta\Big[(k+1)^{2\rho}\beta_k-k^{2\rho}\beta_{k-1}\Big]\langle z^{k+1}-z_*,V(z^{k+1})\rangle+2\lambda\theta k^{2\rho}\beta_{k-1}\langle z^{k+1}-z_*,V(z^{k+1})-V(z^k)\rangle$$

$$+2\lambda\theta k^{2\rho}\beta_{k-1}\langle z^{k+1}-z^k,V(z^k)\rangle$$

$$=2\lambda\theta\Big[(k+1)^{2\rho}\beta_k-k^{2\rho}\beta_{k-1}\Big]\langle z^{k+1}-z_*,V(z^{k+1})\rangle+2\lambda\theta k^{2\rho}\beta_{k-1}\langle z^{k+1}-z_*,V(z^{k+1})-V(z^k)\rangle$$

$$-2\lambda\theta k^{2\rho}\beta_{k-1}\langle z^{k+1}-z^k,V(z^{k+1})-V(z^k)\rangle+2\lambda\theta k^{2\rho}\beta_{k-1}\langle z^{k+1}-z^k,V(z^{k+1})\rangle.$$

Line (76): we obtain

$$\frac{\theta^2}{2}\Big[(k+2)^{2r}(k+1)^{2\rho}\beta_{k+1}\beta_k\|V(z^{k+1})\|^2-(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\|V(z^k)\|^2\Big]$$

$$=\frac{\theta^2}{2}\Big[(k+2)^{2r}(k+1)^{2\rho}\beta_{k+1}\beta_k-(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\Big]\|V(z^{k+1})\|^2$$

$$+\frac{\theta^2}{2}(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\Big[\|V(z^{k+1})\|^2-\|V(z^k)\|^2\Big]$$

$$=\frac{\theta^2}{2}\Big[(k+2)^{2r}(k+1)^{2\rho}\beta_{k+1}\beta_k-(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\Big]\|V(z^{k+1})\|^2$$

$$+(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\Big[\langle V(z^{k+1}),V(z^{k+1})-V(z^k)\rangle-\frac{1}{2}\|V(z^{k+1})-V(z^k)\|^2\Big]$$

$$=\frac{\theta^2}{2}\Big[(k+2)^{2r}(k+1)^{2\rho}\beta_{k+1}\beta_k-(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\Big]\|V(z^{k+1})\|^2$$

$$+\theta^2(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\langle V(z^{k+1}),V(z^{k+1})-V(z^k)\rangle-\frac{\theta^2}{2}(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\big\|V(z^{k+1})-V(z^k)\big\|^2.$$

Now, we combine everything:

$$
\begin{aligned}
&\mathcal{E}_\lambda^{k+1} - \mathcal{E}_\lambda^k \\
&= \Big\{ 2\lambda^2 \Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)} \Big] + 2\lambda(2\rho - r)k^{2(\rho-r)} \Big[k^{r-1} - (k+1)^{r-1} \Big] \\
&\qquad + 2\lambda \Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)} \Big] \big(\alpha - (2\rho - r)(k+1)^{r-1} - \lambda\big) \Big\} \|z^{k+1} - z_*\|^2 && (78) \\
&\quad + \Big\{ 4\lambda k^{2(\rho-r)} \big(\lambda + rk^{r-1} - \alpha\big) + 4\lambda k^{2(\rho-r)} \big(\alpha - (2\rho - r)k^{r-1} - \lambda\big) \\
&\qquad + 4\lambda \Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)} \Big] (k+1)^r \Big\} \langle z^{k+1} - z_*, z^{k+1} - z^k \rangle && (79) \\
&\quad + \Big\{ 2\lambda k^{2(\rho-r)} \eta_k + 2\lambda\theta \Big[(k+1)^{2\rho} \beta_k - k^{2\rho} \beta_{k-1} \Big] \\
&\qquad + 2\lambda\theta \Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)} \Big] (k+1)^{2r} \beta_k \Big\} \langle z^{k+1} - z_*, V(z^{k+1}) \rangle && (80) \\
&\quad + \Big[k^{2(\rho-r)} (-2\lambda\theta k^{2r} \beta_{k-1}) + 2\lambda\theta k^{2\rho} \beta_{k-1} \Big] \langle z^{k+1} - z_*, V(z^{k+1}) - V(z^k) \rangle && (81) \\
&\quad + \Big\{ k^{2(\rho-r)} \Big[4\big(\lambda + rk^{r-1} - \alpha\big)(k+1)^r - 2\big(\lambda + rk^{r-1} - \alpha\big)^2 \Big] + 2\lambda \Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)} \Big] \\
&\qquad - 2\lambda k^{2(\rho-r)} \big(\alpha - (2\rho - r)k^{r-1} - \lambda\big) \Big\} \big\| z^{k+1} - z^k \big\|^2 && (82) \\
&\quad + \Big\{ k^{2(\rho-r)} \Big[2\eta_k (k+1)^r + 2\theta\big(\lambda + rk^{r-1} - \alpha\big)(k+1)^{2r} \beta_k - 2\big(\lambda + rk^{r-1} - \alpha\big)\eta_k \Big] \\
&\qquad + 2\theta \Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)} \Big] (k+1)^{3r} \beta_k + 2\lambda\theta k^{2\rho} \beta_{k-1} \Big\} \langle z^{k+1} - z^k, V(z^{k+1}) \rangle && (83) \\
&\quad + \Big\{ k^{2(\rho-r)} \Big[-2\theta(k+1)^r k^{2r} \beta_{k-1} + 2\theta\big(\lambda + rk^{r-1} - \alpha\big) k^{2r} \beta_{k-1} \Big] \\
&\qquad - 2\lambda\theta k^{2\rho} \beta_{k-1} \Big\} \langle z^{k+1} - z^k, V(z^{k+1}) - V(z^k) \rangle && (84) \\
&\quad + \Big\{ k^{2(\rho-r)} \Big[\theta(k+1)^{2r} \eta_k \beta_k - \frac{1}{2}\eta_k^2 \Big] + \frac{\theta^2}{2} \Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)} \Big] (k+1)^{4r} \beta_k^2 \\
&\qquad + \frac{\theta^2}{2} \Big[(k+2)^{2r} (k+1)^{2\rho} \beta_{k+1} \beta_k - (k+1)^{2r} k^{2\rho} \beta_k \beta_{k-1} \Big] \Big\} \|V(z^{k+1})\|^2 && (85) \\
&\quad + \Big\{ k^{2(\rho-r)} \Big[-\theta^2 (k+1)^{2r} k^{2r} \beta_k \beta_{k-1} + \theta k^{2r} \eta_k \beta_{k-1} \Big] \\
&\qquad + \theta^2 (k+1)^{2r} k^{2\rho} \beta_k \beta_{k-1} \Big\} \langle V(z^{k+1}), V(z^{k+1}) - V(z^k) \rangle && (86)
\end{aligned}
$$

$$+\left[k^{2(\rho-r)}\left(-\frac{\theta^2}{2}k^{4r}\beta_{k-1}^2\right)-\frac{\theta^2}{2}(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\right]\left\|V(z^{k+1})-V(z^k)\right\|^2. \tag{87}$$

We analyze each line separately. Line (79): evidently, we have

$$4\lambda(\lambda+rk^{r-1}-\alpha)+4\lambda(\alpha-(2\rho-r)k^{r-1}-\lambda)+4\lambda\left[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\right](k+1)^r=8\lambda(r-\rho)k^{2\rho-r-1}.$$

Line (80): recalling the definition for η_k, and the fact that $(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\le 0$ (since $2(\rho-r)\le 0$), the coefficient attached to $\langle z^{k+1}-z_*, V(z^{k+1})\rangle$ is less or equal than

$$\begin{aligned}
&2\lambda k^{2(\rho-r)}\eta_k+2\lambda\theta\left[(k+1)^{2\rho}\beta_k-k^{2\rho}\beta_{k-1}\right]\\
&=2\lambda k^{2(\rho-r)}\eta_k+2\lambda\theta k^{2(\rho-r)}\left[(k+1)^{2r}\beta_k-k^{2r}\beta_k\right]\\
&\quad+2\lambda\theta\left\{\left[(k+1)^{2\rho}\beta_k-k^{2\rho}\beta_{k-1}\right]-k^{2(\rho-r)}\left[(k+1)^{2r}\beta_k-k^{2r}\beta_k\right]\right\}\\
&=2\lambda k^{2(\rho-r)}\left\{2k^r\left[(2r\theta k^{r-1}-1)\beta_k+\theta k^r(\beta_k-\beta_{k-1})\right]-\theta\left[(k+1)^{2r}\beta_k-k^{2r}\beta_{k-1}\right]\right\}\\
&\quad+2\lambda\theta k^{2(\rho-r)}\left[(k+1)^{2r}\beta_k-k^{2r}\beta_{k-1}\right]+2\lambda\theta\left[(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\right](k+1)^{2r}\beta_k\\
&\le 4\lambda k^{2\rho-r}\left[(2r\theta k^{r-1}-1)\beta_k+\theta k^r(\beta_k-\beta_{k-1})\right],
\end{aligned}$$

again using that $(k+1)^{2(\rho-r)}-k^{2(\rho-r)}\le 0$.

Line (81): we have

$$k^{2(\rho-r)}(-2\lambda\theta k^{2r})+2\lambda\theta k^{2\rho}=-2\lambda\theta k^{2\rho}+2\lambda\theta k^{2\rho}=0.$$

Line (82): the coefficient accompanying $\left\|z^{k+1}-z^k\right\|^2$ is less than or equal to

$$\begin{aligned}
&k^{2(\rho-r)}\left\{4(\lambda+rk^{r-1}-\alpha)(k+1)^r-2(\lambda+rk^{r-1}-\alpha)^2-2\lambda(\alpha-(2\rho-r)k^{r-1}-\lambda)\right\}\\
&=k^{2(\rho-r)}(\lambda+rk^{r-1}-\alpha)\left[4(k+1)^r-2\lambda-2rk^{r-1}+2\alpha+2\lambda\right]\\
&=k^{2(\rho-r)}(\lambda+rk^{r-1}-\alpha)\left[4(k+1)^r+2(\alpha-rk^{r-1})\right]\\
&\le 4k^{2\rho-r}(\lambda+rk^{r-1}-\alpha),
\end{aligned}$$

where we used the fact that eventually $\lambda+rk^{r-1}-\alpha\le 0$ (argue like at the beginning of the proof), together with $4(k+1)^r\ge 4k^r$ and $\alpha-rk^{r-1}>0$ for large enough k.

Line (84): we notice that

$$k^{2(\rho-r)}\left[-2\theta(k+1)^rk^{2r}+2\theta(\lambda+rk^{r-1}-\alpha)k^{2r}\beta_{k-1}-2\lambda\theta k^{2r}\beta_{k-1}\right]-2\lambda\theta k^{2\rho}\beta_{k-1}$$

is the sum of nonpositive terms for large enough k. Since V is monotone, $\langle z^{k+1}-z^k, V(z^{k+1})-V(z^k)\rangle\ge 0$ for every k.

Line (86): it holds

$$k^{2(\rho-r)}\left[-\theta^2(k+1)^{2r}k^{2r}\beta_k\beta_{k-1}+\theta k^{2r}\eta_k\beta_{k-1}\right]+\theta^2(k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}=\theta k^{2\rho}\beta_{k-1}\eta_k.$$

This proves the lemma. □

In the following lemma, we work with the coefficients attached to $\|z^{k+1}-z_*\|^2$, $\langle z^{k+1}-z_*, z^{k+1}-z^k\rangle$, $\langle z^{k+1}-z^k, V(z^{k+1})\rangle$ and $\|V(z^{k+1})\|^2$, which require further analysis. The proof is quite technical and lengthy; for the sake of readability, we have moved it to the Appendix.

Lemma 3.5. *Consider the upper bound for $\mathcal{E}_\lambda^{k+1}-\mathcal{E}_\lambda^k$ provided by Lemma 3.4. The following hold for large enough k:*

(i) *The coefficient attached to $\|z^{k+1}-z_*\|^2$ is bounded above by*

$$4\lambda(\rho-r)k^{2(\rho-r)-1}+O_k,$$

where $O_k=\mathcal{O}(k^{2(\rho-r)+r-2})$ as $k\to+\infty$.

(ii) *The coefficient attached to $\langle z^{k+1}-z_*, z^{k+1}-z^k\rangle$ can be written as $2P_k$, with $P_k=\mathcal{O}(k^{2\rho-r-2})$ as $k\to+\infty$.*

(iii) *The coefficient attached to $\langle z^{k+1}-z^k, V(z^{k+1})\rangle$ can be written as*

$$k^{2(\rho-r)}\left\{2k^{2r}\left\{\left[2\theta(\lambda+rk^{r-1}-\alpha)+(\theta\alpha+\theta rk^{r-1}-2)\right]\beta_k+\theta k^r(\beta_k-\beta_{k-1})\right\}+2(Q_{1,k}+Q_{2,k})\beta_k\right\},$$

where $|Q_{1,k}|=\mathcal{O}(k^r)$ and $|Q_{2,k}|=\mathcal{O}(k^{3r-1})$ as $k\to+\infty$.

(iv) *The coefficient attached to $\|V(z^{k+1})\|^2$ is bounded above by*

$$k^{2(\rho-r)}\left[-2\delta\theta^2k^{3r}\beta_k^2+R_k\beta_k^2\right],$$

where $R_k=o(k^{3r})$ as $k\to+\infty$.

With these two intermediate lemmas, we can prove the first theorem.

Proof of Theorem 3.3. According to Lemma 3.4 and Lemma 3.5, for large enough k we may write

$$\begin{aligned}
&\mathcal{E}_\lambda^{k+1}-\mathcal{E}_\lambda^k\\
&\leq\left[\lambda(\rho-r)k^{2(\rho-r)-1}+O_k\right]\|z^{k+1}-z_*\|^2\\
&\quad\left.\begin{aligned}&+3\lambda(\rho-r)k^{2(\rho-r)-1}\|z^{k+1}-z_*\|^2\\&+2P_k\langle z^{k+1}-z_*, z^{k+1}-z^k\rangle+\frac{1}{3}k^{2\rho-r}(\lambda+rk^{r-1}-\alpha)\|z^{k+1}-z^k\|^2\end{aligned}\right\} \quad (88)\\
&\quad+4\lambda k^{2\rho-r}\left[(2r\theta k^{r-1}-1)\beta_k+\theta k^r(\beta_k-\beta_{k-1})\right]\langle z^{k+1}-z_*, V(z^{k+1})\rangle\\
&\quad+\frac{2}{3}k^{2\rho-r}(\lambda+rk^{r-1}-\alpha)\|z^{k+1}-z^k\|^2+k^{2(\rho-r)}\left[-\frac{1}{3}\delta\theta^2k^{3r}+R_k\right]\beta_k^2\|V(z^{k+1})\|^2\\
&\quad\left.\begin{aligned}&+3k^{2(\rho-r)}k^r(\lambda+rk^{r-1}-\alpha)\|z^{k+1}-z^k\|^2\\&+2k^{2(\rho-r)}\Big\{k^{2r}\left\{\left[2\theta(\lambda+rk^{r-1}-\alpha)+(\theta\alpha+\theta rk^{r-1}-2)\right]\beta_k+\theta k^r(\beta_k-\beta_{k-1})\right\}\\&\qquad+(Q_{1,k}+Q_{2,k})\beta_k\Big\}\langle z^{k+1}-z^k, V(z^{k+1})\rangle-\frac{4}{3}\delta\theta^2k^{2(\rho-r)}k^{3r}\beta_k^2\|V(z^{k+1})\|^2\end{aligned}\right\} \quad (89)
\end{aligned}$$

$$\left. \begin{aligned} &- \frac{1}{3}\delta\theta^2 k^{2(\rho-r)} k^{3r} \beta_k^2 \|V(z^{k+1})\|^2 + \theta k^{2(\rho-r)} k^{2r} \beta_{k-1} \eta_k \langle V(z^{k+1}), V(z^{k+1}) - V(z^k) \rangle \\ &- k^{2(\rho-r)} \frac{\theta^2}{2} \Big[k^{4r} \beta_{k-1}^2 + (k+1)^{2r} k^{2r} \beta_k \beta_{k-1} \Big] \|V(z^{k+1}) - V(z^k)\|^2. \end{aligned} \right\} \tag{90}$$

We will use Lemma A.2 to show that

- (88) is nonpositive;
- There exist λ such that $0 < \lambda < \alpha$ (when $r \in (0,1)$) or $0 < \lambda < \alpha - 1$ (when $r = 1$) and such that (89) is nonpositive;
- (90) is nonpositive.

To show that (88) is nonpositive, first assume that $r \in (0,1)$. Take

$$B := P_k, \quad A := 3\lambda(\rho - r) k^{2(\rho-r)-1}, \quad C := \frac{1}{3} k^{2\rho-r} (\lambda + r k^{r-1} - \alpha).$$

According to Lemma 3.5, we have $P_k^2 = \mathcal{O}(k^{4\rho-2r-4})$ as $k \to +\infty$. It follows that for large enough k we have

$$B^2 - AC = P_k^2 - \lambda(r-\rho)(\alpha - r k^{r-1} - \lambda) k^{4\rho-3r-1} \leq 0,$$

since $4\rho - 2r - 4 < 4\rho - 3r - 1$. If $r = 1$, the only term of (88) which remains is $\frac{1}{3}k(\lambda + 1 - \alpha)\|z^{k+1} - z^k\|^2 \leq 0$.

Before addressing (89), we deal with (90), since it is simpler. We factor out $k^{2(\rho-r)}$ and take

$$B := \frac{\theta}{2} k^{2r} \beta_{k-1} \eta_k, \quad A := -\frac{1}{3}\delta\theta^2 k^{3r} \beta_k^2, \quad C := -\frac{\theta^2}{2} \Big[k^{4r} \beta_{k-1}^2 + (k+1)^{2r} k^{2r} \beta_k \beta_{k-1} \Big].$$

Therefore,

$$\begin{aligned} B^2 - AC &= \frac{\theta^2}{4} k^{4r} \beta_{k-1}^2 \eta_k^2 - \frac{1}{6}\delta\theta^4 k^{3r} \beta_k^2 \Big[k^{4r} \beta_{k-1}^2 + (k+1)^{2r} k^{2r} \beta_k \beta_{k-1} \Big] \\ &\leq \frac{\theta^2}{4} M_\beta^2 k^{4r} q_k^2 \beta_{k-1}^4 - \frac{1}{3}\delta\theta^4 k^{7r} \beta_{k-1}^4 \\ &= \left[\frac{\theta^2}{4} M_\beta^2 k^{4r} q_{k-1}^2 - \frac{1}{3}\delta\theta^4 k^{7r} \right] \beta_{k-1}^4, \end{aligned}$$

where we used the bound (117) we had on $|\eta_k|$, as well as (70). We had established that $q_k = \mathcal{O}(k^r)$ as $k \to +\infty$, which gives $k^{4r} q_k^2 = \mathcal{O}(k^{6r})$ as $k \to +\infty$. Thus, the term between square brackets becomes negative for large k, which gives $B^2 - AC \leq 0$.

Now, we focus on (89). Again, factor out $k^{2(\rho-r)}$ and define

$$\varepsilon_k := \alpha - r k^{r-1} - \lambda, \quad c_k := \theta\alpha + \theta r k^{r-1} - 2.$$

According to our assumptions, for large enough k we have $\varepsilon_k > 0$ and $c_k \geq 0$. With this notation, we set

$$B := k^{2r} \Big[(-2\theta\varepsilon_k + c_k)\beta_k + \theta k^r (\beta_k - \beta_{k-1}) \Big] + (Q_{1,k} + Q_{2,k})\beta_k, \quad A := -3k^r \varepsilon_k, \quad C := -\frac{4}{3}\delta\theta^2 k^{3r} \beta_k^2.$$

First of all, notice that

$$\begin{aligned}B^2 &= k^{4r}\Big[(-2\theta\varepsilon_k + c_k)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]^2 + (Q_{1,k} + Q_{2,k})^2\beta_k^2 \\ &\quad + 2k^{2r}\Big[(-2\theta\varepsilon_k + c_k)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big](Q_{1,k} + Q_{2,k})\beta_k \\ &= k^{4r}\Big[(-2\theta\varepsilon_k + c_k)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]^2 \\ &\quad + \left\{2k^{2r}\left[(-2\theta\varepsilon_k + c_k) + \theta k^r\frac{\beta_k - \beta_{k-1}}{\beta_k}\right](Q_{1,k} + Q_{2,k}) + (Q_{1,k} + Q_{2,k})^2\right\}\beta_k^2 \\ &= k^{4r}\Big[(-2\theta\varepsilon_k + c_k)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]^2 + S_k\beta_k^2,\end{aligned}$$

where S_k is the term between curly brackets. In Lemma 3.5 we had shown that $Q_{1,k} = \mathcal{O}(k^r)$ and $Q_{2,k} = \mathcal{O}(k^{3r-1})$ as $k \to +\infty$, which gives $S_k = \mathcal{O}(k^{\max\{3r,5r-1\}})$ as $k \to +\infty$. With these observations, we proceed:

$$\begin{aligned}&B^2 - AC - S_k\beta_k^2 \\ &\leq k^{4r}\Big[(-2\theta\varepsilon_k + c_k)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]^2 - 4\delta\theta^2k^{4r}\varepsilon_k\beta_k^2 \\ &= k^{4r}\Big[(-2\theta\varepsilon_k + c_k)^2\beta_k^2 + 2\theta(-2\theta\varepsilon_k + c_k)\beta_k k^r(\beta_k - \beta_{k-1}) + \theta^2k^{2r}(\beta_k - \beta_{k-1})^2\Big] - 4\delta\theta^2k^{4r}\varepsilon_k\beta_k^2 \\ &\leq k^{4r}\Big[(4\theta^2\varepsilon_k^2 - 4\theta c_k\varepsilon_k + c_k^2)\beta_k^2 + 2c_k(1 - 2r\theta k^{r-1} - \delta\theta)\beta_k^2 + (1 - 2r\theta k^{r-1} - \delta\theta)^2\beta_k^2\Big] \\ &\quad - 4\delta\theta^2k^{4r}\varepsilon_k\beta_k^2 \\ &= k^{4r}\beta_k^2\left\{4\theta^2\varepsilon_k^2 - 4\theta(\delta\theta + c_k)\varepsilon_k + \Big[c_k + (1 - 2r\theta k^{r-1} - \delta\theta)\Big]^2\right\} \\ &= k^{4r}\beta_k^2\,p_k, \qquad (91)\end{aligned}$$

where p_k is the term inside curly brackets. We will distinguish between the cases $r \in (0, 1)$ and $r = 1$.

Case $r \in (0, 1)$: define

$$c := \theta\alpha - 2, \quad \varepsilon := \alpha - \lambda, \quad p(t) := 4\theta^2t^2 - 4\theta(c + \delta\theta)t + (c + 1 - \delta\theta)^2.$$

Notice that $c_k \to c$, $\varepsilon_k \to \varepsilon$ and $p_k \to p(\varepsilon)$ as $k \to +\infty$. The discriminant of the quadratic equation $p(t) = 0$ reads

$$\Delta = 16\theta^2\Big[(c + \delta\theta)^2 - (c + 1 - \delta\theta)^2\Big].$$

Notice that δ was chosen such that $\delta < \frac{1}{\theta}$. The growth assumption (16) tells us it can also be chosen such that $\frac{1}{\theta} < 2\delta$, thus giving $\delta\theta > 1 - \delta\theta$ and $c + \delta\theta > c + 1 - \delta\theta > 0$. Therefore, $\Delta > 0$ and p has two distinct roots: setting $\tilde{\Delta} := \frac{1}{8\theta^2}\sqrt{\Delta}$, these two roots are

$$\underline{\varepsilon} = \frac{c + \delta\theta}{2\theta} - \tilde{\Delta}, \quad \overline{\varepsilon} = \frac{c + \delta\theta}{2\theta} + \tilde{\Delta}.$$

The midpoint between the two roots is given by $\varepsilon_{r,1} := \frac{c+\delta\theta}{2\theta}$, which fulfills

$$0 < \varepsilon_{r,1} = \frac{c + \delta\theta}{2\theta} \leq \frac{c + 1}{2\theta} = \frac{\theta\alpha - 1}{2\theta} = \frac{\alpha}{2} - \frac{1}{2\theta} < \alpha.$$

Choose $\varepsilon_{r,2} > 0$ such that $\varepsilon_{r,1} - \tilde{\Delta} < \varepsilon_{r,2} < \varepsilon_{r,1}$. Set $\lambda_{r,i} := \alpha - \varepsilon_{r,i}$, $i = 1, 2$. As we pointed out previously, we have $p_k \to p(\varepsilon_{r,i})$ as $k \to +\infty$. Therefore, for large enough k, we arrive at

$$p_k < \frac{p(\varepsilon_{r,i})}{2} = \frac{p(\alpha - \lambda_{r,i})}{2} < 0 \quad \text{for } i = 1, 2.$$

Going back to (91), this gives

$$B^2 - AC \leq k^{4r}\beta_k^2 \, \frac{p(\alpha - \lambda_{r,i})}{2} + S_k\beta_k^2 < 0 \quad \text{for large enough } k \text{ and } i = 1, 2,$$

since $S_k = \mathcal{O}(k^{\max\{3r, 5r-1\}})$ as $k \to +\infty$ and $4r > 5r - 1$ when $r < 1$.

Case $r = 1$: here, we have

$$\varepsilon_k \equiv \varepsilon := \alpha - 1 - \lambda, \quad \text{and} \quad c_k \equiv c := \theta\alpha + \theta - 2.$$

In this case, we define

$$p(t) := 4\theta^2 t^2 - 4\theta(c + \delta\theta)t + (c + 1 - 2\theta - \delta\theta)^2.$$

With this definition, p_k is nothing else than $p(\varepsilon)$ for every k. The discriminant of the quadratic equation $p(t) = 0$ the reads

$$\Delta = 16\theta^2\Big[(c + \delta\theta)^2 - (c + 1 - 2\theta - \delta\theta)^2\Big],$$

and again, it is positive provided $\delta < \frac{1}{\theta} - 2$ is chosen such that $\frac{1}{\theta} - 2 < 2\delta$, which can be done thanks to the growth condition (16). Define $\tilde{\Delta}$ and the midpoint $\varepsilon_{1,1}$ as before. Now,

$$0 < \varepsilon_{1,1} = \frac{c + \delta\theta}{2\theta} \leq \frac{c + 1 - 2\theta}{2\theta} = \frac{\theta\alpha - \theta - 1}{2\theta} = \frac{\alpha - 1}{2} - \frac{1}{2\theta} < \alpha - 1.$$

Choose $\varepsilon_{1,2} > 0$ such that $\varepsilon_{1,1} - \tilde{\Delta} < \varepsilon_{1,2} < \varepsilon_{1,1}$ and define $\lambda_{1,i} = \alpha - 1 - \varepsilon_{1,i}$, $i = 1, 2$. Going back to (91), this gives

$$B^2 - AC \leq k^4\beta_k^2 \, p(\alpha - 1 - \lambda_{1,i}) + S_k\beta_k^2 < 0 \quad \text{for large enough } k \text{ and } i = 1, 2,$$

since $p(\alpha - 1 - \lambda_{1,i}) = p(\varepsilon_{1,i}) < 0$ and $Q_{2,k} \equiv 0$ when $r = 1$, giving $S_k = \mathcal{O}(k^3)$ as $k \to +\infty$.

To summarize, we have shown that (88), (89), and (90) are nonpositive. Going back to these lines, we have now

$$\begin{aligned}
\mathcal{E}_{\lambda_{r,i}}^{k+1} - \mathcal{E}_{\lambda_{r,i}}^{k} \leq{} & \Big[\lambda_{r,i}(\rho - r)k^{2(\rho-r)-1} + O_k\Big]\|z^{k+1} - z_*\|^2 \\
& + 4\lambda_{r,i}k^{2\rho-r}\Big[(2r\theta k^{r-1} - 1)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]\langle z^{k+1} - z_*, V(z^{k+1})\rangle \\
& + \frac{2}{3}k^{2\rho-r}(\lambda_{r,i} + rk^{r-1} - \alpha)\|z^{k+1} - z^k\|^2 + k^{2(\rho-r)}\left[-\frac{1}{3}\delta\theta^2 k^{3r} + R_k\right]\beta_k^2\|V(z^{k+1})\|^2
\end{aligned}$$

If $r \in (0, 1)$, then $2(\rho - r) - 1 > 2(\rho - r) + r - 2$. Recall that $O_k = k^{2(\rho-r)+r-2}$ as $k \to +\infty$, thus the term with $\|z^{k+1} - z_*\|^2$ eventually becomes nonpositive. If $r = 1$, the $\|z^{k+1} - z_*\|^2$ term vanishes. In both cases the $\|V(z^{k+1})\|^2$ eventually becomes nonpositive since we had $R_k = o(k^{3r})$ as $k \to +\infty$. The $\langle z^{k+1} - z_*, V(z^{k+1})\rangle$ term is nonpositive according to (68), and the $\|z^{k+1} - z^k\|^2$ term is evidently nonpositive. All in all, we have

$$\mathcal{E}_{\lambda_{r,i}}^{k+1} - \mathcal{E}_{\lambda_{r,i}}^{k} \leq 0$$

for k large enough, say $k \geq K \geq k_0$. This means that $(\mathcal{E}^k_{\lambda_{r,i}})_{k\geq K}$ is monotonically decreasing, and thus

$$0 \leq \mathcal{E}^k_{\lambda_{r,i}} \leq \mathcal{E}^K_{\lambda_{r,i}} \text{ for } k \geq K,\ r \in (0,1] \text{ and } i = 1,2. \tag{92}$$

In order to conclude, we separate again the cases $r \in (0,1)$ and $r = 1$.

Case $r \in (0,1)$: according to the definition of the discrete energy function (73)-(76), (92), the fact that $(\beta_k)_{k\geq 1}$ is nondecreasing and (70), for $k \geq K$ we have

$$\frac{\theta^2}{2} k^{2(\rho+r)} \frac{\beta_{k-1}^2}{M_\beta} \|V(z^k)\|^2 \leq \frac{\theta^2}{2} k^{2(\rho+r)} \frac{\beta_k^2}{M_\beta} \|V(z^k)\|^2 \leq \frac{\theta^2}{2} (k+1)^{2r} k^{2\rho} \beta_k \beta_{k-1} \|V(z^k)\|^2 \leq \mathcal{E}^K_{\lambda_{r,1}} \tag{93}$$

and

$$2\lambda_{r,1}\theta k^{2\rho} \frac{\beta_k}{M_\beta} \langle z^{k+1} - z_*, V(z^k)\rangle \leq 2\lambda_{r,1}\theta k^{2\rho} \beta_{k-1} \langle z^k - z_*, V(z^k)\rangle \leq \mathcal{E}^K_{\lambda_{r,1}}, \tag{94}$$

from which we deduce

$$\langle z^k - z_*, V(z^k)\rangle \leq \frac{M_\beta \mathcal{E}^K_{\lambda_{r,1}}}{2\lambda_{r,1}\theta} \cdot \frac{1}{k^{2\rho}\beta_k} \quad \text{and} \quad \|V(z^k)\| \leq \frac{\sqrt{2M_\beta \mathcal{E}^K_{\lambda_{r,1}}}}{\theta} \cdot \frac{1}{k^{\rho+r}\beta_k}. \tag{95}$$

Furthemore, say that for $k \geq K$ we have $(2\rho - r)k^{r-1} < \xi < \alpha - \lambda_{r,1}$. Going back to (74), this means that for $k \geq K$ it holds

$$2\lambda_{r,1} k^{2(\rho-r)} (\alpha - \xi - \lambda_{r,1}) \|z^k - z_*\|^2 \leq 2\lambda_{r,1} k^{2(\rho-r)} (\alpha - (2\rho - r)k^{r-1}) - \lambda_{r,1}) \|z^k - z_*\|^2 \leq \mathcal{E}^K_{\lambda_{r,1}}$$

and thus

$$2\lambda_{r,1} t^{\rho-r} \|z^k - z_*\| \leq \sqrt{\frac{2\mathcal{E}^K_{\lambda_{r,1}}}{(\alpha - \xi - \lambda_{r,1})}}.$$

Combining this together with (73), (92), (93) and (95) yields

$$\begin{aligned} 2k^\rho \|z^k - z^{k-1}\| &\leq 2\lambda_{r,1} k^{\rho-r} \|z^k - z_*\| + \theta k^{\rho+r} \beta_{k-1} \|V(z^k)\| \\ &\quad + \left\| 2\lambda_{r,1} k^{\rho-r} (z - z_*) + 2k^\rho (z^k - z^{k-1}) + \theta k^{\rho+r} \beta_{k-1} V(z^k) \right] \Big\| \\ &\leq \sqrt{\frac{2\mathcal{E}^K_{\lambda_{r,1}}}{(\alpha - \xi - \lambda_{r,1})}} + \sqrt{2M_\beta \mathcal{E}^K_{\lambda_{r,1}}} + \sqrt{2\mathcal{E}^K_{\lambda_{r,1}}}. \end{aligned}$$

All in all, we have obtained that, as $k \to +\infty$,

$$\|V(z^k)\| = \mathcal{O}\left(\frac{1}{k^{\rho+r}\beta_k}\right), \quad \langle z^k - z_*, V(z^k)\rangle = \mathcal{O}\left(\frac{1}{k^{2\rho}\beta_k}\right), \quad \|z^k - z^{k-1}\| = \mathcal{O}\left(\frac{1}{k^\rho}\right). \tag{96}$$

Case $r = 1$: going back to (74) we have, for $k \geq K$,

$$\|z^k - z_*\| \leq \sqrt{\frac{\mathcal{E}^K_{\lambda_{1,1}}}{2\lambda_{1,1}(\alpha - 1 - \lambda_{1,1})}},$$

which gives the boundedness of $(z^k)_{k\geq 2}$. Using the boundedness of $(\mathcal{E}^k_{\lambda_{r,1}})_{k\geq 1}$ and arguing exactly as in the case $r \in (0,1)$ we obtain

$$\|V(z^k)\| = \mathcal{O}\left(\frac{1}{k^2\beta_k}\right), \quad \langle z^k - z_*, V(z^k)\rangle = \mathcal{O}\left(\frac{1}{k^2\beta_k}\right), \quad \|z^k - z^{k-1}\| = \mathcal{O}\left(\frac{1}{k^r}\right) \tag{97}$$

as $k \to +\infty$. □

Similar to the continuous case, in the following theorem we will assume that $(z^k)_{k\geq 2}$ is bounded when $r \in (0,1)$. This will allow us to set $\rho = r$ and improve the rate of convergence results from $\mathcal{O}$ to o in (96), as well as to show weak convergence of the iterates towards a zero of V. Additionally, when $r = 1$, we will can also lift $\mathcal{O}$ to o and again show weak convergence of the iterates towards a zero of V.

Theorem 3.6. *Under the same hypotheses of Theorem 3.3, consider the following convergence rates as $k \to +\infty$:*

$$\|V(z^k)\| = o\left(\frac{1}{k^{2r}\beta_k}\right), \quad \langle z^k - z_*, V(z^k)\rangle = o\left(\frac{1}{k^{2r}\beta_k}\right) \quad \text{and} \quad \|z^k - z^{k-1}\| = o\left(\frac{1}{k^r}\right).$$

If $r = 1$ or $\Bigg(r \in (0,1)$ and $(z^k)_{k\geq 2}$ is bounded$\Bigg)$, then the above rates hold and $(z^k)_{k\geq 2}$ converges weakly to a zero of V as $k \to +\infty$.

Proof. Going back to Lemma 3.4 and 3.5, choosing $\rho \in (0,r)$ when $r \in (0,1)$ is done just so we can guarantee that term attached to $\|z^{k+1} - z_*\|^2$ eventually becomes nonpositive. We can plug $\rho = r$ in our calculations; the term with $\langle z^{k+1} - z_*, z^{k+1} - z^k\rangle$ will vanish, and the nonnegativity of (89) and (90) will not be affected since it doesn't depend on ρ. We now have

$$\begin{aligned}\mathcal{E}^{k+1}_{\lambda_{r,i}} - \mathcal{E}^{k}_{\lambda_{r,i}} \leq{}& 2\lambda_{r,i}\, r\Big[k^{r-1} - (k+1)^{r-1}\Big]\|z^{k+1} - z_*\|^2 \\ &+ 4\lambda_{r,i}k^r\Big[(2r\theta k^{r-1} - 1)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big]\langle z^{k+1} - z_*, V(z^{k+1})\rangle \qquad (98)\\ &+ k^r\left(\lambda_{r,i} + rk^{r-1} - \alpha\right)\|z^{k+1} - z^k\|^2 \qquad (99)\\ &+ \left[-\frac{1}{3}\delta\theta^2 k^{3r} + R_k\right]\beta_k^2\|V(z^{k+1})\|^2 \qquad (100)\\ \leq{}& 2\lambda_{r,i}\, r\Big[k^{r-1} - (k+1)^{r-1}\Big]\|z^{k+1} - z_*\|^2, \qquad (101)\end{aligned}$$

since for large enough k lines (98) - (100) become nonpositive (recall that $R_k = o(k^{3r})$ as $k \to +\infty$, so $R_k - \frac{1}{3}\delta\theta^2 k^{3r}$ eventually becomes negative). If $r \in (0,1)$, recall we assume that $(\|z^k - z_*\|)_{k\geq 0}$ is bounded, say $\|z^k - z_*\|^2 \leq M$ for all k. According to (101), for large k (say $k \geq K \geq 2$ suffices), we have

$$\mathcal{E}^{k+1}_{\lambda_{r,i}} - \mathcal{E}^{k}_{\lambda_{r,i}} \leq 2\lambda_{r,i}\, r\Big[k^{r-1} - (k+1)^{r-1}\Big]\|z^{k+1} - z_*\|^2 \leq 2\lambda_{r,i}\, rM\Big[k^{r-1} - (k+1)^{r-1}\Big]. \qquad (102)$$

Rearranging the previous terms gives

$$\mathcal{E}^{k+1}_{\lambda_{r,i}} + 2\lambda_{r,i}rM(k+1)^{r-1} \leq \mathcal{E}^{k}_{\lambda_{r,i}} + 2\lambda_{r,i}rMk^{r-1} \quad \forall k \geq K.$$

Therefore, the sequence $\left(\mathcal{E}^{k}_{\lambda_{r,i}} + 2\lambda_{r,i}rMk^{r-1}\right)_{k\geq K}$ is nonnegative and nonincreasing, meaning it has a limit as $k \to +\infty$. Since $2\lambda_{r,i}r(1-r)Mk^{r-1}$ also has a limit as $k \to +\infty$, we come to the following: for $\rho = r$, we have

$$0 \leq \mathcal{E}^{k}_{\lambda_{r,i}} \leq \mathcal{E}^{K}_{\lambda_{r,i}} + 2\lambda_{r,i}r(1-r)MK^{r-1} \text{ for } k \geq K,\ r \in (0,1] \text{ and } i = 1,2; \qquad (103)$$

$$\lim_{k\to+\infty}\mathcal{E}^{k}_{\lambda_{r,i}} \text{ exists for } r \in (0,1] \text{ and } i = 1,2. \qquad (104)$$

By going back to (73)-(76), setting $\rho = r$ and reasoning exactly as we did before, we arrive at

$$\|V(z^k)\| = \mathcal{O}\left(\frac{1}{k^{2r}\beta_k}\right), \quad \langle z^k - z_*, V(z^k)\rangle = \mathcal{O}\left(\frac{1}{k^{2r}\beta_k}\right), \quad \|z^k - z^{k-1}\| = \mathcal{O}\left(\frac{1}{k^r}\right) \qquad (105)$$

as $k \to +\infty$, where these rates now hold for all $r \in (0, 1]$.

For $r \in (0, 1]$, we proceed to show three summability results that we will need later. Going back to (101), choosing $i = 1$ and using (68), for $k \geq K$ we have

$$\begin{aligned}
&4\lambda_{r,1}\delta\theta k^r \beta_k \langle z^{k+1} - z_*, V(z^{k+1})\rangle + (\alpha - rk^{r-1} - \lambda_{r,1})k^r \|z^{k+1} - z^k\|^2 + \frac{1}{6}\delta\theta^2 k^{3r}\beta_k^2 \|V(z^{k+1})\|^2 \\
&\leq \mathcal{E}^k_{\lambda_{r,1}} - \mathcal{E}^{k+1}_{\lambda_{r,1}} + 2\lambda_{r,1}\Big[k^{r-1} - (k+1)^{r-1}\Big]\|z^{k+1} - z_*\|^2, \qquad (106)
\end{aligned}$$

where we used the fact that $-\frac{1}{6}\delta\theta^2 k^{3r} + R_k$ eventually becomes negative. Notice that the last line is summable. We have

(I) Using (A1) and the Cauchy-Schwarz inequality, and using that $\|V(z^k)\| = \mathcal{O}\left(\frac{1}{k^{2r}\beta_k}\right)$ as $k \to +\infty$, we obtain

$$\Big[(k+1)^r - k^r\Big]\beta_k \langle z^{k+1} - z_*, V(z^{k+1})\rangle \leq rk^{r-1}\beta_k \|z^{k+1} - z_*\|\|V(z^{k+1})\| = \mathcal{O}\left(\frac{1}{k^{1+r}}\right)$$

as $k \to +\infty$. This implies that the left-hand side is summable. Therefore,

$$\begin{aligned}
&\sum_{k=1}^{+\infty}(k+1)^r \beta_k \langle z^{k+1} - z_*, V(z^{k+1})\rangle \\
&\leq \underbrace{\sum_{k=1}^{+\infty} k^r \beta_k \langle z^{k+1} - z_*, V(z^{k+1})\rangle}_{<+\infty \text{ by } (106)} + \underbrace{\sum_{k=1}^{+\infty}\Big[(k+1)^r - k^r\Big]\beta_k \langle z^{k+1} - z_*, V(z^{k+1})\rangle}_{<+\infty} < +\infty. \qquad (107)
\end{aligned}$$

(II) Arguing like in the previous item and using $\|z^k - z^{k-1}\| = \mathcal{O}\left(\frac{1}{k^r}\right)$ as $k \to +\infty$, (106) implies

$$\sum_{k=}^{+\infty}(k+1)^r \|z^{k+1} - z^k\|^2 < +\infty. \qquad (108)$$

(III) Using (A2) and reasoning as before, (106) entails

$$\sum_{k=1}^{+\infty}(k+1)^{3r}\beta_k^2 \|V(z^{k+1})\|^2 < +\infty. \qquad (109)$$

Consider the $\lambda_{r,i}$ defined in the proof of the previous theorem. Recalling the definition (73) - (76) of our energy function, now with $\rho = r$, we can write

$$\begin{aligned}
&\mathcal{E}^k_{\lambda_{r,2}} - \mathcal{E}^k_{\lambda_{r,1}} \\
&= \frac{1}{2}\Big\|2\lambda_{r,2}(z^k - z_*) + 2k^r(z^k - z^{k-1}) + \theta k^{2r}\beta_{k-1}V(z^k)\Big\|^2 \\
&\quad - \frac{1}{2}\Big\|2\lambda_{r,1}(z^k - z_*) + 2k^r(z^k - z^{k-1}) + \theta k^{2r}\beta_{k-1}V(z^k)\Big\|^2 \\
&\quad + \Big[2\lambda_{r,2}(\alpha - rk^{r-1} - \lambda_{r,2}) - 2\lambda_{r,1}(\alpha - rk^{r-1} - \lambda_{r,1})\Big]\|z^k - z_*\|^2 \\
&\quad + 2(\lambda_{r,2} - \lambda_{r,1})\theta k^{2r}\beta_{k-1}\langle z^k - z_*, V(z^k)\rangle \\
&= \Big\langle 2\lambda_{r,2}(z^k - z_*) + 2k^r(z^k - z^{k-1}) + \theta k^{2r}\beta_{k-1}V(z^k),\ 2(\lambda_{r,2} - \lambda_{r,1})(z^k - z_*)\Big\rangle
\end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{1}{2}\|2(\lambda_{r,2} - \lambda_{r,1})(z^k - z_*)\|^2 + 2(\lambda_{r,2} - \lambda_{r,1})(\alpha - rk^{r-1})\|z^k - z_*\|^2 \\
&\quad - 2(\lambda_{r,2}^2 - \lambda_{r,1}^2)\|z^k - z_*\|^2 + \langle 2(\lambda_{r,2} - \lambda_{r,1})(z^k - z_*),\ \theta k^{2r}\beta_{k-1}V(z^k)\rangle \\
&= \Big\langle 2(\lambda_{r,2} - \lambda_{r,1})(z^k - z_*),\ 2k^r(z^k - z^{k-1}) + 2\theta k^{2r}\beta_{k-1}V(z^k)\Big\rangle \\
&\quad + 2(\lambda_{r,2} - \lambda_{r,1})(\alpha - rk^{r-1})\|z^k - z_*\|^2 \\
&\quad + \underbrace{\Big[4\lambda_{r,2}(\lambda_{r,2} - \lambda_{r,1}) - 2(\lambda_{r,2} - \lambda_{r,1})^2 - 2(\lambda_{r,2}^2 - \lambda_{r,1}^2)\Big]}_{=0}\|z^k - z_*\|^2 \\
&= 4(\lambda_{r,2} - \lambda_{r,1})\left[\frac{\alpha - rk^{r-1}}{2}\|z^k - z_*\|^2 + k^r\Big\langle z^k - z_*,\ z^k - z^{k-1} + \theta k^r\beta_{k-1}V(z^k)\Big\rangle\right].
\end{aligned}
$$

Call p_k the term between square brackets. We showed that $\lim_{k\to+\infty}\mathcal{E}^k_{\lambda_{r,i}}$ exists for $i = 1, 2$, and $\lambda_{r,2} - \lambda_{r,1} \neq 0$, thus

$$
\lim_{k\to+\infty} p_k \quad \text{exists.} \tag{110}
$$

With this at hand, we can rewrite $\mathcal{E}^k_{\lambda_{r,1}}$ as

$$
\begin{aligned}
\mathcal{E}^k_{\lambda_{r,1}} &= \frac{1}{2}\Big\|2\lambda_{r,1}(z^k - z_*) + 2k^r(z^k - z^{k-1}) + \theta k^{2r}\beta_{k-1}V(z^k)\Big\|^2 + 2\lambda_{r,1}(\alpha - rk^{r-1} - \lambda_{r,1})\|z^k - z_*\|^2 \\
&\quad + 2\lambda_{r,1}\theta^{2r}\beta_{k-1}\langle z^k - z_*, V(z^k)\rangle + \frac{\theta^2}{2}(k+1)^{2r}k^{2r}\beta_k\beta_{k-1}\|V(z^k)\|^2 \\
&= \frac{1}{2}\Bigg[4\lambda_{r,1}^2\|z^k - z_*\|^2 + 2\lambda_{r,1}k^r\Big\langle z^k - z_*,\ 2(z^k - z^{k-1}) + \theta k^r\beta_{k-1}V(z^k)\Big\rangle \\
&\qquad + k^{2r}\Big\|2(z^k - z^{k-1}) + \theta k^r\beta_{k-1}V(z^k)\Big\|^2\Bigg] \\
&\quad + \Big[2\lambda_{r,1}(\alpha - rk^{r-1}) - 2\lambda_{r,1}^2\Big]\|z^k - z_*\|^2 + 2\lambda_{r,1}\theta k^{2r}\langle z^k - z_*, V(z^k)\rangle \\
&\quad + \frac{\theta^2}{2}(k+1)^{2r}k^{2r}\beta_k\beta_{k-1}\|V(z^k)\|^2 \\
&= 4\lambda_{r,1}p_k + \frac{k^{2r}}{2}\Big\|2(z^k - z^{k-1}) + \theta k^r\beta_{k-1}V(z^k)\Big\|^2 + \frac{\theta^2}{2}k^{4r}\beta_{k-1}^2\|V(z^k)\|^2 \\
&\quad + \frac{\theta^2}{2}\Big[(k+1)^{2r}k^{2r}\beta_k\beta_{k-1} - k^{4r}\beta_{k-1}^2\Big]\|V(z^k)\|^2 \\
&= 4\lambda_{r,1}p_k + k^{2r}\Big\|z^k - z^{k-1} + \theta k^r\beta_{k-1}V(z^k)\Big\|^2 + k^{2r}\|z^k - z^{k-1}\|^2 \\
&\quad + \frac{\theta^2}{2}\Big[(k+1)^{2r}k^{2r}\beta_k\beta_{k-1} - k^{4r}\beta_{k-1}^2\Big]\|V(z^k)\|^2.
\end{aligned} \tag{111}
$$

Let us analyze (111). We have

$$
\begin{aligned}
&\quad (k+1)^{2r}k^{2r}\beta_k\beta_{k-1} - k^{4r}\beta_{k-1}^2 \\
&= k^{2r}\beta_{k-1}\Big[(k+1)^{2r}\beta_k - k^{2r}\beta_{k-1}\Big] \\
&= k^{2r}\beta_{k-1}\Big\{\Big[(k+1)^{2r} - k^{2r}\Big]\beta_k + k^{2r}(\beta_k - \beta_{k-1})\Big\} \\
&\leq k^{2r}\beta_{k-1}\Big[(2rk^{2r-1} + rk^{2r-2})\beta_k + k^r(1 - 2r\theta k^{r-1} - \delta\theta)\beta_k\Big]
\end{aligned}
$$

$$\leq (2rk^{4r-1} + rk^{4r-2})M_\beta \beta_{k-1}^2 + (1-\delta\theta)k^{3r} M_\beta \beta_{k-1}^2,$$

where we used (A3), (69) and (70). According to Theorem 3.3 and the previous inequality, we obtain that line (111) converges to zero as $k \to +\infty$. Since we know that $\lim_{k\to+\infty} p_k$ exists, we obtain the existence of $\lim_{k\to+\infty} h_k$, where

$$h_k := k^{2r}\left\|z^k - z^{k-1} + \theta k^r \beta_{k-1} V(z^k)\right\|^2 + k^{2r}\|z^k - z^{k-1}\|^2.$$

Furthermore, we have

$$\sum_{k=1}^{+\infty} \frac{h_k}{k^r} \leq 3\sum_{k=1}^{+\infty} k^r \|z^k - z^{k-1}\|^2 + 2\theta \sum_{k=1}^{+\infty} k^{3r}\beta_{k-1}^2 \|V(z^k)\|^2 < +\infty,$$

according to (108) and (80). Since $\sum_{k=1}^{+\infty} \frac{1}{k^r} = +\infty$, it must be the case that

$$\lim_{k\to+\infty} h_k = 0.$$

In particular, this implies

$$\lim_{k\to+\infty} k^{2r}\|z^k - z^{k-1}\|^2 = 0, \text{ equivalently } \|z^k - z^{k-1}\| = o\left(\frac{1}{k^r}\right) \text{ as } k \to +\infty.$$

Using again that $\lim_{k\to+\infty} h_k = 0$ together with the previous statement, (70) and the triangle inequality, gives

$$\lim_{k\to+\infty} k^{4r}\beta_{k-1}^2 \|V(z^k)\|^2 = 0, \text{ equivalently } \|V(z^k)\| = o\left(\frac{1}{k^{2r}\beta_k}\right) \text{ as } k \to +\infty.$$

This last statement, together with the boundedness of the trajectories and the Cauchy-Schwarz inequality, gives

$$\langle z^k - z_*, V(z^k)\rangle = o\left(\frac{1}{k^{2r}\beta_k}\right) \text{ as } k \to +\infty.$$

This produces the stated o rates.

We now deal with the weak convergence of the iterates. Just like in the continuous case, we use Opial's lemma (see Lemma A.5). Define, for $k \geq K$,

$$q_k := \frac{1}{2}\|z^k - z_*\|^2 + \theta \sum_{j=1}^{k} j^r \beta_{j-1}\langle z^j - z_*, V(z^j)\rangle.$$

It follows that for $k \geq K$ we have

$$\begin{aligned} q_k - q_{k-1} &= \frac{1}{2}\left[\|z^k - z_*\|^2 - \|z^{k-1} - z_*\|^2\right] + \theta k^r \beta_{k-1}\langle z^k - z_*, V(z^k)\rangle \\ &= \langle z^k - z_*, z^k - z^{k-1} + \theta k^r \beta_{k-1} V(z^k)\rangle - \frac{1}{2}\|z^k - z^{k-1}\|^2. \end{aligned}$$

We distinguish between the cases $r \in (0,1)$ and $r = 1$.

Case $r \in (0,1)$: we have

$$\alpha q_k + k^r(q_k - q_{k-1}) = \frac{\alpha}{2}\|z^k - z_*\|^2 + \langle z^k - z_*, z^k - z^{k-1} + \theta k^r \beta_{k-1} V(z^k)\rangle$$

$$+\theta\alpha\sum_{j=1}^{k} j^r\beta_{j-1}\langle z^j - z_*, V(z^j)\rangle - \frac{1}{2}\|z^k - z^{k-1}\|^2$$

$$= \frac{r}{2}k^{r-1}\|z^k - z_*\|^2 + p_k + \theta\alpha\sum_{j=1}^{k} j^r\beta_{j-1}\langle z^j - z_*, V(z^j)\rangle - \frac{1}{2}\|z^k - z^{k-1}\|^2.$$

Since $\lim_{k\to+\infty} k^{r-1} = 0$, and according to (107), (110) and Theorem 3.3, the last line has a limit as $k \to +\infty$. We use Lemma A.4 to ensure the existence of $\lim_{k\to+\infty} q_k$. Going back to the definition of q_k and again using (107), we finally obtain that

$$\lim_{k\to+\infty} \|z^k - z_*\| \quad \text{exists.}$$

Case $r = 1$: we similarly obtain

$$(\alpha - 1)q_k + k^r(q_k - q_{k-1}) = p_k + \theta(\alpha - 1)\sum_{j=1}^{k} j\beta_{j-1}\langle z^j - z_*, V(z^j)\rangle - \frac{1}{2}\|z^k - z^{k-1}\|^2,$$

and with the same reasoning as before come to the existence of $\lim_{k\to+\infty} \|z^k - z_*\|$. This verifies the first condition of Opial's Lemma.

For the second condition, assume that $\overline{z}$ is a sequential cluster point of $(z^k)_{k\geq 2}$. Therefore, we have $z^{k_n} \rightharpoonup \overline{z}$ as $n \to +\infty$ for some subsequence $(z^{k_n})_{n\geq 0}$. According to Theorem 3.3, we have $V(z^{k_n}) \to 0$ as $n \to +\infty$. Since the graph of V is sequentially closed in $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$, we conclude that

$$V(\overline{z}) = 0,$$

therefore checking the second condition of Opial's Lemma and thus ending the proof for this theorem. □

A Auxiliary results

Lemma A.1. *For $k \geq 1$, $r \in [0, 1]$ and $\sigma \leq 0$, the following hold:*

$$(k+1)^r - k^r \leq rk^{r-1}; \tag{A1}$$

$$(k+1)^{3r} - k^{3r} \leq 3rk^{3r-1} + 3rk^{3r-2} + rk^{3r-3}; \tag{A2}$$

$$(k+1)^{2r} - k^{2r} \leq 2rk^{2r-1} + rk^{2r-2}; \tag{A3}$$

$$(k+1)^{\sigma} - k^{\sigma} \leq \sigma k^{\sigma-1} + \sigma(\sigma - 1)k^{\sigma-2}; \tag{A4}$$

$$\left|(k+1)^{\sigma} - k^{\sigma}\right| \leq |\sigma|k^{\sigma-1}; \tag{A5}$$

$$\left|2rk^{2r-1} - \left[(k+1)^{2r} - k^{2r}\right]\right| \leq 2r|2r - 1|k^{2r-2}. \tag{A6}$$

Proof. We first show (A3). According to the mean value theorem, for some $\xi \in \left(k^2, (k+1)^2\right)$ we have

$$(k+1)^{2r} - k^{2r} = \left((k+1)^2\right)^r - \left(k^2\right)^r = r\xi^{r-1}\left[(k+1)^2 - k^2\right] = r\xi^{r-1}(2k+1).$$

Notice that

$$k^2 < \xi < (k+1)^2 \quad \Rightarrow \quad \frac{1}{(k+1)^2} < \frac{1}{\xi} < \frac{1}{k^2} \quad \Rightarrow \quad \frac{1}{(k+1)^{2(1-r)}} \leq \frac{1}{\xi^{1-r}} \leq \frac{1}{k^{2(1-r)}}$$

and so

$$r\xi^{r-1}(2k+1) \leq \frac{r(2k+1)}{k^{2(1-r)}} = 2rk^{2r-1} + rk^{2r-2}.$$

The inequalities (A1) and (A2) are shown in the same way. For (A6), there exists $\mu \in (k, k+1)$ such that $(k+1)^{2r} - k^{2r} = 2r\mu^{2r-1}$. Again applying the mean value theorem, there exists $\xi \in [k, \mu]$ such that

$$\left|2rk^{2r-1} - \left[(k+1)^{2r} - k^{2r}\right]\right| = \left|2rk^{2r-1} - 2r\mu^{2r-1}\right| = 2r|2r-1|\xi^{2r-2}(\mu - k) \leq 2r|2r-1|k^{2r-2},$$

since $\mu - k \leq (k+1) - k = 1$ and $2r - 2 \leq 0$. For showing (A4), again write, for some $\xi \in (k, k+1)$,

$$(k+1)^{\sigma} - k^{\sigma} = \sigma\xi^{\sigma-1}. \tag{112}$$

Arguing similarly as before, we have $(k+1)^{\sigma-1} \leq \xi^{\sigma-1} \leq k^{\sigma-1}$, thus

$$\sigma k^{\sigma-1} \leq \sigma\xi^{\sigma-1} \leq \sigma(k+1)^{\sigma-1} = \sigma\left[(k+1)^{\sigma-1} - k^{\sigma-1}\right] + \sigma k^{\sigma-1} \leq \sigma k^{\sigma-1} + \sigma(\sigma-1)k^{\sigma-2},$$

where the $\sigma(\sigma-1)k^{\sigma-2}$ is obtained following the same reasoning. Inequality (A5) is obtained by taking the absolute value of both sides of (112). □

The following elementary result is used several times in the paper.

Lemma A.2. *Let $A, B, C \in \mathbb{R}$ be such that $A \neq 0$ and $B^2 - AC \leq 0$. The following statements are true:*

(i) *if $A > 0$, then it holds*

$$A\|X\|^2 + 2B\langle X, Y\rangle + C\|Y\|^2 \geq 0 \quad \forall X, Y \in \mathcal{H};$$

(ii) *if $A < 0$, then it holds*

$$A\|X\|^2 + 2B\langle X, Y\rangle + C\|Y\|^2 \leq 0 \quad \forall X, Y \in \mathcal{H}.$$

Regarding the following lemma, the case where $r = 1$ has already appeared in the literature, but we haven't found a proof when $r \in [0, 1)$. For the sake of completeness, we provide a proof for $r \in [0, 1]$.

Lemma A.3. *Let $a > 0, r \in [0, 1]$ and $q\colon [t_0, +\infty) \to \mathbb{R}$ be a continuously differentiable function such that*

$$\lim_{t\to+\infty}\left(q(t) + \frac{t^r}{a}\dot{q}(t)\right) = \ell \in \mathbb{R}.$$

Then it holds $\lim_{t\to+\infty} q(t) = \ell$.

Proof. By considering $q(\cdot) - \ell$ if necessary, we may assume that $\ell = 0$, i.e., we need to show that $\lim_{t\to+\infty} q(t) = 0$. Define, for $t \geq t_0$,

$$F(t) := \begin{cases} \dfrac{1}{a}\exp\left(\dfrac{a}{1-r}t^{1-r}\right) & r \in [0, 1) \\ t^a & r = 1. \end{cases}$$

Therefore, for every $t \geq t_0$ we have

$$\dot{F}(t) = \begin{cases} \frac{1}{t^r} \exp\left(\frac{a}{1-r} t^{1-r}\right) & r \in [0,1) \\ a t^{a-1} & r = 1. \end{cases}$$

In any case, $\dot{F}(t) > 0$ for all $t \geq t_0$, and it holds

$$F(t) = \frac{t^r}{a} \dot{F}(t),$$

which means that

$$\dot{F}(t)\left(q(t) + \frac{t^r}{a}\dot{q}(t)\right) = \dot{F}(t)q(t) + \frac{t^r}{a}\dot{F}(t)\dot{q}(t) = \dot{F}(t)q(t) + F(t)\dot{q}(t) = \frac{d}{dt}(F(t)q(t)). \tag{113}$$

By assumption we have $q(t) + \frac{t^r}{a}\dot{q}(t) \to 0$ as $t \to +\infty$. Take $\varepsilon > 0$. Then, there exists $T \geq t_0$ such that

$$\left|q(t) + \frac{t^r}{a}\dot{q}(t)\right| \leq \varepsilon \quad \forall t \geq T.$$

Using (113), for $t \geq T$ it holds

$$\left|\frac{d}{dt}(F(t)q(t))\right| = \left|\dot{F}(t)\left(q(t) + \frac{t^r}{a}\dot{q}(t)\right)\right| \leq \varepsilon \dot{F}(t),$$

so integration from T to $t \geq T$ yields

$$\begin{aligned} |F(t)q(t) - F(T)q(T)| &= \left|\int_T^t \frac{d}{ds}(F(s)q(s))ds\right| \leq \int_T^t \left|\frac{d}{ds}(F(s)q(s))\right| ds \\ &\leq \varepsilon \int_T^t \dot{F}(s)ds = \varepsilon\big(F(t) - F(T)\big). \end{aligned}$$

It follows that

$$|F(t)q(t)| \leq |F(t)q(t) - F(T)q(T)| + |F(T)q(T)| \leq \varepsilon\big(F(t) - F(T)\big) + |F(T)q(T)|,$$

from which we deduce

$$|q(t)| \leq \varepsilon + \frac{F(T)(|q(T)| - \varepsilon)}{F(t)} \quad \forall t \geq T.$$

Now, we use the fact that $F(t) \to +\infty$ as $t \to +\infty$ to finally obtain

$$\limsup_{t \to +\infty} |q(t)| \leq \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the desired result is shown. □

The following result is the discrete counterpart of the previous lemma. Again, the case $r = 1$ has already been previously addressed, but we found no proof when $r \in [0,1)$. We provide a complete proof.

Lemma A.4. *Let $a > 0$, $r \in [0,1]$ and let $(q_k)_{k \geq 1}$ be a sequence of real numbers such that*

$$\lim_{k \to +\infty} \left[q_{k+1} + \frac{k^r}{a}(q_{k+1} - q_k)\right] = \ell \in \mathbb{R}.$$

Then it holds $\lim_{k \to +\infty} q_k = \ell$.

Proof. By taking $q_k - \ell$ instead of q_k, we may assume w.l.o.g. that $\ell = 0$. First of all, we recall a small auxiliary result. For any sequence $(a_n)_{n\geq 1}$ of nonnegative numbers, it holds

$$\prod_{k=1}^{n}(1+a_k) \geq 1 + \sum_{k=1}^{n} a_k. \tag{114}$$

Now, define $F_1 = 1$, and inductively,

$$F_{k+1} = \left(1 + \frac{a}{k^r}\right) F_k.$$

We readily see that for every $n \geq 1$, the previous definition together with (114) entail

$$F_{n+1} = \prod_{k=1}^{n}\left(1 + \frac{a}{k^r}\right) \geq 1 + a\sum_{k=1}^{n}\frac{1}{k^r} \geq 1 + a\sum_{k=1}^{n}\frac{1}{k},$$

and since the right-hand side grows to $+\infty$ as $n \to +\infty$, we have $\lim_{n\to+\infty} F_n = +\infty$. Furthermore, notice that $(F_k)_{k\geq 1}$ is increasing, since $F_{k+1} = F_k + \frac{a}{k^r}F_k \geq F_k$. Additionally, we have

$$F_k = \frac{k^r}{a}(F_{k+1} - F_k),$$

which gives

$$\begin{aligned}
(F_{k+1} - F_k)\left[q_{k+1} + \frac{k^r}{a}(q_{k+1} - q_k)\right] &= (F_{k+1} - F_k)q_{k+1} + \frac{k^r}{a}(F_{k+1} - F_k)(q_{k+1} - q_k) \\
&= (F_{k+1} - F_k)q_{k+1} + F_k(q_{k+1} - q_k) \\
&= F_{k+1}q_{k+1} - F_k q_k. \qquad (115)
\end{aligned}$$

Let $\varepsilon > 0$. Then, there exists $k_0 \geq 1$ such that for $k \geq k_0$ it holds

$$\left|q_{k+1} + \frac{k^r}{a}(q_{k+1} - q_k)\right| \leq \varepsilon.$$

Multiplying both sides by $F_{k+1} - F_k \geq 0$ and using (115) yields

$$|F_{k+1}q_{k+1} - F_k q_k| = (F_{k+1} - F_k)\left|q_{k+1} + \frac{k^r}{a}(q_{k+1} - q_k)\right| \leq \varepsilon(F_{k+1} - F_k).$$

Summing the previous inequality from k_0 to $n - 1 \geq k_0$ leads to

$$\begin{aligned}
|F_n q_n - F_{k_0}q_{k_0}| &= \left|\sum_{k=k_0}^{n-1}(F_{k+1}q_{k+1} - F_k q_k)\right| \leq \sum_{k=k_0}^{n-1}|F_{k+1}q_{k+1} - F_k q_k| \\
&\leq \varepsilon\sum_{k=k_0}^{n-1}(F_{k+1} - F_k) = \varepsilon(F_n - F_{k_0}).
\end{aligned}$$

It follows that for $n \geq k_0 + 1$ we have

$$|F_n q_n| \leq |F_n q_n - F_{k_0}q_{k_0}| + |F_{k_0}q_{k_0}| \leq \varepsilon(F_n - F_{k_0}) + |F_{k_0}q_{k_0}|$$

and therefore

$$|q_n| \leq \varepsilon + \frac{F_{k_0}(|q_{k_0}| - \varepsilon)}{F_n}.$$

Since we have already established that $F_n \to +\infty$ as $n \to +\infty$, we come to

$$\limsup_{n\to+\infty}|q_n| \leq \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we conclude that $\lim_{n\to+\infty} q_n = 0$. □

The proof for the following lemma can be found in [18].

Lemma A.5 (Opial's Lemma)**.** *Let $\mathcal{H}$ be a real Hilbert space, $S \subseteq \mathcal{H}$ a nonempty set, $t_0 > 0$ and $z : [t_0, +\infty) \to \mathcal{H}$ a mapping that satisfies*

(i) *for every $z_* \in S$, $\lim_{t\to+\infty} \|z(t) - z_*\|$ exists;*

(ii) *every weak sequential cluster point of the trajectory $z(t)$ as $t \to +\infty$ belongs to S.*

Then, $z(t)$ converges weakly to an element of S as $t \to +\infty$.

B Proof of Lemma 3.5

For the sake of readability, we include the proof here.

Proof of Lemma 3.5. (i) The coefficient reads

$$\begin{aligned}2\lambda^2\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big] + \overbrace{2\lambda(2\rho-r)k^{2(\rho-r)}\Big[k^{r-1} - (k+1)^{r-1}\Big]}^{=:O_k} \\ + 2\lambda\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big]\big(\alpha - (2\rho-r)(k+1)^{r-1} - \lambda\big).\end{aligned}$$

After dropping the nonpositive term of the second line and using (A4) and (A5), the previous quantity is less or equal to

$$\begin{aligned}&2\lambda^2\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big] + |O_k| \\ &\leq 4\lambda^2(\rho-r)k^{2(\rho-r)-1} + \Big\{4\lambda^2(\rho-r)(2(\rho-r)-1)k^{2(\rho-r)-2} + 2\lambda|2\rho-r|(1-r)k^{2(\rho-r)+r-2}\Big\}.\end{aligned}$$

(ii) Recall the term in question reads

$$\begin{aligned}2P_k &= 8\lambda(r-\rho)k^{2(\rho-r)-1}k^r + 4\lambda\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big](k+1)^r \\ &= 4\lambda k^r\Big\{2(r-\rho)k^{2(\rho-r)-1} - \Big[k^{2(\rho-r)} - (k+1)^{2(\rho-r)}\Big]\Big\} + 4\lambda\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big]\Big[(k+1)^r - k^r\Big],\end{aligned}$$

which means that

$$|P_k| \leq 2\lambda k^r \cdot 2(r-\rho)(1+2(r-\rho))k^{2(\rho-r)-2} + 2\lambda\cdot 2(r-\rho)k^{2(\rho-r)-1}\cdot rk^{r-1},$$

and the right-hand side is of order $\mathcal{O}(k^{2\rho-r-2})$ as $k \to +\infty$.

(iii) We can rewrite the term as

$$\begin{aligned}\eta_k &= 2k^r\Big[(2r\theta k^{r-1} - 1)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big] - \theta\Big[(k+1)^{2r}\beta_k - k^{2r}\beta_{k-1}\Big] \\ &= 2k^r(2r\theta k^{r-1} - 1)\beta_k + 2\theta k^{2r}(\beta_k - \beta_{k-1}) - \theta\Big\{\Big[(k+1)^{2r} - k^{2r}\Big]\beta_k + k^{2r}(\beta_k - \beta_{k-1})\Big\} \\ &= 2k^r(2r\theta k^{r-1} - 1)\beta_k + \theta k^{2r}(\beta_k - \beta_{k-1}) - \theta\Big[(k+1)^{2r} - k^{2r}\Big]\beta_k. \qquad (116)\end{aligned}$$

According to our assumptions, for large enough k we have $|2r\theta k^{r-1} - 1| = 1 - 2r\theta k^{r-1} \leq 1$. Using this, together with (A3) and (69), yields

$$|\eta_k| \leq 2k^r\beta_k + k^r(1 - 2r\theta k^{r-1} - \delta\theta)\beta_k + (2r\theta k^{2r-1} + r\theta k^{2r-2})\beta_k.$$

Since $2r-2 \le 2r-1 \le r$, the previous inequality tells us that

$$|\eta_k| \le q_k \beta_k, \tag{117}$$

where $q_k = \mathcal{O}(k^r)$ as $k \to +\infty$. We rewrite the term accompanying $\langle z^{k+1} - z^k, V(z^{k+1})\rangle$ as

$$k^{2(\rho-r)}\Big\{2\eta_k(k+1)^r + 2\theta(\lambda + rk^{r-1} - \alpha)(k+1)^{2r}\beta_k - 2(\lambda + rk^{r-1} - \alpha)\eta_k + 2\lambda\theta k^{2r}\beta_{k-1}\Big\} \tag{118}$$

$$+ 2\theta\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big](k+1)^{3r}\beta_k. \tag{119}$$

We first focus on (118), namely, the term inside curly brackets:

$$\begin{aligned}
&2\eta_k(k+1)^r + 2\theta(\lambda + rk^{r-1} - \alpha)(k+1)^{2r}\beta_k - 2(\lambda + rk^{r-1} - \alpha)\eta_k + 2\lambda\theta k^{2r}\beta_{k-1} \\
&= 2\Big[(k+1)^r - k^r - (\lambda + rk^{r-1} - \alpha)\Big]\eta_k + 2k^r\eta_k \\
&\quad + 2\theta\Big[(k+1)^{2r} - k^{2r}\Big](\lambda + rk^{r-1} - \alpha)\beta_k + 2\theta k^{2r}(\lambda + rk^{r-1} - \alpha)\beta_k \\
&\quad + 2\lambda\theta k^{2r}(\beta_{k-1} - \beta_k) + 2\lambda\theta k^{2r}\beta_k \\
&= 2k^r\eta_k + 2\theta k^{2r}(\lambda + rk^{r-1} - \alpha)\beta_k + 2\lambda\theta k^{2r}\beta_k \\
&\quad + 2\Big[(k+1)^r - k^r - (\lambda + rk^{r-1} - \alpha)\Big]\eta_k + 2\theta\Big[(k+1)^{2r} - k^{2r}\Big](\lambda + rk^{r-1} - \alpha)\beta_k \\
&\quad + 2\lambda\theta k^{2r}(\beta_{k-1} - \beta_k) \\
&= 2k^r\Big\{2k^r(r\theta k^{r-1} - 1)\beta_k + 2r\theta k^{2r-1}\beta_k + \theta k^{2r}(\beta_k - \beta_{k-1}) - \theta\Big[(k+1)^{2r} - k^{2r}\Big]\beta_k\Big\} \\
&\quad + 2\theta k^{2r}(\lambda + rk^{r-1} - \alpha)\beta_k + 2\lambda\theta k^{2r}\beta_k \\
&\quad + 2\Big[(k+1)^r - k^r - (\lambda + rk^{r-1} - \alpha)\Big]\eta_k + 2\theta\Big[(k+1)^{2r} - k^{2r}\Big](\lambda + rk^{r-1} - \alpha)\beta_k \\
&\quad + 2\lambda\theta k^{2r}(\beta_{k-1} - \beta_k) \\
&= 2k^r\Big[2k^r(r\theta k^{r-1} - 1)\beta_k + \theta k^{2r}(\beta_k - \beta_{k-1})\Big] + 2\theta k^{2r}(\lambda + rk^{r-1} - \alpha)\beta_k + 2\lambda\theta k^{2r}\beta_k \\
&\quad + 2\theta k^r\Big\{2rk^{2r-1} - \Big[(k+1)^{2r} - k^{2r}\Big]\Big\}\beta_k \\
&\quad + 2\Big[(k+1)^r - k^r - (\lambda + rk^{r-1} - \alpha)\Big]\eta_k + 2\theta\Big[(k+1)^{2r} - k^{2r}\Big](\lambda + rk^{r-1} - \alpha)\beta_k \\
&\quad + 2\lambda\theta k^{2r}(\beta_{k-1} - \beta_k) \\
&= 2k^{2r}\Big[2(r\theta k^{r-1} - 1)\beta_k + \theta k^r(\beta_k - \beta_{k-1}) + \theta(\lambda + rk^{r-1} - \alpha)\beta_k + \lambda\theta\beta_k\Big] && (120)\\
&\quad + 2\theta k^r\Big\{2rk^{2r-1} - \Big[(k+1)^{2r} - k^{2r}\Big]\Big\}\beta_k + 2\Big[(k+1)^r - k^r - (\lambda + rk^{r-1} - \alpha)\Big]\frac{\eta_k}{\beta_k}\cdot\beta_k && (121)\\
&\quad + 2\theta\Big[(k+1)^{2r} - k^{2r}\Big](\lambda + rk^{r-1} - \alpha)\beta_k + 2\lambda\theta k^{2r}\frac{\beta_{k-1} - \beta_k}{\beta_k}\cdot\beta_k && (122)
\end{aligned}$$

Line (120) reads

$$\begin{aligned}
&2k^{2r}\Big[(2r\theta k^{r-1} - 2 + \lambda\theta + \theta rk^{r-1} - \theta\alpha + \lambda\theta)\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big] \\
&= 2k^{2r}\Big\{\Big[-2\theta(\alpha - rk^{r-1} - \lambda) + (\theta\alpha + \theta rk^{r-1} - 2)\Big]\beta_k + \theta k^r(\beta_k - \beta_{k-1})\Big\}.
\end{aligned}$$

For lines (121) and (122), we factor out $2\beta_k$ and write these lines as $2Q_{1,k}\beta_k$. We use (A6), (A1), the fact that $|\lambda + rk^{r-1} - \alpha| = \alpha - rk^{r-1} - \lambda \le \alpha$ for large enough k, (117), (A3) and (69) to obtain

$$|Q_{1,k}| \le \theta k^r\Big|2rk^{2r-1} - \Big[(k+1)^{2r} - k^{2r}\Big]\Big| + \Big|(k+1)^r - k^r - (\lambda + rk^{r-1} - \alpha)\Big|\left|\frac{\eta_k}{\beta_k}\right|$$

$$
\begin{aligned}
&+ \theta\Big|(k+1)^{2r} - k^{2r}\Big|\big|\lambda + rk^{r-1} - \alpha\big| + \lambda\theta k^{2r}\left|\frac{\beta_{k-1}-\beta_k}{\beta_k}\right| \\
&\leq 2r|2r-1|\theta k^{3r-2} + \big|rk^{r-1} + \alpha - rk^{r-1} - \lambda\big| q_k \\
&\quad + \big(2r\theta k^{2r-1} + r\theta k^{2r-2}\big)\big(\alpha - rk^{r-1} - \lambda\big) + \lambda k^r\big(1 - 2r\theta k^{r-1} - \delta\theta\big) \\
&\leq 2r|2r-1|\theta k^{3r-2} + (\alpha-\lambda)q_k + \alpha\big(2r\theta k^{2r-1} + r\theta k^{2r-2}\big) + \lambda k^r(1-\delta\theta).
\end{aligned}
$$

The right-hand side is of order $\mathcal{O}(k^r)$ as $k \to +\infty$. Now, line (119): we write it as $2k^{2(\rho-r)}Q_{2,k}$ and we have, according to (A5),

$$
\begin{aligned}
|Q_{2,k}| &\leq \frac{\theta}{k^{2(\rho-r)}}\Big|(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big|(k+1)^{3r} \leq \frac{1}{k^{2(\rho-r)}} \cdot 2\theta(r-\rho)k^{2(\rho-r)-1}(k+1)^{3r} \\
&\leq 2\theta(r-\rho)k^{-1}\left[k^{3r} + 3k^{2r} + 3k^r + 1\right]
\end{aligned}
$$

and this last term is of order $\mathcal{O}\left(k^{3r-1}\right)$ as $k \to +\infty$.

(iv) First, notice that

$$
\begin{aligned}
(k+1)^{2\rho} &= (k+1)^{2\rho} - k^{2(\rho-r)}(k+1)^{2r} + k^{2(\rho-r)}(k+1)^{2r} \\
&= (k+1)^{2r}\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big] + k^{2(\rho-r)}(k+1)^{2r} \\
&\leq k^{2(\rho-r)}(k+1)^{2r}.
\end{aligned}
$$

It follows that the coefficient attached to $\|V(z^{k+1})\|^2$ is less or equal than

$$
\begin{aligned}
& k^{2(\rho-r)}\left[\theta(k+1)^{2r}\eta_k\beta_k - \frac{1}{2}\eta_k^2\right] + \frac{\theta^2}{2}\Big[(k+2)^{2r}k^{2(\rho-r)}(k+1)^{2r}\beta_{k+1}\beta_k - (k+1)^{2r}k^{2\rho}\beta_k\beta_{k-1}\Big] \\
& + \frac{\theta^2}{2}\Big[(k+1)^{2(\rho-r)} - k^{2(\rho-r)}\Big](k+1)^{4r}\beta_k^2 \\
&\leq k^{2(\rho-r)}\left\{\theta(k+1)^{2r}\eta_k\beta_k - \frac{1}{2}\eta_k^2 + \frac{\theta^2}{2}\Big[(k+2)^{2r}(k+1)^{2r}\beta_{k+1}\beta_k - (k+1)^{2r}k^{2r}\beta_k\beta_{k-1}\Big]\right\} \\
&\leq k^{2(\rho-r)}\left\{\theta k^{2r}\beta_k\eta_k + \frac{\theta^2}{2}(k+1)^{2r}\beta_k\Big[(k+2)^{2r}\beta_{k+1} - k^{2r}\beta_{k-1}\Big]\right\}, \qquad (123)
\end{aligned}
$$

where we used the fact that for large enough k, $\eta_k \leq 0$ (a consequence of (68)) and we dropped the nonpositive term $-\frac{1}{2}\eta_k^2$. We focus on the term inside curly brackets and address the cases $r \in (0,1)$ and $r = 1$ separately.

Case $r \in (0,1)$: since the inequality (67) is strict, there exists $\tilde{\delta}$ such that $0 < \delta < \tilde{\delta} < \frac{1}{\theta}$ and such that inequalities (68), (69) and (70) hold replacing δ by $\tilde{\delta}$. Taking into account (116), the first of summand of (123) reads

$$
\begin{aligned}
\theta k^{2r}\beta_k\eta_k &= \theta k^{2r}\beta_k\Big\{2k^r\big(2r\theta k^{r-1} - 1\big)\beta_k + \theta k^{2r}(\beta_k - \beta_{k-1}) - \theta\Big[(k+1)^{2r} - k^{2r}\Big]\beta_k\Big\} \\
&\leq \theta k^{2r}\beta_k\Big[2k^r\big(2r\theta k^{r-1} - 1\big)\beta_k + k^r\big(1 - 2r\theta k^{r-1} - \tilde{\delta}\theta\big)\beta_k\Big] \\
&= \theta k^{2r}\beta_k\Big[k^r\big(2r\theta k^{r-1} - 1\big)\beta_k - \tilde{\delta}\theta k^r\beta_k\Big] \\
&= \theta(-1 - \tilde{\delta}\theta)k^{3r}\beta_k^2 + 2r\theta^2 k^{4r-1}\beta_k^2. \qquad (124)
\end{aligned}
$$

Moving on to the second summand of (123), we have

$$
(k+2)^{2r}\beta_{k+1} - k^{2r}\beta_{k-1}
$$

$$
\begin{aligned}
&= (k+2)^{2r}\beta_{k+1} - (k+1)^{2r}\beta_k + (k+1)^{2r}\beta_k - k^{2r}\beta_{k-1} \\
&= \Big[(k+2)^{2r} - (k+1)^{2r}\Big]\beta_{k+1} + (k+1)^{2r}(\beta_{k+1} - \beta_k) + \Big[(k+1)^{2r} - k^{2r}\Big]\beta_k + k^{2r}(\beta_k - \beta_{k-1}) \\
&\leq (2r(k+1)^{2r-1} + r(k+1)^{2r-2})\beta_{k+1} + (k+1)^r\left(\frac{1}{\theta} - 2r(k+1)^{r-1} - \tilde{\delta}\right)\beta_{k+1} \\
&\quad + (2rk^{2r-1} + rk^{2r-2})\beta_k + k^r\left(\frac{1}{\theta} - 2rk^{r-1} - \tilde{\delta}\right)\beta_k \\
&\leq \Big[M_\beta(2r(k+1)^{2r-1} + r(k+1)^{2r-2}) + 2rk^{2r-1} + rk^{2r-2}\Big]\beta_k \\
&\quad + (k+1)^r\left(\frac{1}{\theta} - \tilde{\delta}\right)M_\beta\beta_k + k^r\left(\frac{1}{\theta} - \tilde{\delta}\right)\beta_k \\
&\leq \left\{M_\beta(2r(k+1)^{2r-1} + r(k+1)^{2r-2}) + 2rk^{2r-1} + rk^{2r-2} + M_\beta\left(\frac{1}{\theta} - \tilde{\delta}\right)\right\}\beta_k \\
&\quad + (1 + M_\beta)\left(\frac{1}{\theta} - \tilde{\delta}\right)k^r\beta_k \\
&= r_k\beta_k + (1 + M_\beta)\left(\frac{1}{\theta} - \tilde{\delta}\right)k^r\beta_k, \qquad (125)
\end{aligned}
$$

where r_k is the term between curly brackets. Notice that $r_k = \mathcal{O}(k^{\max\{2r-1,0\}})$ as $k \to +\infty$. Here, we used (A3), (69), (70), we dropped the nonpositive terms $-2r(k+1)^{r-1}$ and $-2rk^{r-1}$ and we used the fact that $(k+1)^r \leq k^r + 1$, since $t \mapsto t^r$ is subadditive on $[0, +\infty)$. With (124) and (125) at hand, we can bound (123):

$$
\begin{aligned}
&\theta k^{2r}\beta_k\eta_k + \frac{\theta^2}{2}(k+1)^{2r}\beta_k\Big[(k+2)^{2r}\beta_{k+1} - k^{2r}\beta_{k-1}\Big] \\
&\leq \theta(-1 - \tilde{\delta}\theta)k^{3r}\beta_k^2 + 2r\theta^2k^{4r-1}\beta_k^2 + \frac{\theta^2}{2}(k+1)^{2r}\left[r_k + (1 + M_\beta)\left(\frac{1}{\theta} - \tilde{\delta}\right)k^r\right]\beta_k^2 \\
&\leq \theta k^{3r}\left[-1 + \frac{1 + M_\beta}{2} - \left(1 + \frac{1 + M_\beta}{2}\right)\tilde{\delta}\theta\right]\beta_k^2 \\
&\quad + \theta^2\left\{2rk^{4r-1} + \frac{(k+1)^{2r}}{2}r_k + k^r(2k^r + 1)\frac{1 + M_\beta}{2}\left(\frac{1}{\theta} - \tilde{\delta}\right)\right\}\beta_k^2 \\
&\leq \theta k^{3r}\left[\frac{M_\beta - 1}{2} - 2\tilde{\delta}\theta\right]\beta_k^2 + R_k\beta_k^2,
\end{aligned}
$$

where R_k is the term between curly brackets multiplied by θ^2. We used the fact that $M_\beta > 1$ for the bound $-\frac{1+M_\beta}{2} < -1$, and again the subadditivity of $t \mapsto t^r$ to obtain $(k+1)^{2r} \leq \big(k^r + 1\big)^2 = k^{2r} + 2k^r + 1$. Since $r < 1$, we have $3r > 4r - 1$, and from here we deduce that $R_k = o(k^{3r})$ as $k \to +\infty$. According to Remark 3.1, M_β can be taken as close to 1 as desired, provided k is large enough. In particular, it can be chosen such that $\frac{M_\beta - 1}{2} < 2\theta(\tilde{\delta} - \delta)$, and thus

$$
\frac{M_\beta - 1}{2} - 2\tilde{\delta}\theta < 2\theta(\tilde{\delta} - \delta) - 2\tilde{\delta}\theta = -2\delta\theta,
$$

and this is what we wanted to show.

Case $r = 1$: again, according to (116), we have

$$
\begin{aligned}
\eta_k &= 2k(2\theta - 1)\beta_k + \theta k^2(\beta_k - \beta_{k-1}) - \theta\Big[(k+1)^2 - k^2\Big]\beta_k \\
&= 2k(2\theta - 1)\beta_k + \theta k^2(\beta_k - \beta_{k-1}) - \theta(2k+1)\beta_k
\end{aligned}
$$

$$= \Big[2(\theta-1)k-\theta\Big]\beta_k + \theta k^2(\beta_k-\beta_{k-1}),$$

so using (69) yields

$$\begin{aligned} \theta k^2\beta_k\eta_k &\le \theta k^2\beta_k\Big\{\Big[2(\theta-1)k-\theta\Big]\beta_k + k(1-2\theta-\delta\theta)\beta_k\Big\} \\ &= \theta k^3\Big[2(\theta-1)+(1-2\theta-\delta\theta)\Big]\beta_k^2 - \theta^2k^2\beta_k^2. \end{aligned} \tag{126}$$

For the second summand of (123), we write

$$\begin{aligned} &(k+2)^2\beta_{k+1} - k^2\beta_{k-1} \\ &= (k+2)^2(\beta_{k+1}-\beta_k) + k^2(\beta_k-\beta_{k-1}) + \Big[(k+2)^2-k^2\Big]\beta_k \\ &= \Big[(k+1)^2+2k+3\Big](\beta_{k+1}-\beta_k) + k^2(\beta_k-\beta_{k-1}) + (4k+4)\beta_k \\ &\le \left(\frac{1}{\theta}-2-\delta\right)\Big[(k+1)\beta_{k+1}+k\beta_k\Big] + (2k+3)(\beta_{k+1}-\beta_k) + 4(k+1)\beta_k \\ &= \left(\frac{1}{\theta}-2-\delta\right)(k+1)(\beta_{k+1}-\beta_k) + (2k+3)(\beta_{k+1}-\beta_k) + \left[\left(\frac{1}{\theta}-2-\delta\right)(2k+1)+4(k+1)\right]\beta_k \\ &\le \left(\frac{1}{\theta}-2-\delta\right)^2\beta_{k+1} + \frac{2k+3}{k+1}\left(\frac{1}{\theta}-2-\delta\right)\beta_{k+1} + \left[\left(\frac{1}{\theta}-2-\delta\right)(2k+1)+4(k+1)\right]\beta_k \\ &\le \left[\left(\frac{1}{\theta}-2-\delta\right)^2 M_\beta + \frac{2k+3}{k+1}\left(\frac{1}{\theta}-2-\delta\right)M_\beta\right]\beta_k + \left[\left(\frac{1}{\theta}-2-\delta\right)(2k+1)+4(k+1)\right]\beta_k \\ &= 2\left(\frac{1}{\theta}-\delta\right)k\beta_k + \left\{\left(\frac{1}{\theta}-2-\delta\right)^2 M_\beta + \frac{2k+3}{k+1}\left(\frac{1}{\theta}-2-\delta\right)M_\beta + \frac{1}{\theta}+2-\delta\right\}\beta_k \\ &= 2\left(\frac{1}{\theta}-\delta\right)k\beta_k + r_k\beta_k, \end{aligned} \tag{127}$$

where r_k is the term between curly brackets. We have $r_k = \mathcal{O}(1)$ as $k \to +\infty$. To obtain the bounds, we repeatedly used (69) and (70). Using (126) and (127), we have a bound for (123):

$$\begin{aligned} &\theta k^{2r}\beta_k\eta_k + \frac{\theta^2}{2}(k+1)^{2r}\beta_k\Big[(k+2)^{2r}\beta_{k+1} - k^{2r}\beta_{k-1}\Big] \\ &\le \theta k^3\Big[2(\theta-1)+(1-2\theta-\delta\theta)\Big]\beta_k^2 - \theta^2k^2\beta_k^2 + \frac{\theta^2}{2}(k+1)^2\beta_k\left[2\left(\frac{1}{\theta}-\delta\right)k\beta_k + r_k\beta_k\right] \\ &\le \theta k^3\Big[2(\theta-1)+(1-2\theta-\delta\theta)+(1-\delta\theta)\Big]\beta_k^2 + \frac{\theta^2}{2}(2k+1)r_k\beta_k^2 \\ &= -2\delta\theta^2k^3\beta_k^2 + R_k\beta_k^2, \end{aligned}$$

where $R_k = \frac{\theta^2}{2}(2k+1)r_k = o(k^3)$ as $k \to +\infty$. This concludes the proof. □

References

[1] F. Álvarez. On the minimizing property of a second-order dissipative system in Hilbert spaces. *SIAM Journal on Control Optimization* 38(4), 1102–1119 (2000).

[2] F. Álvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Analysis* 9(1–2), 3–11 (2001).

[3] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics. *Journal of Optimization Theory and Applications* 193, 704–736 (2022).

[4] H. Attouch and S. C. László. Continuous Newton-like inertial dynamics for monotone inclusions. *Set-Valued and Variational Analysis* 29, 555–581 (2021).

[5] H. Attouch and P. E. Maingé. Asymptotic behavior of second order dissipative evolution equations combining potential with non-potential effects. *ESAIM Control Optimization and Calculus of Variations* 17(3), 836–857 (2011).

[6] H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators. *Mathematical Programming* 174(1–2), 391–432 (2019).

[7] H. Attouch, J. Peypouquet, and P. Redont. Fast convex minimization via inertial dynamics with Hessian driven damping. *Journal of Differential Equations* 210(10), 5734–5783 (2016).

[8] H. Attouch and B. F. Svaiter. A continuous dynamical Newton-like approach to solving monotone inclusions. *SIAM Journal on Control and Optimization* 49(2), 574–598 (2011).

[9] A. Böhm, M. Sedlmayer, E. R. Csetnek, and R. I. Boţ. Two steps at a time—taking GAN training in stride with Tseng's method. *SIAM Journal on Mathematics of Data Science* 4(2), 750–771 (2022).

[10] R. I. Boţ and E. R. Csetnek. Second order forward-backward dynamical systems for monotone inclusion problems. *SIAM Journal on Control and Optimization* 54, 1423–1443 (2016).

[11] R. I. Boţ, E. R. Csetnek, and D.-K. Nguyen. Fast Optimistic Gradient Descent Ascent (OGDA) method in continuous and discrete time. *Foundations of Computational Mathematics* (2023).

[12] R. I. Boţ and D.-K. Nguyen. Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping. *Journal of Differential Equations* 303, 369–406 (2021).

[13] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *NIPS 2014: Advances in Neural Information Processing Systems* 27, 2672–2680 (2014).

[14] X. He, R. Hu, and Y.-P. Fang. Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems. *SIAM Journal on Control and Optimization* 59(5), 3278–3301 (2021).

[15] X. He, R. Hu, and Y.-P. Fang. Inertial primal-dual dynamics with damping and scaling for linearly constrained convex optimization problems. *Applicable Analysis* 102(15), 4114–4139 (2023).

[16] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards deep learning models resistant to adversarial attacks. *ICLR 2018: International Conference on Learning Representations* (2018).

[17] S. Omidshafiei, J. Pazis, C. Amato, J. P. How, and J. Vian. Deep decentralized multi-task multi-agent reinforcement learning under partial observability. *Proceedings of the 34th International Conference on Machine Learning* PLMR 70, 2681–2690 (2017).

[18] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society* 73(4), 591–597 (July 1967).

[19] W. Su, S. Boyd, and E. Candès. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *Journal of Machine Learning Research* 17(153), 1–43 (2016).

[20] X. Zeng, J. Lei, and J. Chen. Dynamical primal-dual accelerated method with applications to network optimization. *IEEE Transactions on Automatic Control* 68(3) (2023).