
Continuous-time Analysis of Anchor Acceleration

Jaewook J. Suh
Seoul National University
jacksuhkr@snu.ac.kr

Jisun Park
Seoul National University
colleenp0515@snu.ac.kr

Ernest K. Ryu
Seoul National University
ernestryu@snu.ac.kr

Abstract

Recently, the anchor acceleration, an acceleration mechanism distinct from Nesterov’s, has been discovered for minimax optimization and fixed-point problems, but its mechanism is not understood well, much less so than Nesterov acceleration. In this work, we analyze continuous-time models of anchor acceleration. We provide tight, unified analyses for characterizing the convergence rate as a function of the anchor coefficient $\beta(t)$, thereby providing insight into the anchor acceleration mechanism and its accelerated $\mathcal{O}(1/k^2)$ -convergence rate. Finally, we present an adaptive method inspired by the continuous-time analyses and establish its effectiveness through theoretical analyses and experiments.

1 Introduction

Nesterov acceleration [51] is foundational to first-order optimization theory, but the mechanism and its convergence proof are not transparent. One approach to better understand the mechanism is the continuous-time analysis: derive an ODE model of the discrete-time algorithm and analyze the continuous-time dynamics [65, 66]. This approach provides insight into the accelerated dynamics and has led to a series of follow-up work [71, 62, 29].

Recently, a new acceleration mechanism, distinct from Nesterov’s, has been discovered. This *anchor acceleration* for minimax optimization and fixed-point problems [35, 75, 54] has been an intense subject of study, but its mechanism is understood much less than Nesterov acceleration. The various analytic techniques developed to understand Nesterov acceleration, including continuous-time analyses, have only been applied in a very limited manner [59].

Contribution. In this work, we present continuous-time analyses of anchor acceleration. The continuous-time model is the differential inclusion

$$\dot{X} \in -\mathbf{A}(X) - \beta(t)(X - X_0)$$

with initial condition $X(0) = X_0 \in \text{dom } \mathbf{A}$, maximal monotone operator \mathbf{A} , and scalar-valued function $\beta(t)$. The case $\beta(t) = \frac{1}{t}$ corresponds to the prior anchor-accelerated methods APPM [35], EAG [75], and FEG [42].

We first establish that the differential inclusion is well-posed, despite the anchor coefficient $\beta(t)$ blowing up at $t = 0$. We then provide tight, unified analyses for characterizing the convergence rate as a function of the anchor coefficient $\beta(t)$. This is the first formal and rigorous treatment of this anchored dynamics, and it provides insight into the anchor acceleration mechanism and its accelerated $\mathcal{O}(1/k^2)$ -convergence rate. Finally, we present an adaptive method inspired by the continuous-time analyses and establish its effectiveness through theoretical analyses and experiments.

1.1 Preliminaries and notation

We provide the organization for prior works in Appendix A. Here, we review standard definitions and set up the notation.

Monotone and set-valued operators. We follow the standard definitions of Bauschke and Combettes [15], Ryu and Yin [58]. For the underlying space, consider \mathbb{R}^n with standard inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. Define domain of \mathbf{A} as $\text{dom } \mathbf{A} = \{x \in \mathbb{R}^n \mid \mathbf{A}x \neq \emptyset\}$. We say \mathbf{A} is an operator on \mathbb{R}^n and write $\mathbf{A}: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ if \mathbf{A} maps a point in \mathbb{R}^n to a subset of \mathbb{R}^n . We say $\mathbf{A}: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is monotone if

$$\langle \mathbf{A}x - \mathbf{A}y, x - y \rangle \geq 0, \quad \forall x, y \in \mathbb{R}^n,$$

where the notation means that $\langle u - v, x - y \rangle \geq 0$ for all $u \in \mathbf{A}x$ and $v \in \mathbf{A}y$. For $\mu \in (0, \infty)$, say $\mathbf{A}: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is μ -strongly monotone if

$$\langle \mathbf{A}x - \mathbf{A}y, x - y \rangle \geq \mu \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

Write $\text{Gra } \mathbf{A} = \{(x, u) \mid u \in \mathbf{A}x\}$ for the graph of \mathbf{A} . An operator \mathbf{A} is maximally monotone if there is no other monotone \mathbf{B} such that $\text{Gra } \mathbf{A} \subset \text{Gra } \mathbf{B}$ properly, and is maximally μ -strongly monotone if there is no other μ -strongly monotone \mathbf{B} such that $\text{Gra } \mathbf{A} \subset \text{Gra } \mathbf{B}$ properly.

For $L \in (0, \infty)$, single-valued operator $\mathbf{T}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is L -Lipschitz if

$$\|\mathbf{T}x - \mathbf{T}y\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

Write $\mathbf{J}_{\mathbf{A}} = (\mathbf{I} + \mathbf{A})^{-1}$ for the resolvent of \mathbf{A} , while $\mathbf{I}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the identity operator. When \mathbf{A} is maximally monotone, it is well known that $\mathbf{J}_{\mathbf{A}}$ is single-valued with $\text{dom } \mathbf{J}_{\mathbf{A}} = \mathbb{R}^n$.

We say $x_* \in \mathbb{R}^n$ is a zero of \mathbf{A} if $0 \in \mathbf{A}x_*$. We say y_* is a fixed-point of \mathbf{T} if $\mathbf{T}y_* = y_*$. Write $\text{Zer } \mathbf{A}$ for the set of all zeros of \mathbf{A} and $\text{Fix } \mathbf{T}$ for the set of all fixed-points of \mathbf{T} .

Monotonicity with continuous curves. We say an operator is differentiable if it is single-valued, continuous, and differentiable as a function. If a differentiable operator \mathbf{A} is monotone and $X: [0, \infty) \rightarrow \mathbb{R}^n$ is a differentiable curve, then taking limit $h \rightarrow 0$ of

$$\frac{1}{h^2} \langle \mathbf{A}(X(t+h)) - \mathbf{A}(X(t)), X(t+h) - X(t) \rangle \geq 0$$

leads to

$$\left\langle \frac{d}{dt} \mathbf{A}(X(t)), \dot{X}(t) \right\rangle \geq 0. \quad (1)$$

Similarly if \mathbf{A} is furthermore μ -strongly monotone, then

$$\left\langle \frac{d}{dt} \mathbf{A}(X(t)), \dot{X}(t) \right\rangle \geq \mu \|\dot{X}(t)\|^2. \quad (2)$$

2 Derivation of differential inclusion model of anchor acceleration

2.1 Anchor ODE

Suppose $\mathbf{A}: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a maximal monotone operator and $\beta: (0, \infty) \rightarrow [0, \infty)$ is a twice differentiable function. Consider differential inclusion

$$\dot{X}(t) \in -\mathbf{A}(X(t)) - \beta(t)(X(t) - X_0) \quad (3)$$

with initial condition $X(0) = X_0 \in \text{dom } (\mathbf{A})$. We refer to this as the *anchor ODE*.¹ We say $X: [0, \infty) \rightarrow \mathbb{R}^n$ is a solution, if it is absolutely continuous and satisfies (3) for $t \in (0, \infty)$ almost everywhere.

Denote S as the subset of $[0, \infty)$ on which X satisfies the differential inclusion. Define

$$\tilde{\mathbf{A}}(X(t)) = -\dot{X}(t) - \beta(t)(X(t) - X_0)$$

for $t \in S$. Since $\tilde{\mathbf{A}}(X(t)) \in \mathbf{A}(X(t))$ for $t \in S$, we say $\tilde{\mathbf{A}}$ is a *selection* of \mathbf{A} for $t \in S$. If $\|\tilde{\mathbf{A}}(X(t))\|$ is bounded on all bounded subsets of S , then we can extend $\tilde{\mathbf{A}}$ to $[0, \infty)$ while retaining certain favorable properties. We discuss the technical details of this extension in Appendix E.1. The statements of Section 3 are stated with this extension.

¹Strictly speaking, this is a differential inclusion, not a differential equation, but we nevertheless refer to it as an ODE.

2.2 Derivation from discrete methods

We now show that the following instance of the anchor ODE

$$\dot{X}(t) = -\mathbf{A}(X(t)) - \frac{1}{t}(X(t) - X_0), \quad (4)$$

where $X(0) = X_0$ is the initial condition and $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuous operator, is a continuous-time model of APPM [35], EAG [75], and FEG [42], which are accelerated methods for monotone inclusion and minimax problems.

Consider APPM with operator $h\mathbf{A}$

$$\begin{aligned} x^k &= \mathbf{J}_{h\mathbf{A}}y^{k-1} \\ y^k &= \frac{k}{k+1}(2x^k - y^{k-1}) + \frac{1}{k+1}y^0 \end{aligned} \quad (5)$$

with initial condition $y^0 = x^0$. Assume $h > 0$ and $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuous monotone operator. Using $y^{k-1} = x^k + h\mathbf{A}x^k$ obtained from the first line, substituting y^k and y^{k-1} in the second line we get,

$$x^{k+1} + h\mathbf{A}x^{k+1} = \frac{k}{k+1}(x^k - h\mathbf{A}x^k) + \frac{1}{k+1}x^0.$$

Then reorganizing and dividing both sides by h , we have

$$\frac{x^{k+1} - x^k}{h} = -\mathbf{A}x^{k+1} - \frac{k}{k+1}\mathbf{A}x^k - \frac{1}{h(k+1)}(x^k - x^0).$$

Identifying $x^0 = X_0$, $2hk = t$, and $x^k = X(t)$, we have $\frac{k}{k+1} = 1 - \frac{h}{hk+h} = 1 + \mathcal{O}(h)$ and so

$$2\dot{X}(t) + \mathcal{O}(h) = -\mathbf{A}(X(t+2h)) - (1 + \mathcal{O}(h))\mathbf{A}(X(t)) - \frac{2}{t + \mathcal{O}(h)}(X(t) - X_0).$$

Taking limit $h \rightarrow 0^+$ and dividing both sides by 2, we get the anchor ODE (4). The correspondence with EAG and FEG are provided in Appendix D.4.

The following theorem establishes a rigorous correspondence between APPM and the anchor ODE for general maximal monotone operators.

Theorem 2.1. *Let \mathbf{A} be a (possibly set-valued) maximal monotone operator and assume $\text{Zer}\mathbf{A} \neq \emptyset$. Let x^k be the sequence generated by APPM (5) and X be the solution of the differential inclusion (3) with $\beta(t) = \frac{1}{t}$. For all fixed $T > 0$,*

$$\lim_{h \rightarrow 0^+} \max_{0 \leq k \leq \frac{T}{2h}} \|x^k - X(2kh)\| = 0.$$

We provide the proof in Appendix D.2.

2.3 Existence of the solution for $\beta(t) = \frac{\gamma}{t^p}$

To get further insight into the anchor acceleration, we generalize anchor coefficient to $\beta(t) = \frac{\gamma}{t^p}$ for $p, \gamma > 0$. We first establish the uniqueness and existence of the solution.

Theorem 2.2. *Consider (3) with $\beta(t) = \frac{\gamma}{t^p}$, i.e.*

$$\dot{X}(t) \in -\mathbf{A}(X(t)) - \frac{\gamma}{t^p}(X(t) - X_0). \quad (6)$$

for $p, \gamma > 0$. Then solution of (6) uniquely exists.

We provide the proof in Appendix B.

2.4 Additional properties of anchor ODE

We state a regularity lemma of the differential inclusion (3), which we believe may be of independent interest. In particular, we use this result several times throughout our various proofs.

Lemma 2.3. *Let $X(\cdot)$ and $Y(\cdot)$ are solutions of the differential inclusion (3) respectively with initial values and anchors X_0 and Y_0 . Then for all $t \in [0, \infty)$,*

$$\|X(t) - Y(t)\| \leq \|X_0 - Y_0\|.$$

We provide the proof in Appendix B.1.

Boundedness of trajectories is an immediate corollary of Lemma 2.3. Specifically, suppose $X(\cdot)$ is the solution of differential inclusion (3) with initial value X_0 . Then for all $X_* \in \text{Zer}\mathbf{A}$ and $t \in [0, \infty)$,

$$\|X(t) - X_*\| \leq \|X_0 - X_*\|.$$

This follows from setting $Y_0 = X_*$ in Lemma 2.3.

3 Convergence analysis

We now analyze the convergence rate of $\|\tilde{\mathbf{A}}(X(t))\|^2$ for the anchor ODE (3) with $\beta(t) = \frac{\gamma}{t^p}$ and $\gamma, p > 0$. The results are organized in Table 1.

Case	$p = 1, \gamma \geq 1$	$p = 1, \gamma < 1$	$p < 1$	$p > 1$
$\ \tilde{\mathbf{A}}(X(t))\ ^2$	$\mathcal{O}\left(\frac{1}{t^2}\right)$	$\mathcal{O}\left(\frac{1}{t^{2\gamma}}\right)$	$\mathcal{O}\left(\frac{1}{t^{2p}}\right)$	$\mathcal{O}(1)$

Table 1: Convergence rates of Theorem 3.1.

Let β be the anchor coefficient function of (3). Define $C: [0, \infty) \rightarrow \mathbb{R}$ as $C(t) = e^{\int_v^t \beta(s) ds}$ for some $v \in [0, \infty]$. Note that $\dot{C} = C\beta$ and C is unique up to scalar multiple. We call $\mathcal{O}(\beta(t))$ the *vanishing speed* and $\mathcal{O}\left(\frac{1}{C(t)}\right)$ the *contracting speed*, and we describe their trade-off in the following.

Loosely speaking, the *contracting speed* describes how fast the anchor term alone contracts the dynamical system. Consider $\dot{X}(t) = -\beta(t)(X(t) - a)$ for $a \in \mathbb{R}^n$, a system only with the anchor. Then, $X(t) = \frac{C(0)}{C(t)}(X(0) - a) + a$ is the solution, so the flow contracts towards the anchor a with rate $\frac{1}{C(t)}$. Intuitively speaking, this contracting behavior leads to stability and convergence. On the other hand, the anchor must eventually vanish, since our goal is to converge to an element in $\text{Zer}\mathbf{A}$, not the anchor. Thus the *vanishing speed* must be fast enough to not slow down the convergence of the flow to $\text{Zer}\mathbf{A}$.

This observation is captured in Figure 1. Consider a monotone linear operator $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ on \mathbb{R}^2 and $\beta(t) = \frac{\gamma}{t^p}$ with $\gamma = 1$ and $p > 0$. Note if there is no anchor, the ODE reduces to $\dot{X} = -\mathbf{A}(X)$ which do not converge [31, Chapter 8.2]. Figure 1 shows that with $p > 1$, the anchor vanished too early before the flow is contracted enough to result in converging flow. With $p < 1$, the flow does converge but the anchor vanished too late, slowing down the convergence. With $p = 1$, the convergence is fastest.

The following theorem formalizes this insight and produces the results of Table 1.

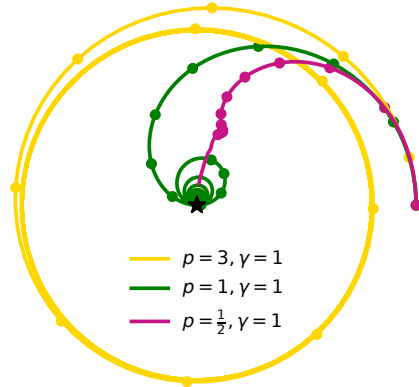


Figure 1: Flows of the solution of (6) with $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $\gamma = 1$, $X_0 = (1, 0)$ and various p . Flow is from $t = 0$ to $t = 100$. The marker is plotted every 0.8 units of time until $t = 9.6$. Note the last marker of the flow for $p = \frac{1}{2}$ is farther from the optimal point \star than that of the flow for $p = 1$.

Theorem 3.1. *Suppose \mathbf{A} is a maximal monotone operator with $\text{Zer}\mathbf{A} \neq \emptyset$. Consider (3) with $\beta(t) = \frac{\gamma}{t^p}$. Let $\tilde{\mathbf{A}}(X(t))$ be the selection of $\mathbf{A}(X(t))$ as in Section 2.1. Then,*

$$\|\tilde{\mathbf{A}}(X(t))\|^2 = \mathcal{O}\left(\frac{1}{C(t)^2}\right) + \mathcal{O}(\beta(t)^2) + \mathcal{O}(\dot{\beta}(t)).$$

Note that

$$C(t) = \begin{cases} t^\gamma & p = 1 \\ e^{\frac{\gamma}{1-p}t^{1-p}} & p \neq 1. \end{cases}$$

We expect the convergence rate of Theorem 3.1 to be optimized when the terms are balanced. When $\beta(t) = \frac{1}{t}$,

$$\frac{1}{C(t)^2} = \frac{1}{(e^{\int_1^t \frac{1}{s} ds})^2} = \frac{1}{t^2} = \beta(t)^2 = -\dot{\beta}(t)$$

and all three terms are balanced. Indeed, the choice $\beta(t) = \frac{1}{t}$ corresponds to the optimal discrete-time choice $\frac{1}{k+2}$ of APPM or other accelerated methods.

3.1 Proof outline of Theorem 3.1

The proof of Theorem 3.1 follows from Lemma 3.4, which we will introduce later in this section. To derive Lemma 3.4, we introduce a conservation law.

Proposition 3.2. *Suppose $\tilde{\mathbf{A}}$ is Lipschitz continuous and monotone. For $t_0 > 0$, define $E : (0, \infty) \rightarrow \mathbb{R}$ as*

$$E = \frac{C(t)^2}{2} \left(\|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \right) - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds + \int_{t_0}^t C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds.$$

Then E is a constant function.

The proof of Proposition 3.2 uses dilated coordinate $W(t) = C(t)(X(t) - X_0)$ to derive its conservation law in the style of Suh et al. [67]. We provide the details in Appendix E.2.

Recall from (1) that $\left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle \geq 0$, the integrand of the last term of E is nonnegative. This motivates us to define

$$V(t) = E - \int_{t_0}^t C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds$$

as our Lyapunov function.

Corollary 3.3. *Let \mathbf{A} be maximal monotone and $\beta(t) = \frac{\gamma}{t^p}$ with $p > 0$, $\gamma > 0$. Let $\tilde{\mathbf{A}}(X(t))$ be the selection of $\mathbf{A}(X(t))$ as in Section 2.1. For $t_0 \geq 0$, define $V : [0, \infty) \rightarrow \mathbb{R}$ as*

$$V(t) = \frac{C(t)^2}{2} \left(\|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \right) - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds.$$

for $t > 0$ and $V(0) = \lim_{t \rightarrow 0^+} V(t)$. Then $V(t) \leq V(0)$ holds for $t \geq 0$.

A technical detail is that all terms involving $\frac{d}{ds} \tilde{\mathbf{A}}(X(s))$ have been excluded in the definition of V and this is what allows \mathbf{A} to not be Lipschitz continuous. We provide the details in Appendix E.3.

Lemma 3.4. *Consider the setup of Corollary 3.3. Assume $\text{Zer}\mathbf{A} \neq \emptyset$. Then for $t > 0$ and $X_\star \in \text{Zer}\mathbf{A}$,*

$$\begin{aligned} \|\tilde{\mathbf{A}}(X(t))\|^2 &\leq 4\beta(t)^2 \|X_0 - X_\star\|^2 + \frac{4V(0)}{C(t)^2} - 2(\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \\ &\quad + \frac{2}{C(t)^2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds. \end{aligned} \tag{7}$$

Proof outline of Lemma 3.4. Define

$$\Phi(t) = \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle.$$

Then, from monotonicity of $\tilde{\mathbf{A}}$ and Young's inequality,

$$\begin{aligned} \Phi(t) &\geq \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X_* - X_0 \rangle \\ &\geq \|\tilde{\mathbf{A}}(X(t))\|^2 - 2 \left(\left\| \frac{1}{2} \tilde{\mathbf{A}}(X(t)) \right\|^2 + \|\beta(t) (X_* - X_0)\|^2 \right) \\ &= \frac{1}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 - 2\beta(t)^2 \|X_0 - X_*\|^2. \end{aligned} \quad (8)$$

By Corollary 3.3, $\frac{2V(0)}{C(t)^2} - \frac{2V(t)}{C(t)^2} + \Phi(t) \geq \Phi(t)$ for $t > 0$. Applying (8) and organizing, we can get the desired result. The details are provided in Appendix E.4. \square

Proof outline of Theorem 3.1. It remains to show that last integral term of Lemma 3.4 is $\mathcal{O}\left(\frac{1}{C(t)^2}\right) + \mathcal{O}(\beta(t)^2) + \mathcal{O}(\dot{\beta}(t))$. The details are provided in Appendix E.5. \square

Before we end this section, we observe how our analysis simplifies in the special case $\beta(t) = \frac{1}{t}$. In this case,

$$V(t) = \frac{t^2}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 + t \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle,$$

and this corresponds to the Lyapunov function of [59, Section 4] for the case $\gamma = 1$. As $V(0) = 0$, the conclusion of Lemma 3.4 becomes

$$\|\tilde{\mathbf{A}}(X(t))\|^2 \leq \frac{4}{t^2} \|X_0 - X_*\|^2 = \mathcal{O}\left(\frac{1}{t^2}\right),$$

which to the best rate in Table 1.

3.2 Point convergence

APPM is an instance of the Halpern method [54, Lemma 3.1], which iterates converge to the element in $\text{Zer}\mathbf{A}$ closest to X_0 [34, 73]. The anchor ODE also exhibits this behavior.

Theorem 3.5. *Let \mathbf{A} be a maximal monotone operator with $\text{Zer}\mathbf{A} \neq \emptyset$ and X be the solution of (3). If $\lim_{t \rightarrow \infty} \|\tilde{\mathbf{A}}(X(t))\| = 0$ and $\lim_{t \rightarrow \infty} 1/C(t) = 0$, then, as $t \rightarrow \infty$,*

$$X(t) \rightarrow \underset{z \in \text{Zer}\mathbf{A}}{\text{argmin}} \|z - X_0\|.$$

We provide the proof in Appendix E.6.

4 Tightness of analysis

In this section, we show that the convergence rates of Table 1 are actually tight by considering the dynamics under the explicit example $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Throughout this section, we denote \mathbf{A} as A when the operator is linear.

4.1 Explicit solution for linear A

Lemma 4.1. *Let $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear operator and let $\beta(t) = \frac{\gamma}{t}$. The series*

$$X(t) = \sum_{n=0}^{\infty} \frac{(-tA)^n}{\Gamma(n + \gamma + 1)} \Gamma(\gamma + 1) X_0,$$

where Γ denotes the gamma function, is the solution for (3) with $\mathbf{A} = A$.

Note that when $\gamma = 0$, this is the series definition of the matrix exponential and $X(t) = e^{-tA}$. The solution also has an integral form, which extends to general $\beta(t)$.

Lemma 4.2. *Suppose $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a monotone linear operator. Then*

$$X(t) = \frac{e^{-tA}}{C(t)} \left(\int_0^t e^{sA} C(s) \beta(s) ds + C(0)I \right) X_0 \quad (9)$$

is the solution for (3) with $\mathbb{A} = A$.

See Appendix F.1.1 and Appendix F.1.2 for details.

4.2 The rates in Table 1 are tight

First, we consider $p > 1$ for $\beta(t) = \frac{\gamma}{t^p}$.

Theorem 4.3. *Suppose $\lim_{t \rightarrow \infty} \frac{1}{C(t)} \neq 0$, i.e., suppose $\beta(t) \in L^1[t_0, \infty)$ for some $t_0 > 0$. Then there exists an operator \mathbb{A} such that*

$$\lim_{t \rightarrow \infty} \|\tilde{\mathbb{A}}(X(t))\| \neq 0,$$

where X is the solution of (3).

Note that $\frac{\gamma}{t^p} \in L^1[t_0, \infty)$ when $p > 1$. The proof of Theorem 4.3 considers $\mathbb{A} = 2\pi\xi \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ for $\xi \in \mathbb{R}$ and uses the Fourier inversion formula. See Appendix F.2 for details.

Next, we consider $\beta(t) = \frac{\gamma}{t^p}$ for cases other than $p > 1$.

Theorem 4.4. *Let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $\beta(t) = \frac{\gamma}{t^p}$, $0 < p \leq 1$, and $\gamma > 0$. Let X be the solution given by (9) and $X_0 \neq 0$. Let*

$$r(t) = \begin{cases} t^2 & \text{for } p = 1, \gamma \geq 1 \\ t^{2\gamma} & \text{for } p = 1, \gamma < 1, \\ t^{2p} & \text{for } 0 < p < 1. \end{cases}$$

Then,

$$\lim_{t \rightarrow \infty} r(t) \|A(X(t))\|^2 \neq 0.$$

We provide the proof in Appendix F.3.

5 Discretized algorithms

In this section, we provide discrete-time convergence results that match the continuous-time rate of Section 3.

Theorem 5.1. *Suppose \mathbb{A} be a maximal monotone operator; $p > 0$, and $\gamma > 0$. Consider*

$$\begin{aligned} x^k &= \mathbb{J}_{\mathbb{A}} y^{k-1} \\ y^k &= \frac{k^p}{k^p + \gamma} (2x^k - y^{k-1}) + \frac{\gamma}{k^p + \gamma} x^0 \end{aligned}$$

for $k = 1, 2, \dots$, with initial condition $y^0 = x^0 \in \mathbb{R}^n$. Let $\tilde{\mathbb{A}}x^k = y^{k-1} - x^k$ for $k = 1, 2, \dots$. Then this method exhibits the rates of convergence in Table 2.

Case	$\begin{matrix} p = 1, \\ \gamma \geq 1 \end{matrix}$	$\begin{matrix} p = 1, \\ \gamma < 1 \end{matrix}$	$p < 1$	$p > 1$
$\ \tilde{\mathbb{A}}(x^k)\ ^2$	$\mathcal{O}\left(\frac{1}{k^2}\right)$	$\mathcal{O}\left(\frac{1}{k^{2\gamma}}\right)$	$\mathcal{O}\left(\frac{1}{k^{2p}}\right)$	$\mathcal{O}(1)$

Table 2: Rates for the discrete-time method of Theorem 5.1.

Note that the method of Theorem 5.1 reduces to APPM when $\gamma = 1, p = 1$.

Proof outline of Theorem 5.1. The general strategy is to find discretized counterparts of corresponding continuous-time analyses. However, directly discretizing the conservation law of Proposition 3.2 was difficult due to technical reasons. Instead, we obtain differently scaled but equivalent conservation laws using dilated coordinates and then performed the discretization. The specific dilated coordinates, inspired by [67], are $W_1(t) = X(t) - X_0$ for $p > 1$, $W_2(t) = t^p (X(t) - X_0)$ for $0 < p < 1$, $W_3(t) = t(X(t) - X_0)$ for $p = 1, \gamma \geq 1$ and $W_4(t) = t^\gamma (X(t) - X_0)$ for $p = 1, 0 < \gamma < 1$.

In the discrete-time analyses, the behavior of the leading-order terms is predictable as they match the continuous-time counterpart. The difficult part is, however, controlling the higher-order terms that were not present in the continuous-time analyses. Through our detailed analyses, we bound such higher-order terms and show that they do not affect the convergence rate in the end. We provide the details in Appendix G.3. \square

6 Convergence analysis under strong monotonicity

In this section, we analyze the dynamics of the anchor ODE (3) for μ -strongly monotone \mathbf{A} . When $\beta(t) = \frac{1}{t}$ and $\mathbf{A} = \begin{pmatrix} \mu & 0 \\ 0 & \mu \end{pmatrix}$, Lemma 4.1 tells us that $\mathbf{A}(X(t)) = \frac{1}{t} (I - e^{-t\mathbf{A}}) X_0$ and therefore that $\|\mathbf{A}(X(t))\|^2 = \Theta\left(\frac{1}{t^2}\right)$, which is a slow rate for the strongly monotone setup. On the other hand, we will see that $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1}$ is a better choice leading to a faster rate in this setup.

Our analysis of this section is also based on a conservation law, but we use a slightly modified version to exploit strong monotonicity.

Proposition 6.1. *Suppose $\tilde{\mathbf{A}}$ is monotone and Lipschitz continuous. Let X be the solution of (3) and let $R : [0, \infty) \rightarrow (0, \infty)$ be a differentiable function. For $t_0 > 0$, define $E : (0, \infty) \rightarrow \mathbb{R}$ as*

$$E = \frac{C(t)^2 R(t)^2}{2} \left(\|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \left(\beta(t)^2 + \dot{\beta}(t) \right) \|X(t) - X_0\|^2 \right) - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 R(s)^2 \dot{\beta}(s)}{2} \right) \|X(t) - X_0\|^2 ds + \int_{t_0}^t C(s)^2 R(s)^2 \left(\left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle - \frac{\dot{R}(s)}{R(s)} \|\dot{X}(s)\|^2 \right) ds.$$

Then E is a constant function for $t \in [0, \infty)$.

Proposition 6.1 generalizes Proposition 3.2, since it corresponds to the special case with $R(t) \equiv 1$.

Recall from (2), when \mathbf{A} is μ -strongly monotone we have

$$\left\langle \frac{d}{ds} \mathbf{A}(X(t)), \dot{X}(t) \right\rangle - \mu \|\dot{X}(t)\|^2 \geq 0.$$

This motivates the choice $R(t) = e^{\mu t}$, since $\frac{\dot{R}(s)}{R(s)} = \mu$. From calculation provided in Appendix H.2, the choice $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1}$ makes $\frac{d}{ds} \left(\frac{C(s)^2 R(s)^2 \dot{\beta}(s)}{2} \right) = 0$. Plugging these choices into Proposition 6.1 and following arguments of Section 3, we arrive at the following theorem.

Theorem 6.2. *Let \mathbf{A} be a μ -strongly maximal monotone operator with $\mu > 0$ and assume $\text{Zer}\mathbf{A} \neq \emptyset$. Let X be a solution of the differential inclusion (3) with $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1}$, i.e. for almost all t ,*

$$\dot{X} \in -\mathbf{A}(X) - \frac{2\mu}{e^{2\mu t} - 1} (X - X_0). \quad (10)$$

Let $\tilde{\mathbf{A}}(X(t))$ be the selection of $\mathbf{A}(X(t))$ as in Section 2.1. Define $V : [0, \infty) \rightarrow \mathbb{R}$ as

$$V(t) = \frac{(e^{\mu t} - e^{-\mu t})^2}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\mu (1 - e^{-2\mu t}) \left(\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle - \mu \|X(t) - X_0\|^2 \right).$$

Then $V(t) \leq V(0)$ holds for $t \geq 0$. Furthermore for $X_\star \in \text{Zer}\mathbf{A}$,

$$\|\tilde{\mathbf{A}}(X(t))\|^2 \leq \left(\frac{2\mu}{e^{\mu t} - 1} \right)^2 \|X_0 - X_\star\|^2 = \mathcal{O}\left(\frac{1}{e^{2\mu t}}\right).$$

In Appendix H.2.3, we show that (10) is a continuous-time model for OS-PPM of Park and Ryu [54]. In Appendix C, we show the existence and uniqueness of the solution.

Since $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1} \in L^1[t_0, \infty)$ for any $t_0 > 0$, Theorem 4.3 implies that $\tilde{\mathbf{A}}(X(t)) \rightarrow 0$ when \mathbf{A} is merely monotone. This tells us that the optimal choice of $\beta(t)$ for should depend on the properties of \mathbf{A} . In the following section, we describe how $\beta(t)$ can be chosen to adapt to the operator's properties.

7 Adaptive anchor acceleration and experiments

In this section, we present an adaptive method for choosing the anchor coefficient β , and we theoretically and experimentally show that this choice allows the dynamics to adapt to the operator's properties. It achieves the optimal $\mathcal{O}(1/k^2)$ -convergence rate when \mathbf{A} is monotone and an exponential convergence rate when \mathbf{A} is furthermore μ -strongly monotone and Lipschitz continuous.

Theorem 7.1. *Suppose $\tilde{\mathbf{A}}$ is Lipschitz continuous and monotone. Consider the anchor ODE*

$$\dot{X} = -\tilde{\mathbf{A}}(X) + \underbrace{\frac{\|\tilde{\mathbf{A}}(X)\|^2}{2\langle \tilde{\mathbf{A}}(X), X - X_0 \rangle}}_{=-\beta(t)}(X - X_0) \quad (11)$$

with initial condition $X(0) = X_0$ and $\|\tilde{\mathbf{A}}(X_0)\| \neq 0$. Suppose the solution exists and \dot{X} is continuous at $t = 0$. Moreover, suppose $\beta: (0, \infty) \rightarrow \mathbb{R}$ is well-defined, i.e., no division by zero occurs in the definition of $\beta(t)$. Then for $t > 0$ and $X_* \in \text{Zer}\tilde{\mathbf{A}}$, we have $\beta(t) > 0$ and

$$\begin{aligned} \|\tilde{\mathbf{A}}(X(t))\|^2 &\leq 4\beta(t)^2 \|X_0 - X_*\|^2 \\ \beta(t)^2 &\leq \frac{1}{t^2}. \end{aligned}$$

If $\tilde{\mathbf{A}}$ is furthermore μ -strongly monotone, then for $t > 0$,

$$\beta(t)^2 \leq \left(\frac{\mu/2}{e^{\mu t/2} - 1} \right)^2.$$

We provide the proof in Appendix I.1. Note that anchor coefficient (11) is chosen so that

$$\Phi(t) = \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle = 0.$$

So left-hand side of (8) is zero and a $\mathcal{O}(\beta(t)^2)$ convergence rate is immediate. An analogous discrete-time result is shown in the following theorem.

Theorem 7.2. *Let \mathbf{A} be a maximal monotone operator. Let $x^0 = y^0 \in \mathbb{R}^n$. Consider*

$$\begin{aligned} x^k &= \mathbf{J}_{\mathbf{A}} y^{k-1} \\ y^k &= (1 - \beta_k)(2x^k - y^{k-1}) + \beta_k x^0 \end{aligned}$$

with

$$\beta_k = \begin{cases} \frac{\|\tilde{\mathbf{A}}x^k\|^2}{-\langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle + \|\tilde{\mathbf{A}}x^k\|^2} & \text{if } \|\tilde{\mathbf{A}}x^k\|^2 \neq 0 \\ 0 & \text{if } \|\tilde{\mathbf{A}}x^k\|^2 = 0, \end{cases}$$

for $k = 1, 2, \dots$, where $\tilde{\mathbf{A}}x^k = y^{k-1} - x^k$.

Then $\beta_k \geq 0$ and

$$\begin{aligned} \|\tilde{\mathbf{A}}x^{k+1}\|^2 &\leq \beta_k^2 \|x^0 - x^*\|^2 \\ \beta_k^2 &\leq \frac{1}{(k+1)^2} \end{aligned}$$

for $k = 1, 2, \dots$ and $x^* \in \text{Zer}\mathbf{A}$.

If \mathbf{A} is furthermore μ -strongly monotone and L -Lipschitz continuous, then for $k = 1, 2, \dots$,

$$\beta_k^2 \leq \left(\frac{\mu/(1+L^2)}{(1+\mu/(1+L^2))^k - 1 + \mu/(1+L^2)} \right)^2.$$

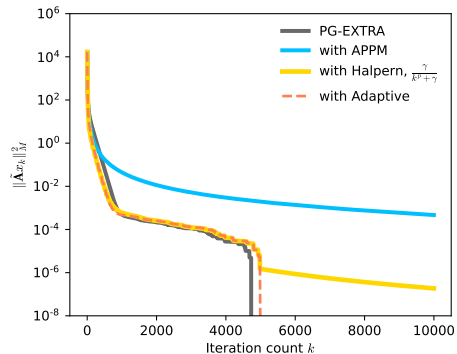
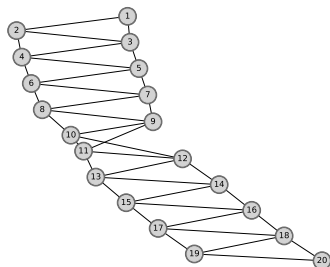


Figure 2: (Left) Network graph. (Right) Squared M -norm $\|\tilde{\mathbf{A}}x^k\|_M^2$ vs. k . Halpern corresponds to the method in Theorem 5.1, we use $p = 1.5$ and $\gamma = 2.0$.

For the monotone setup, the rate $\|\tilde{\mathbf{A}}x^{k+1}\|^2 \leq \frac{1}{(k+1)^2} \|x^0 - x^*\|^2$ matches the exact optimal rate of APPM [54]. In the limit $\mu \rightarrow 0$, the result for the μ -strongly monotone case reduces to the result for the monotone case. We provide the proof and details of Theorem 7.2 in Appendix I.3.

The method of Theorem 7.2 is a discrete-time counterpart of the ODE of (11). The extra term $\|\tilde{\mathbf{A}}x^k\|^2$ of the denominator in the definition of β_k vanishes in the continuous-time limit. We provide further details in Appendix I.2. Analogous to the continuous-time case, a key property of the discrete-time adaptive method is that the counterpart of $\Phi(t)$ is kept nonpositive. In the proof of Lemma I.3, the fact $\beta_k < 1$ plays the key role in proving this property. The extra term $\|\tilde{\mathbf{A}}x^k\|^2$ in the denominator and the fact that $\langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle < 0$ when $\|\tilde{\mathbf{A}}x^k\|^2 \neq 0$ leads to $\beta_k < 1$.

7.1 Experiment details

We now show an experiment with the method of Theorem 7.2 applied to a decentralized compressed sensing problem Shi et al. [63]. We assume that we have the measurement $b_i = A_{(i)}x + e_i$, where $A_{(i)}$ is a measurement matrix available for each local agent i , x is an unknown shared signal we hope to recover, and e_i is an error in measurement. We solve this problem in a decentralized manner in which the local agents keep their measurements private and only communicate with their neighbors.

As in Shi et al. [63], we formulate the problem into an unconstrained ℓ_1 -regularized least squares problem

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{2} \|A_{(i)}x - b_i\|^2 + \rho \|x\|_1 \right\},$$

and apply PG-EXTRA. We compare vanilla PG-EXTRA with the various anchored versions of PG-EXTRA with β_k as in Theorem 5.1 and Theorem 7.2. We show the results in Figure 2. Further details of the experiment are provided in Appendix J.

8 Conclusion

This work introduces a continuous-time model of anchor acceleration, the anchor ODE $\dot{X} \in -\mathbf{A}(X) - \beta(t)(X - X_0)$. We characterize the convergence rate as a function of $\beta(t)$ and thereby obtain insight into the anchor acceleration mechanism. Finally, inspired by the continuous-time analyses, we present an adaptive method and establish its effectiveness through theoretical analyses and experiments.

Prior work analyzing continuous-time models of Nesterov acceleration had inspired various follow-up research, such as analyses based on Lagrangian and Hamiltonian mechanics [71, 72, 26], high-resolution ODE model [62], and continuized framework [29]. Carrying out similar analyses for the anchor ODE are interesting directions of future work.

Acknowledgments and Disclosure of Funding

This work was supported by the Samsung Science and Technology Foundation (Project Number SSTF-BA2101-02). We thank Jaeyeon Kim for providing valuable feedback. We thank Hangjun Cho for the helpful discussions on well-posedness of ODEs. We also thank anonymous reviewers for the constructive comments.

References

- [1] J. K. Alcala, Y. T. Chow, and M. Sunkula. Moving anchor extragradient methods for smooth structured minimax problems. *arXiv:2308.12359*, 2023.
- [2] V. Apidopoulos, J.-F. Aujol, and C. Dossal. The differential inclusion modeling FISTA algorithm and optimality of convergence rate in the case $b \leq 3$. *SIAM Journal on Optimization*, 28(1): 551–574, 2018.
- [3] H. Attouch and A. Cabot. Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. *Journal of Differential Equations*, 263(9):5412–5458, 2017.
- [4] H. Attouch and S. C. László. Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. *SIAM Journal on Optimization*, 30(4):3252–3283, 2020.
- [5] H. Attouch and S. C. László. Convex optimization via inertial algorithms with vanishing Tikhonov regularization: Fast convergence to the minimum norm solution. *arXiv:2104.11987*, 2021.
- [6] H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Mathematical Programming*, 174(1):391–432, 2019.
- [7] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168(1):123–175, 2018.
- [8] H. Attouch, Z. Chbani, and H. Riahi. Combining fast inertial dynamics for convex optimization with Tikhonov regularization. *Journal of Mathematical Analysis and Applications*, 457(2): 1065–1094, 2018.
- [9] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:2, 2019.
- [10] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping. *Optimization*, pages 1–40, 2021.
- [11] J.-P. Aubin and A. Cellina. *Differential Inclusions*. Springer, 1984.
- [12] J.-F. Aujol, C. Dossal, and A. Rondepierre. Optimal convergence rates for Nesterov acceleration. *SIAM Journal on Optimization*, 29(4):3131–3153, 2019.
- [13] J.-B. Baillon and R. E. Bruck. Optimal rates of asymptotic regularity for averaged nonexpansive mappings. *Fixed Point Theory and Applications*, 128:27–66, 1992.
- [14] S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922.
- [15] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer International Publishing, second edition, 2017.
- [16] R. I. Boţ and E. R. Csetnek. A second-order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities. *Optimization*, 68(7):1265–1277, 2019.

- [17] R. I. Boş and D. A. Hulett. Second order splitting dynamics with vanishing damping for additively structured monotone inclusions. *Journal of Dynamics and Differential Equations*, 2022.
- [18] R. I. Boş, E. R. Csetnek, and S. C. László. A second-order dynamical approach with variable damping to nonconvex smooth minimization. *Applicable Analysis*, 99(3):361–378, 2020.
- [19] R. I. Boş, E. R. Csetnek, and S. C. László. Tikhonov regularization of a second order dynamical system with Hessian driven damping. *Mathematical Programming*, 189(1):151–186, 2021.
- [20] R. I. Bot, E. R. Csetnek, and D.-K. Nguyen. Fast OGDAs in continuous and discrete time. *arXiv:2203.10947*, 2022.
- [21] M. Bravo and R. Cominetti. Sharp convergence rates for averaged nonexpansive maps. *Israel Journal of Mathematics*, 227:163–188, 2018.
- [22] A.-L. Cauchy. Méthode générale pour la résolution des systèmes d’équations simultanées. *Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences*, 25:536–538, 1847.
- [23] R. Cominetti, J. A. Soto, and J. Vaisman. On the rate of convergence of Krasnosel’skiĭ–Mann iterations and their connection with sums of Bernoullis. *Israel Journal of Mathematics*, 199(2): 757–772, 2014.
- [24] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. Training GANs with optimism. *International Conference on Learning Representations*, 2018.
- [25] J. Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. *Conference on Learning Theory*, 2020.
- [26] J. Diakonikolas and M. I. Jordan. Generalized momentum-based methods: A Hamiltonian perspective. *SIAM Journal on Optimization*, 31(1):915–944, 2021.
- [27] Y. Drori. The exact information-based complexity of smooth convex minimization. *Journal of Complexity*, 39:1–16, 2017.
- [28] J. Eckstein and D. P. Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.
- [29] M. Even, R. Berthier, F. Bach, N. Flammarion, H. Hendriks, P. Gaillard, L. Massoulié, and A. Taylor. Continuized accelerations of deterministic and stochastic gradient descents, and of gossip algorithms. *Neural Information Processing Systems*, 2021.
- [30] George J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.
- [31] I. Goodfellow. NIPS 2016 tutorial: Generative adversarial networks. *arXiv:1701.00160*, 2016.
- [32] E. Gorbunov, N. Loizou, and G. Gidel. Extragradient method: $O(1/K)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity. *International Conference on Artificial Intelligence and Statistics*, 2022.
- [33] G. Gu and J. Yang. Tight sublinear convergence rate of the proximal point algorithm for maximal monotone inclusion problems. *SIAM Journal on Optimization*, 30(3):1905–1921, 2020.
- [34] B. Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957–961, 1967.
- [35] D. Kim. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190(1–2):57–87, 2021.
- [36] D. Kim and J. A. Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical Programming*, 159(1–2):81–107, 2016.

- [37] D. Kim and J. A. Fessler. Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions. *Journal of Optimization Theory and Applications*, 188(1): 192–219, 2021.
- [38] J. Kim and I. Yang. Unifying Nesterov’s accelerated gradient methods for convex and strongly convex objective functions. *International Conference on Machine Learning*, 202, 2023.
- [39] U. Kohlenbach. On quantitative versions of theorems due to F. E. Browder and R. Wittmann. *Advances in Mathematics*, 226(3):2764–2795, 2011.
- [40] M. A. Krasnosel’skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1):123–127, 1955.
- [41] J. Lee and E. K. Ryu. Accelerating value iteration with anchoring. *Neural Information Processing Systems*, 2023.
- [42] S. Lee and D. Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. *Neural Information Processing Systems*, 2021.
- [43] L. Leustean. Rates of asymptotic regularity for Halpern iterations of nonexpansive mappings. *Journal of Universal Computer Science*, 13(11):1680–1691, 2007.
- [44] J. Liang, J. Fadili, and G. Peyré. Convergence rates with inexact non-expansive operators. *Mathematical Programming*, 159(1):403–434, 2016.
- [45] F. Lieder. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15(2): 405–418, 2021.
- [46] T. Lin and M. I. Jordan. Monotone Inclusions, Acceleration, and Closed-Loop Control. *Mathematics of Operations Research*, 2023.
- [47] W. R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4(3):506–510, 1953.
- [48] B. Martinet. Régularisation d’inéquations variationnelles par approximations successives. *Revue Française d’Informatique et de Recherche Opérationnelle, Série Rouge*, 4(3):154–158, 1970.
- [49] B. Martinet. *Algorithmes Pour La Résolution de Problèmes d’optimisation et de Minimax*. PhD thesis, Université Scientifique et Médicale de Grenoble, 1972.
- [50] S.-Y. Matsushita. On the convergence rate of the Krasnosel’skiĭ–Mann iteration. *Bulletin of the Australian Mathematical Society*, 96(1):162–170, 2017.
- [51] Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Doklady Akademii Nauk SSSR*, 269(3):543–547, 1983.
- [52] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 2004.
- [53] C. Park, J. Park, and E. K. Ryu. Factor- $\sqrt{2}$ acceleration of accelerated gradient methods. *Applied Mathematics and Optimization*, 2022.
- [54] J. Park and E. K. Ryu. Exact optimal accelerated complexity for fixed-point iterations. *International Conference on Machine Learning*, 2022.
- [55] J. Park and E. K. Ryu. Accelerated infeasibility detection of constrained optimization and fixed-point iterations. *International Conference on Machine Learning*, 2023.
- [56] L. D. Popov. A modification of the Arrow–Hurwicz method for search of saddle points. *Mathematical Notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.
- [57] A. Rakhlin and K. Sridharan. Online learning with predictable sequences. *Conference on Learning Theory*, 2013.
- [58] E. K. Ryu and W. Yin. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge University Press, 2022.

- [59] E. K. Ryu, K. Yuan, and W. Yin. ODE analysis of stochastic gradient methods with optimism and anchoring for minimax problems and GANs. *arXiv:1905.10899*, 2019.
- [60] S. Sabach and S. Shtern. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27(2):640–660, 2017.
- [61] A. Salim, L. Condat, D. Kovalev, and P. Richtárik. An optimal algorithm for strongly convex minimization under affine constraints. *AISTATS*, 2022.
- [62] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*, 2021.
- [63] W. Shi, Q. Ling, G. Wu, and W. Yin. A proximal gradient algorithm for decentralized composite optimization. *IEEE Transactions on Signal Processing*, 63(22):6013–6023, 2015.
- [64] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient – proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, 7(4):323–345, 1999.
- [65] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights. *Neural Information Processing Systems*, 2014.
- [66] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.
- [67] J. J. Suh, G. Roh, and E. K. Ryu. Continuous-time analysis of AGM via conservation laws in dilated coordinate systems. *International Conference on Machine Learning*, 2022.
- [68] A. Taylor and Y. Drori. An optimal gradient method for smooth strongly convex minimization. *Mathematical Programming*, 2022.
- [69] Q. Tran-Dinh and Y. Luo. Halpern-type accelerated and splitting algorithms for monotone inclusions. *arXiv:2110.08150*, 2021.
- [70] B. Van Scoy, R. A. Freeman, and K. M. Lynch. The fastest known globally convergent first-order method for minimizing strongly convex functions. *IEEE Control Systems Letters*, 2(1):49–54, 2018.
- [71] A. Wibisono, A. C. Wilson, and M. I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.
- [72] A. C. Wilson, B. Recht, and M. I. Jordan. A Lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, 22(113):1–34, 2021.
- [73] R. Wittmann. Approximation of fixed points of nonexpansive mappings. *Archiv der Mathematik*, 58(5):486–491, 1992.
- [74] T. Wu, K. Yuan, Q. Ling, W. Yin, and A. H. Sayed. Decentralized consensus optimization with asynchrony and delays. *IEEE Transactions on Signal and Information Processing over Networks*, 4(2):293–307, 2018.
- [75] T. Yoon and E. K. Ryu. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. *International Conference on Machine Learning*, 2021.
- [76] T. Yoon and E. K. Ryu. Accelerated minimax algorithms flock together. *arXiv:2205.11093*, 2022.

A Prior work

Acceleration for smooth convex function in discrete setting. There had been rich amount of research on acceleration about smooth convex functions. Nesterov [51] introduced accelerated gradient method (AGM), which has a faster $\mathcal{O}(1/k^2)$ rate than $\mathcal{O}(1/k)$ rate of gradient descent [22] in reducing the function value. Optimized gradient method (OGM) [36] improved AGM’s rate by a constant factor, and is proven to be optimal [27]. For smooth strongly convex setup, strongly convex AGM [52] achieves an accelerated rate, and further improvements were studied [70, 53, 68, 61]. Recently, OGM-G [37] was introduced as an accelerated method reducing squared gradient magnitude for smooth convex minimization.

Acceleration for smooth convex function in continuous setting. Continuous-time analysis of Nesterov acceleration has been thoroughly studied as well. Su et al. [65, 66] introduced an ODE model of AGM $\dot{X}(t) + \frac{r}{t}X + \nabla f(X) = 0$, providing $f(X(t)) - f_* \in \mathcal{O}(1/t^2)$ rate for $r \geq 3$. Attouch et al. [7] improved the constant of bound for $r > 3$ and proved convergence of the trajectories. Attouch et al. [9] achieved $\mathcal{O}(1/t^{-2r/3})$ rate for $0 < r < 3$. Apidopoulos et al. [2] generalized their results to differential inclusion with non-differentiable convex function. Furthermore, wide range of variations of AGM ODE has been studied [3, 8, 12, 16, 18, 5, 10, 19]. Also, applications to monotone inclusion problem were studied by Attouch and Peypouquet [6], Attouch and László [4], Boţ and Hulett [17].

Motivated from above continuous-time analysis for accelerated methods, tools analyzing ODEs have further developed. Wibisono et al. [71], Wilson et al. [72] and Kim and Yang [38] adopted Lagrangian mechanics and introduced first, second, unified Bregman Lagrangian to provide unified analysis for generalized family of ODE, where the latter provided analysis for strongly convex AGM. Systemical approach to obtain Lyapunov functions exploiting Hamiltonian mechanics [26] and dilated coordinate system [67] were proposed, and analysis of OGM-G was provided by dilated coordinate framework. Different forms of continuous-time models such as high-resolution ODE [62] and continuized framework [29] were developed.

On the other hand, another type of acceleration called *anchor acceleration* recently gained attention. As Yoon and Ryu [76] focused, many recently discovered accelerated methods for both minimax optimization and fixed-point problems are based on anchor acceleration. More recently, practical applications of anchor acceleration to detecting infeasibility for constrained optimization problems [55], and accelerating value iteration for dynamic programming and reinforcement learning [41] were introduced as well.

Fixed-point problem. The history of studies on fixed-point problem dates back to the work of Banach [14], which established that the Picard iteration with contractive operator is convergent. Krasnosel’skii-Mann iteration (KM) [40, 47] was introduced, which is a generalization of Picard iteration. Convergence of KM iteration with general nonexpansive operators was proven by Martinet [49]. For iteration of Halpern [34], convergence with wide choice of parameter were shown by Wittmann [73].

The squared norm $\|y_k - \mathbb{T}y_k\|^2$ of fixed-point residual is a common error measure for fixed-point problems. KM iteration was shown to exhibit $\mathcal{O}(1/k)$ rate [23, 44, 21] and $o(1/k)$ -rate [13, 50]. For Halpern iteration, $\mathcal{O}(1/(\log k)^2)$ -rate was established by Leustean [43], then improve to $\mathcal{O}(1/k)$ rate by Kohlenbach [39]. First accelerated $\mathcal{O}(1/k^2)$ rate was achieved by Sabach and Shtern [60] and the constant was improved by Lieder [45] by a factor of 16.

It is known that there is an equivalence between solving fixed-point problem and solving monotone inclusion problem [30, 28, 54]. Proximal point method (PPM) [48] achieves $\mathcal{O}(1/k)$ -rate in terms of $\|\tilde{A}x_k\|^2$ [33]. Accelerated proximal point method (APPM) [35] improved the rate to accelerated $\mathcal{O}(1/k^2)$ -rate. Park and Ryu [54] showed APPM is exactly optimal method for this problem and provided exactly optimal method for μ -strongly monotone operator named OS-PPM, which achieved $\mathcal{O}(1/e^{4\mu k})$ rate. The optimal methods APPM and OS-PPM are based on anchor acceleration [76].

Minimax problems. Minimax optimization problem of the form $\min_x \max_y \mathbf{L}(x, y)$ have recently gained attention in machine learning society. One of the commonly considered theoretical setting is smooth convex-concave setup, with squared gradient norm as error measure. In terms of $\|\partial \mathbf{L}(x, y)\|^2$, classical EG [64] and OG [56, 57, 24] was shown to achieved $\mathcal{O}(1/k)$ -rate [32]. SGDA [59] achieved $\mathcal{O}(1/k^{2-2p})$ rate for $p > 1/2$ with introducing the term *anchor*. With introducing a parameter-free Halpern type method, Diakonikolas [25] achieved $\mathcal{O}(\log k/k^2)$. Recently, EAG [75] first achieved accelerated rate $\mathcal{O}(1/k^2)$ with anchor acceleration, followed by FEG [42], anchored Popov’s scheme [69] and moving anchor methods [1]. Fast ODGA [20] also achieved accelerated $o(1/k^2)$ rate. For $\partial \mathbf{L}$ is furthermore strongly monotone with condition number κ , SM-EAG+ [76] achieved accelerated $\mathcal{O}(1/e^{2k\kappa})$ rate.

However, continuous-time analysis for anchor acceleration is, to the best of our knowledge, insufficient. Continuous-time analyses of acceleration for monotone inclusion problem were studied by Bot et al. [20], Lin and Jordan [46], but they did not consider anchor acceleration. Ryu et al. [59] considered continuous-time analysis of anchor acceleration, but only with limited cases $\dot{X}(t) = -\mathbf{A}(X(t)) - \frac{\gamma}{t}(X - X_0)$ for $\gamma \geq 1$. In this paper, we provide a unified continuous-time analysis for anchor acceleration with generalized anchor coefficient.

B Proof of Theorem 2.2

B.1 Proof of uniqueness

Proof of uniqueness is immediate from Lemma 2.3, we first prove the lemma.

Proof of Lemma 2.3. It is trivial for $t = 0$, so we may assume $t > 0$.

By monotonicity of \mathbb{A} and Young's inequality, we get the following inequality.

$$\begin{aligned} \frac{d}{dt} \|X(t) - Y(t)\|^2 &= 2 \langle \dot{X}(t) - \dot{Y}(t), X(t) - Y(t) \rangle \\ &= -2 \langle \tilde{\mathbb{A}}(X(t)) - \tilde{\mathbb{A}}(Y(t)) + \beta(t)(X(t) - Y(t)) - \beta(t)(X_0 - Y_0), X(t) - Y(t) \rangle \\ &\leq -2\beta(t) \|X(t) - Y(t)\|^2 + 2\beta(t) \langle X_0 - Y_0, X(t) - Y(t) \rangle \\ &\leq -2\beta(t) \|X(t) - Y(t)\|^2 + \beta(t) \left(\|X_0 - Y_0\|^2 + \|X(t) - Y(t)\|^2 \right) \\ &= -\beta(t) \|X(t) - Y(t)\|^2 + \beta(t) \|X_0 - Y_0\|^2. \end{aligned}$$

Now define $C(t) = e^{\int_0^t \beta(s) ds}$ for some $v > 0$, then we see $\dot{C}(t) = C(t)\beta(t)$. Moving $-\beta(t) \|X(t) - Y(t)\|^2$ to the left hand side and multiplying both sides by $C(t)$, we have

$$\begin{aligned} \frac{d}{dt} \left(C(t) \|X(t) - Y(t)\|^2 \right) &= C(t) \frac{d}{dt} \|X(t) - Y(t)\|^2 + C(t)\beta(t) \|X(t) - Y(t)\|^2 \\ &\leq C(t)\beta(t) \|X_0 - Y_0\|^2 = \frac{d}{dt} C(t) \|X_0 - Y_0\|^2 \end{aligned}$$

Integrating both sides from $\epsilon > 0$ to t we have

$$C(t) \|X(t) - Y(t)\|^2 - C(\epsilon) \|X(\epsilon) - Y(\epsilon)\|^2 \leq C(t) \|X_0 - Y_0\|^2 - C(\epsilon) \|X_0 - Y_0\|^2.$$

As C is nonnegative and nondecreasing, $\lim_{\epsilon \rightarrow 0+} C(\epsilon)$ exists. Taking limit $\epsilon \rightarrow 0+$ both sides we have

$$C(t) \|X(t) - Y(t)\|^2 \leq C(t) \|X_0 - Y_0\|^2.$$

Finally, dividing both sides by $C(t)$ we conclude

$$\|X(t) - Y(t)\|^2 \leq \|X_0 - Y_0\|^2. \quad \square$$

Proof for uniqueness can be done to generalized case (3).

Theorem B.1 (Uniqueness of solutions). *If the solution for (3) exists, it is unique.*

Proof. Suppose X_1, X_2 are solutions of (3) with same initial value X_0 . By Lemma 2.3, we have

$$\|X_1(t) - X_2(t)\| \leq \|X_0 - X_0\| = 0$$

Therefore $X_1(t) = X_2(t)$ for all $t \in [0, \infty)$, we get the desired result. \square

As (6) is special case of (3), uniqueness proof for Theorem 2.2 follows from Theorem B.1.

B.2 Proof of existence

The proof of existence needs tedious work due to the singularity of $\frac{\gamma}{t^p}$ at $t = 0$, we provide our proof through subsections. Before we start, we provide a short outline of the proof. The proof is basically based on the proof provided in [65] and [11].

We first prove the case \mathbb{A} is Lipschitz. The differential inclusion becomes ODE when \mathbb{A} is Lipschitz, we can adopt similar argument done in [65]. We consider series of ODEs Lipschitz with respect to X , approximating $\dot{X}(t) = -\mathbb{A}(X(t)) - \frac{\gamma}{t^p}(X_0 - X(t))$. The approximated ODEs have solutions by classical theory of ODE, we obtain the true solution by considering proper subsequence of solutions.

As the solution for Lipschitz \mathbb{A} is obtained, we can adopt similar argument done in [11]. We first consider solution X_λ with Yosida approximation $\mathbb{A}_\lambda = \frac{1}{\lambda}(\mathbb{I} - (\mathbb{I} + \lambda\mathbb{A})^{-1})$, which is an approximation of \mathbb{A} that is Lipschitz continuous. Then we can obtain a subsequence X_{λ_n} converging to original differential inclusion.

B.2.1 Existence proof for Lipschitz A

Since we will approximate A with Lipschitz continuous functions, we first consider the ODE with Lipschitz continuous A.

Theorem B.2. (Existence of solution for Lipschitz A)

Suppose $\tilde{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is L -Lipschitz continuous function. Consider the differential equation, with initial value condition $\tilde{X}(0) = X_0 \in \text{dom}(\tilde{A})$,

$$\dot{\tilde{X}}(t) = -\tilde{A}(\tilde{X}(t)) - \frac{\gamma}{t^p}(\tilde{X}(t) - X_0). \quad (12)$$

where $\gamma, p > 0$. Then there exists a unique solution $\tilde{X} \in C^1([0, \infty), \mathbb{R}^n)$ that satisfies (12) for all $t \in (0, \infty)$. Moreover, for $\dot{\tilde{X}}(0)$ defined by $\dot{\tilde{X}}(0) = \lim_{t \rightarrow 0^+} \frac{\tilde{X}(t) - X_0}{t}$, following is true

$$\dot{\tilde{X}}(0) = \begin{cases} -\tilde{A}(X_0) & \text{if } 0 < p < 1 \\ -\frac{1}{\gamma+1}\tilde{A}(X_0) & \text{if } p = 1 \\ 0 & \text{if } p > 1. \end{cases}$$

The proof need some preparation. We will think of approximated solutions \tilde{X}_δ , obtain a sequence that converges to solution \tilde{X} . Thus we first define \tilde{X}_δ . From now, we will denote \tilde{A} as a L -Lipschitz monotone function.

Definition B.3. Let $0 < \delta < 1$. Consider

$$\dot{\tilde{X}}_\delta(t) = \begin{cases} -\tilde{A}(\tilde{X}_\delta(t)) - \frac{\gamma}{\delta^p}(\tilde{X}_\delta(t) - X_0) & 0 \leq t < \delta \\ -\tilde{A}(\tilde{X}_\delta(t)) - \frac{\gamma}{t^p}(\tilde{X}_\delta(t) - X_0) & t > \delta \end{cases} \quad (13)$$

Since right hand side above is $(L + \frac{\gamma}{\delta^p})$ -Lipschitz with respect to \tilde{X}_δ , the solution uniquely exists by classical ODE theory. Define the solution as \tilde{X}_δ . Then for positive sequence $\{\delta_m\}_{m \in \mathbb{N}}$ such that $\delta_m < 1$ and $\lim_{m \rightarrow \infty} \delta_m = 0$, consider sequence $\{\tilde{X}_{\delta_m}\}_{m \in \mathbb{N}}$.

Before we start, we prove a useful lemma we will widely use for the cases that operator is Lipschitz continuous.

Lemma B.4. Let $\tilde{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ a Lipschitz continuous function. Suppose $\beta : D \rightarrow [0, \infty)$ be a continuous function with $D \subset [0, \infty)$. Consider differential equation

$$\dot{\tilde{X}} = -\tilde{A}(\tilde{X}) - \beta(t)(\tilde{X} - X_0),$$

with $X_0 \in \mathbb{R}^n$. Let $\tilde{X} : D \rightarrow \mathbb{R}^n$ be a differentiable curve that satisfies above equation for $t \in D$. Then for all $0 \leq a < b$ such that $[a, b] \subset D$, \tilde{X} and the composition $\tilde{A} \circ \tilde{X} : [a, b] \rightarrow \mathbb{R}^n$ are Lipschitz continuous.

Moreover if $\dot{\beta}(t)$ is well-defined and bounded for almost all $t \in [a, b]$, then $\dot{\tilde{X}}$ is Lipschitz continuous in $[a, b]$. Thus if β is twice differentiable, $\dot{\tilde{X}}$ is Lipschitz continuous.

As Lipschitz continuous functions, \tilde{X} , $\tilde{A} \circ \tilde{X}$ and $\dot{\tilde{X}}$ are absolutely continuous functions.

Proof. We first prove \tilde{X} is Lipschitz continuous. As $\tilde{A}, \tilde{X}, \beta$ are continuous in $[a, b]$ we see $\tilde{A}(\tilde{X}(t)) + \beta(t)(\tilde{X}(t) - X_0)$ is continuous in $[a, b]$, thus

$$M_1 = \max_{t \in [a, b]} \left\| \dot{\tilde{X}}(t) \right\| = \max_{t \in [a, b]} \left\| \tilde{A}(\tilde{X}(t)) + \beta(t)(\tilde{X}(t) - X_0) \right\|$$

exists. Since its derivative is bounded by $M_1 < \infty$, we have \tilde{X} is M_1 -Lipshitz continuous. As composition of two Lipschitz continuous functions, $\tilde{A} \circ \tilde{X}$ is Lipschitz continuous.

First observe if β is twice differentiable, $\dot{\beta}$ is bounded in $[a, b]$ as it is continuous. Now if $\dot{\beta}(t)$ is bounded for $t \in [a, b]$, i.e. $|\dot{\beta}(t)| \leq M_2$ for some $M_2 > 0$, for $M_3 = \max_{t \in [a, b]} |\beta(t)|$ we have

$$\left\| \frac{d}{dt} \left(\beta(t)(\tilde{X}(t) - X_0) \right) \right\| = \left\| \dot{\beta}(t)(\tilde{X}(t) - X_0) \right\| + \left\| \beta(t)\dot{\tilde{X}}(t) \right\| \leq M_2 M_1 |b - a| + M_3 M_1$$

Therefore $\beta(t)(\tilde{X}(t) - X_0)$ is $(M_2 M_1 |b - a| + M_3 M_1)$ -Lipschitz continuous. Thus as a sum of Lipschitz continuous functions, $\dot{\tilde{X}}$ is Lipschitz continuous. \square

We will show for every $T > 0$, the set of derivatives $\{\dot{\tilde{X}}_\delta \mid \delta \in (0, 1)\}$ is uniformly bounded on $[0, T]$ and $\{\tilde{X}_{\delta_m}\}_{m \in \mathbb{N}}$ converges uniformly on $[0, T]$. We first prove the boundedness of derivatives. To do so, we first prove a useful lemma.

Lemma B.5. For $0 < a < b$, suppose $\tilde{X} : [a, b] \rightarrow \mathbb{R}^n$ satisfies ODE (12) for $t \in [a, b]$. Define $\tilde{U} : [a, b] \rightarrow \mathbb{R}$ as

$$\begin{aligned}\tilde{U}_1(t) &= \left\| \dot{\tilde{X}}(t) \right\|^2 + \frac{\gamma p}{t^{p+1}} \left\| \tilde{X}(t) - X_0 \right\|^2 \\ \tilde{U}_2(t) &= \frac{1}{t^{p-1}} \left\| \dot{\tilde{X}}(t) \right\|^2 + \frac{\gamma p}{t^{2p}} \left\| \tilde{X}(t) - X_0 \right\|^2\end{aligned}$$

Then \tilde{U}_1 is nonincreasing if for $0 < p \leq 1$, and \tilde{U}_2 is nonincreasing for $p > 1$.

Proof. From Lemma B.4 we can check \tilde{U}_1 and \tilde{U}_2 are absolutely continuous in $[a, b]$. Therefore it is enough to check the derivative is nonpositive for almost all t . From Lemma B.4, we can differentiate (12) both sides. Thus for almost all t we have

$$\ddot{\tilde{X}}(t) = -\frac{d}{dt} \tilde{A}(\tilde{X}(t)) + \frac{\gamma p}{t^{p+1}} (\tilde{X}(t) - X_0) - \frac{\gamma}{t^p} \dot{\tilde{X}}(t).$$

Recall from monotonicity of \tilde{A} and (1), we know $\left\langle \dot{\tilde{X}}(t), \frac{d}{dt} \tilde{A}(\tilde{X}(t)) \right\rangle \geq 0$ for almost all t . Therefore for almost all t ,

(i) $0 < p \leq 1$

$$\begin{aligned}\dot{\tilde{U}}_1(t) &= 2 \left\langle \dot{\tilde{X}}(t), \ddot{\tilde{X}}(t) \right\rangle + \frac{2\gamma p}{t^{p+1}} \left\langle \dot{\tilde{X}}(t), \tilde{X}(t) - X_0 \right\rangle - \frac{\gamma p(p+1)}{t^{p+2}} \left\| \tilde{X}(t) - X_0 \right\|^2 \\ &= -2 \left\langle \dot{\tilde{X}}(t), \frac{d}{dt} \tilde{A}(\tilde{X}(t)) \right\rangle - \frac{2\gamma}{t^p} \left\| \dot{\tilde{X}}(t) \right\|^2 + \frac{2\gamma p}{t^{p+1}} \left\langle \dot{\tilde{X}}(t), \tilde{X}(t) - X_0 \right\rangle \\ &\quad + \frac{2\gamma p}{t^{p+1}} \left\langle \dot{\tilde{X}}(t), \tilde{X}(t) - X_0 \right\rangle - \frac{2\gamma p^2}{t^{p+2}} \left\| \tilde{X}(t) - X_0 \right\|^2 - \frac{\gamma p(1-p)}{t^{p+2}} \left\| \tilde{X}(t) - X_0 \right\|^2 \\ &= -2 \left\langle \dot{\tilde{X}}(t), \frac{d}{dt} \tilde{A}(\tilde{X}(t)) \right\rangle - \frac{2\gamma}{t^p} \left\| \dot{\tilde{X}}(t) - \frac{p}{t} (\tilde{X} - X_0) \right\|^2 - \frac{\gamma p(1-p)}{t^{p+2}} \left\| \tilde{X}(t) - X_0 \right\|^2 \\ &\leq 0.\end{aligned}$$

(ii) $p > 1$

$$\begin{aligned}\dot{\tilde{U}}_2(t) &= -\frac{p-1}{t^p} \left\| \dot{\tilde{X}}(t) \right\|^2 + \frac{2}{t^{p-1}} 2 \left\langle \dot{\tilde{X}}(t), \ddot{\tilde{X}}(t) \right\rangle + \frac{2\gamma p}{t^{2p}} \left\langle \dot{\tilde{X}}(t), \tilde{X} - X_0 \right\rangle - \frac{2\gamma p}{t^{2p+1}} \left\| \tilde{X}(t) - X_0 \right\|^2 \\ &= -\frac{p-1}{t^p} \left\| \dot{\tilde{X}}(t) \right\|^2 - \frac{2}{t^{p-1}} \left\langle \dot{\tilde{X}}(t), \frac{d}{dt} \tilde{A}(\tilde{X}(t)) \right\rangle \\ &\quad - \frac{2\gamma}{t^{2p-1}} \left\| \dot{\tilde{X}}(t) \right\|^2 + \frac{2\gamma p}{t^{2p}} \left\langle \dot{\tilde{X}}(t), \tilde{X} - X_0 \right\rangle + \frac{2\gamma p}{t^{2p}} \left\langle \dot{\tilde{X}}(t), \tilde{X} - X_0 \right\rangle - \frac{2\gamma p}{t^{2p+1}} \left\| \tilde{X}(t) - X_0 \right\|^2 \\ &= -\frac{p-1}{t^p} \left\| \dot{\tilde{X}}(t) \right\|^2 - \frac{2}{t^{p-1}} \left\langle \dot{\tilde{X}}(t), \frac{d}{dt} \tilde{A}(\tilde{X}(t)) \right\rangle - \frac{2\gamma}{t^{2p-1}} \left\| \dot{\tilde{X}}(t) - \frac{p}{t} (\tilde{X} - X_0) \right\|^2 \\ &\leq 0.\end{aligned}$$

□

We now are ready to prove uniform boundedness of derivatives $\dot{\tilde{X}}_\delta$.

Lemma B.6. Take $T > 0$. For all $t \in [0, T]$ and $\delta \in (0, 1)$, below inequality holds.

$$\left\| \dot{\tilde{X}}_\delta(t) \right\| \leq \tilde{M}_{dot}(T) = \begin{cases} \sqrt{1 + \gamma p} \left\| \tilde{A}(X_0) \right\| & 0 < p \leq 1 \\ \sqrt{T^{p-1} \left(\frac{1}{2\gamma} + 2p(2L^2 + 2\gamma + 1) \right)} \left\| \tilde{A}(X_0) \right\| & p \geq 1. \end{cases}$$

Proof. We prove first statement by considering two cases.

(1) $t \in [0, \delta]$

From Lemma B.4, we know \dot{X}_δ is absolutely continuous and so $\|\dot{X}_\delta\|^2$ is absolutely continuous as well. Differentiating (13), for almost all $t \in (0, \delta)$ we have

$$\ddot{X}_\delta = -\frac{d}{dt}\tilde{A}(\tilde{X}_\delta(t)) - \frac{\gamma}{\delta^p}\dot{X}_\delta.$$

Now for almost all $t \in (0, \delta)$,

$$\frac{d}{dt}\|\dot{X}_\delta\|^2 = 2\langle \dot{X}_\delta, \ddot{X}_\delta \rangle = -2\left\langle \dot{X}_\delta, \frac{d}{dt}\tilde{A}(\tilde{X}_\delta(t)) \right\rangle - \frac{2\gamma}{\delta^p}\|\dot{X}_\delta\|^2 \leq -\frac{2\gamma}{\delta^p}\|\dot{X}_\delta\|^2 \leq 0.$$

(i) $0 < p \leq 1$

From above we know $\|\dot{X}_\delta\|$ is nonincreasing in $t \in [0, \delta]$, we have $\|\dot{X}_\delta(t)\| \leq \|\dot{X}_\delta(0)\| = \|\tilde{A}(X_0)\|$.

(ii) $p > 1$

Integrating $\frac{d}{dt}\|\dot{X}_\delta\|^2 \leq -\frac{2\gamma}{\delta^p}\|\dot{X}_\delta\|^2$ we have

$$\begin{aligned} \|\dot{X}_\delta(t)\|^2 - \|\dot{X}_\delta(0)\|^2 &\leq -\frac{2\gamma}{\delta^p} \int_0^t \|\dot{X}_\delta(s)\|^2 ds \\ &\leq -\frac{2\gamma}{\delta^p} \int_0^t \|\dot{X}_\delta(t)\|^2 ds = -\frac{2\gamma t}{\delta^p} \|\dot{X}_\delta(t)\|^2. \end{aligned}$$

Organizing with respect to $\|\dot{X}_\delta(t)\|$ we have

$$\|\dot{X}_\delta(t)\| \leq \frac{\|\dot{X}_\delta(0)\|}{\sqrt{1 + \frac{2\gamma t}{\delta^p}}} = \sqrt{\frac{\delta^p}{\delta^p + 2\gamma t}} \|\tilde{A}(X_0)\| \leq \sqrt{\frac{\delta^p}{2\gamma t}} \|\tilde{A}(X_0)\|.$$

(2) $t \geq \delta$

Note for $t > \delta$ (12) holds, we can apply Lemma B.5.

(i) $0 < p \leq 1$

From (i) we know $\|\dot{X}_\delta(t)\| \leq \|\tilde{A}(X_0)\|$ for $t \in [0, \delta]$. Therefore $\|\dot{X}_\delta(\delta)\| \leq \|\tilde{A}(X_0)\|$ and we see

$$\|\tilde{X}_\delta(\delta) - X_0\| = \left\| \int_0^\delta \dot{X}_\delta(s) ds \right\| \leq \int_0^\delta \|\dot{X}_\delta(s)\| ds \leq \delta \|\tilde{A}(X_0)\|.$$

From Lemma B.5 we know $\tilde{U}_1(t) = \|\dot{X}_\delta(t)\|^2 + \frac{\gamma p}{t^{p+1}} \|\tilde{X}_\delta(t) - X_0\|^2$ is nonincreasing. Therefore $\|\dot{X}_\delta(t)\|^2 \leq \tilde{U}_1(\delta)$ for $t \geq \delta$. Since $\delta < 1$ we have

$$\begin{aligned} \|\dot{X}_\delta(t)\| &\leq \sqrt{\tilde{U}_1(\delta)} = \sqrt{\|\dot{X}_\delta(\delta)\|^2 + \frac{\gamma p}{\delta^{p+1}} \|\tilde{X}_\delta(\delta) - X_0\|^2} \\ &\leq \sqrt{\|\tilde{A}(X_0)\|^2 + \gamma p \delta^{1-p} \|\tilde{A}(X_0)\|^2} \leq \sqrt{1 + \gamma p} \|\tilde{A}(X_0)\|. \end{aligned}$$

(ii) $p > 1$

From (i) we know $\|\dot{X}_\delta(t)\| \leq \sqrt{\frac{\delta^p}{2\gamma t}} \|\tilde{A}(X_0)\|$ for $t \in [0, \delta]$. Therefore $\|\dot{X}_\delta(\delta)\| \leq \sqrt{\frac{\delta^{p-1}}{2\gamma}} \|\tilde{A}(X_0)\|$ and we see

$$\|\tilde{X}_\delta(\delta) - X_0\| \leq \int_0^\delta \|\dot{X}_\delta(s)\| ds \leq \int_0^\delta \sqrt{\frac{\delta^p}{2\gamma s}} \|\tilde{A}(X_0)\| ds = \sqrt{\frac{2\delta^{p+1}}{\gamma}} \|\tilde{A}(X_0)\|.$$

Applying (13) and recalling the fact \tilde{A} is L -Lipschitz, we have

$$\begin{aligned}
\frac{\gamma^2}{\delta^{2p}} \|X_\delta(\delta) - X_0\|^2 &= \left\| \tilde{A}(X_\delta(\delta)) + \dot{X}_\delta(\delta) \right\|^2 = \left\| \tilde{A}(X_\delta(\delta)) - \tilde{A}(X_0) + \tilde{A}(X_0) + \dot{X}_\delta(\delta) \right\|^2 \\
&\leq 2 \left\| \tilde{A}(X_\delta(\delta)) - \tilde{A}(X_0) \right\|^2 + 2 \left\| \tilde{A}(X_0) + \dot{X}_\delta(\delta) \right\|^2 \\
&\leq 2L^2 \left\| \tilde{X}_\delta(\delta) - X_0 \right\|^2 + 4 \left\| \tilde{A}(X_0) \right\|^2 + 4 \left\| \dot{X}_\delta(\delta) \right\|^2 \\
&\leq \left(\frac{4}{\gamma} L^2 \delta^{p+1} + 4 + \frac{2\delta^{p-1}}{\gamma} \right) \left\| \tilde{A}(X_0) \right\|^2.
\end{aligned}$$

From Lemma B.5 we know $\tilde{U}_2(t) = \frac{1}{t^{p-1}} \left\| \dot{X}(t) \right\|^2 + \frac{\gamma p}{t^{2p}} \left\| \tilde{X}(t) - X_0 \right\|^2$ is nonincreasing. Therefore $\frac{\left\| \dot{X}_\delta(t) \right\|^2}{t^{p-1}} \leq \tilde{U}_2(\delta)$ for $t \geq \delta$. Since $\delta < 1$ we have

$$\begin{aligned}
\frac{\left\| \dot{X}_\delta(t) \right\|}{\sqrt{t^{p-1}}} &\leq \sqrt{\tilde{U}_2(\delta)} = \sqrt{\frac{\left\| \dot{X}_\delta(\delta) \right\|^2}{\delta^{p-1}} + \frac{\gamma p}{\delta^{2p}} \left\| \tilde{X}_\delta(\delta) - X_0 \right\|^2} \\
&\leq \sqrt{\frac{\left\| \tilde{A}(X_0) \right\|^2}{2\gamma} + p(4L^2\delta^{p+1} + 4\gamma + 2\delta^{p-1}) \left\| \tilde{A}(X_0) \right\|^2} \leq \sqrt{\frac{1}{2\gamma} + 2p(2L^2 + 2\gamma + 1)} \left\| \tilde{A}(X_0) \right\|.
\end{aligned}$$

Therefore for $t \in [0, T]$ we have

$$\left\| \dot{X}_\delta(t) \right\| \leq \sqrt{T^{p-1} \left(\frac{1}{2\gamma} + 2p(2L^2 + 2\gamma + 1) \right)} \left\| \tilde{A}(X_0) \right\|$$

From (1) and (2), we get the desired result. \square

We now show the sequence $\left\{ \tilde{X}_{\delta_m} \right\}_{m \in \mathbb{N}}$ converges uniformly on $[0, T]$ for every $T > 0$. It suffices to prove following proposition.

Proposition B.7. *For $T > 0$, the sequence $\left\{ \tilde{X}_{\delta_m} \right\}_{m \in \mathbb{N}}$ is a Cauchy sequence with respect to supremum norm on $[0, T]$.*

Proof. Take $\epsilon > 0$. We want to show, there is $N \in \mathbb{N}$ such that if $n, m > N$ then $\left\| \tilde{X}_{\delta_n}(t) - \tilde{X}_{\delta_m}(t) \right\| \leq \epsilon$ for all $t \in [0, \infty)$.

Define $d_{\delta, \nu}(t) = \frac{1}{2} \left\| \tilde{X}_\delta(t) - \tilde{X}_\nu(t) \right\|^2$. Without loss of generality, we may assume $\delta \geq \nu$. With $M_{dot}(T)$ defined in Lemma B.6, we will show $d_{\delta, \nu}(t) \leq 2\delta^2 \tilde{M}_{dot}(T)^2$.

First consider the case $0 \leq t \leq \delta$. By Lemma B.6, we have

$$\begin{aligned}
\left\| \tilde{X}_\delta(t) - \tilde{X}_\nu(t) \right\| &\leq \int_0^t \left\| \dot{X}_\delta(s) - \dot{X}_\nu(s) \right\| ds \leq \int_0^t \left(\left\| \dot{X}_\delta(s) \right\| + \left\| \dot{X}_\nu(s) \right\| \right) ds \\
&\leq \int_0^t 2\tilde{M}_{dot}(T) ds = 2t\tilde{M}_{dot}(T)
\end{aligned}$$

Thus for $t \in [0, \delta]$ we have

$$d_{\delta, \nu}(t) \leq 2t^2 \tilde{M}_{dot}(T)^2 \leq 2\delta^2 \tilde{M}_{dot}(T)^2.$$

Now we consider the case $t \geq \delta$. By monotonicity of \tilde{A} , we have

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \left\| \tilde{X}_\delta(t) - \tilde{X}_\nu(t) \right\|^2 &= \left\langle \dot{X}_\delta(t) - \dot{X}_\nu(t), \tilde{X}_\delta(t) - \tilde{X}_\nu(t) \right\rangle \\ &= \left\langle -\left(\tilde{A}(\tilde{X}_\delta(t)) - \tilde{A}(\tilde{X}_\nu(t)) \right) - \frac{\gamma}{t^p} (X_\delta(t) - X_0) + \frac{\gamma}{t^p} (X_\nu(t) - X_0), X_\delta(t) - X_\nu(t) \right\rangle \\ &\leq \left\langle -\frac{\gamma}{t^p} (X_\delta(t) - X_0) + \frac{\gamma}{t^p} (X_\nu(t) - X_0), X_\delta(t) - X_\nu(t) \right\rangle \\ &= -\frac{\gamma}{t^p} \|X_\delta(t) - X_\nu(t)\|^2 \\ &\leq 0. \end{aligned}$$

Thus we have $d_{\delta,\nu}(t) \leq d_{\delta,\nu}(\delta)$ for $t \geq \delta$.

Now combining two cases, we have

$$d_{\delta,\nu}(t) \leq 2\delta^2 \tilde{M}_{dot}(T)^2 \quad (14)$$

Since $\lim_{k \rightarrow \infty} \delta_n = 0$, there is $N \in \mathbb{N}$ such that $m > n > N$ implies $d_{\delta_n, \delta_m} \leq \frac{\epsilon^2}{2}$, we're done. \square

We are now ready to prove Theorem B.2.

Proof. (1) *Existence of solution.*

From Proposition B.7, we know $\left\{ \tilde{X}_{\delta_m} \right\}_{m \in \mathbb{N}}$ converging uniformly on $[0, T]$ for every $T > 0$. Denote the limit as \tilde{X} , i.e. define $\tilde{X}: [0, \infty) \rightarrow \mathbb{R}^n$ as

$$\lim_{m \rightarrow \infty} \tilde{X}_{\delta_m}(t) = \tilde{X}(t).$$

We can check $\tilde{X}(0) = X_0$ easily since $\tilde{X}_{\delta_m}(0) = X_0$ for all $m \in \mathbb{N}$. It remains to show \tilde{X} satisfies (12).

Take $t > 0$. We wish to show

$$\lim_{h \rightarrow 0} \left\| \frac{\tilde{X}(t+h) - \tilde{X}(t)}{h} + \tilde{A}(\tilde{X}(t)) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right\| = 0.$$

Consider h, δ, T such that $0 < |h| < t, 0 < \delta < \min\{1, t - |h|\}$ and $T > t + |h|$. Then $(t - |h|, t + |h|) \in [0, T]$ and $\dot{X}_\delta(t) = -\tilde{A}(\tilde{X}_\delta(t)) - \frac{\gamma}{t^p} (\tilde{X}_\delta(t) - X_0)$. Consider inequality

$$\begin{aligned} &\left\| \frac{\tilde{X}(t+h) - \tilde{X}(t)}{h} + \tilde{A}(\tilde{X}(t)) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right\| \\ &\leq \left\| \frac{\tilde{X}(t+h) - \tilde{X}(t)}{h} - \frac{\tilde{X}_\delta(t+h) - \tilde{X}_\delta(t)}{h} \right\| + \left\| \frac{\tilde{X}_\delta(t+h) - \tilde{X}_\delta(t)}{h} - \dot{X}_\delta(t) \right\| + \left\| \dot{X}_\delta(t) + \tilde{A}(\tilde{X}(t)) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right\|. \end{aligned}$$

We now show right hand side goes to zero as $h \rightarrow 0$. The point of the proof is, \tilde{X}_δ converges uniformly to \tilde{X} and \tilde{X}_δ is uniformly bounded on $[t - |h|, T]$.

From (14) we have

$$\begin{aligned} &\left\| \frac{\tilde{X}(t+h) - \tilde{X}(t)}{h} - \frac{\tilde{X}_\delta(t+h) - \tilde{X}_\delta(t)}{h} \right\| \\ &\leq \frac{1}{|h|} \left(\left\| \tilde{X}(t+h) - \tilde{X}_\delta(t+h) \right\| + \left\| \tilde{X}(t) - \tilde{X}_\delta(t) \right\| \right) \leq 4 \frac{\delta}{|h|} \tilde{M}_{dot}(T). \end{aligned}$$

Also from (14) and since \tilde{A} is L -Lipschitz continuous,

$$\begin{aligned} \left\| \dot{X}_\delta(t) + \tilde{A}(\tilde{X}(t)) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right\| &= \left\| -\left(\tilde{A}(\tilde{X}_\delta(t)) + \frac{\gamma}{t^p} (\tilde{X}_\delta(t) - X_0) \right) + \tilde{A}(\tilde{X}(t)) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right\| \\ &\leq L \left\| \tilde{X}_\delta(t) - \tilde{X}(t) \right\| + \frac{\gamma}{t^p} \left\| \tilde{X}_\delta(t) - \tilde{X}(t) \right\| \leq 2\delta \left(L + \frac{\gamma}{t^p} \right) \tilde{M}_{dot}(T). \end{aligned}$$

Now from Lemma B.4 we have $\dot{X}_\delta(t)$ is absolutely continuous, thus

$$\begin{aligned} \left\| \frac{\tilde{X}_\delta(t+h) - \tilde{X}_\delta(t)}{h} - \dot{X}_\delta(t) \right\| &= \left\| \frac{\int_t^{t+h} (\dot{\tilde{X}}_\delta(s) - \dot{\tilde{X}}_\delta(t)) ds}{h} \right\| \\ &= \frac{1}{|h|} \left\| \int_t^{t+h} \int_t^s \ddot{\tilde{X}}_\delta(u) du ds \right\| \leq \frac{1}{|h|} \left| \int_t^{t+h} \int_t^s \|\ddot{\tilde{X}}_\delta(u)\| du ds \right| \end{aligned}$$

Observe, for almost every $u \in [t - |h|, t + |h|] \subset [0, T]$ by Lemma B.6 we have

$$\begin{aligned} \|\ddot{\tilde{X}}_\delta(u)\| &= \left\| -\frac{d}{du} \tilde{A}(\tilde{X}_\delta(u)) + \frac{p\gamma}{u^{p+1}} (\tilde{X}_\delta(u) - X_0) - \frac{\gamma}{u^p} \dot{\tilde{X}}_\delta(u) \right\| \\ &= \left\| \frac{d}{du} \tilde{A}(\tilde{X}_\delta(u)) \right\| + \left\| \frac{p\gamma}{u^{p+1}} \int_0^u \dot{\tilde{X}}_\delta(v) dv \right\| + \left\| \frac{\gamma}{u^p} \dot{\tilde{X}}_\delta(u) \right\| \\ &\leq L\tilde{M}_{dot}(T) + \frac{p\gamma}{(t-|h|)^{p+1}} \int_0^T \tilde{M}_{dot}(T) dv + \frac{\gamma}{(t-|h|)^p} \tilde{M}_{dot}(T) \\ &= \left(L + \frac{p\gamma T}{(t-|h|)^{p+1}} + \frac{\gamma}{(t-|h|)^p} \right) \tilde{M}_{dot}(T) =: M. \end{aligned}$$

Note M is independent of h or δ . While obtaining the inequality, we used the fact that $\tilde{A}(\tilde{X}_\delta(\cdot))$ is $L\tilde{M}_{dot}(T)$ -Lipschitz continuous in $[0, T]$. Now

$$\frac{1}{|h|} \left| \int_t^{t+h} \int_t^s \|\ddot{\tilde{X}}_\delta(u)\| du ds \right| \leq \frac{1}{|h|} \left| \int_t^{t+h} \int_t^s M du ds \right| = \frac{1}{|h|} M \left| \int_t^{t+h} (s-t) ds \right| = \frac{|h|}{2} M.$$

Now consider $\delta = h^2$ with h small enough that satisfies $|h| < 1$ and $|h| + h^2 < t$. Then the conditions $|h| < t$, $0 < \delta < \min\{1, t - |h|\}$ hold, above arguments are valid. Gathering above results, we have

$$\begin{aligned} &\left\| \frac{\tilde{X}(t+h) - \tilde{X}(t)}{h} + \tilde{A}(\tilde{X}(t)) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right\| \\ &\leq 2\sqrt{2}|h|\tilde{M}_{dot}(T) + \sqrt{2}|h|^2 \left(L + \frac{\gamma}{t^p} \right) \tilde{M}_{dot}(T) + \frac{|h|}{2} M = \mathcal{O}(|h|). \end{aligned}$$

which implies the desired result.

- (2) *The value and continuity of $\dot{\tilde{X}}(t)$ at $t = 0$.*
Define $C(t)$ as

$$C(t) = \begin{cases} t^\gamma & p = 1 \\ e^{\frac{\gamma}{1-p} t^{1-p}} & p > 0, p \neq 1. \end{cases}$$

Then we see for $t > 0$

$$\frac{d}{dt} \left(C(t) (\tilde{X}(t) - X_0) \right) = C(t) \left(\dot{\tilde{X}}(t) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right) = -C(t) \tilde{A}(\tilde{X}(t)).$$

Integrating both sides from $\epsilon > 0$ to t we have

$$C(t)(\tilde{X}(t) - X_0) - C(\epsilon)(\tilde{X}(\epsilon) - X_0) = - \int_\epsilon^t C(s) \tilde{A}(\tilde{X}(s)) ds.$$

As C is a nondecreasing function and bounded below by 0, $\lim_{\epsilon \rightarrow 0+} C(\epsilon)$ exists, taking limit $\epsilon \rightarrow 0+$ we have

$$C(t)(\tilde{X}(t) - X_0) = - \int_0^t C(s) \tilde{A}(\tilde{X}(s)) ds.$$

Dividing both sides by $tC(t)$, with change of variable $v = s/t$, we have

$$\frac{\tilde{X}(t) - X_0}{t} = - \int_0^t \frac{C(s)}{C(t)} \tilde{A}(\tilde{X}(s)) \frac{ds}{t} = - \int_0^1 \frac{C(tv)}{C(t)} \tilde{A}(\tilde{X}(tv)) dv.$$

Observe

$$\frac{C(tv)}{C(t)} = \begin{cases} v^\gamma & p = 1 \\ e^{\frac{\gamma}{1-p} t^{1-p} (v^{1-p} - 1)} & p \neq 1, p > 0. \end{cases}$$

Note $\frac{\gamma}{1-p} t^{1-p} (v^{1-p} - 1) \leq 0$ for $v \in [0, 1]$, since $\frac{\gamma}{1-p}$ and $(v^{1-p} - 1)$ has opposite sign either $0 < p < 1$ or $p > 1$. Therefore $e^{\frac{\gamma}{1-p} t^{1-p} (v^{1-p} - 1)} \leq 1$. Also, $\tilde{A}(\tilde{X}(tv))$ is bounded for $v \in [0, 1]$ since $A(\tilde{X}(\cdot))$ is continuous by Lemma B.4.

So we can apply dominated convergence theorem and take limit $t \rightarrow 0+$. Since

$$\lim_{t \rightarrow 0+} \frac{C(tv)}{C(t)} = \begin{cases} 1 & 0 < p < 1 \\ v^\gamma & p = 1 \\ 0 & p > 1 \end{cases}$$

for $v \neq 0$, we have

$$\dot{\tilde{X}}(0) = \lim_{t \rightarrow 0+} \frac{\tilde{X}(t) - X_0}{t} = - \int_0^1 \lim_{t \rightarrow 0+} \left(\frac{C(tv)}{C(t)} \tilde{A}(\tilde{X}(tv)) \right) dv = \begin{cases} -\tilde{A}(X_0) & 0 < p < 1 \\ -\frac{1}{\gamma+1} \tilde{A}(X_0) & p = 1 \\ 0 & p > 1. \end{cases} \quad (15)$$

We now check $\dot{\tilde{X}}(t)$ is continuous at $t = 0$.

(i) $0 < p \leq 1$

For $t > 0$, we know

$$\dot{\tilde{X}}(t) = -\tilde{A}(\tilde{X}(t)) - \frac{\gamma}{t^p} (\tilde{X}(t) - X_0).$$

Observe

$$\lim_{t \rightarrow 0+} \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) = \lim_{t \rightarrow 0+} \left(\gamma t^{1-p} \cdot \frac{\tilde{X}(t) - X_0}{t} \right) = \begin{cases} 0 & 0 < p < 1 \\ -\frac{\gamma}{\gamma+1} \tilde{A}(X_0) & p = 1. \end{cases} \quad (16)$$

Now from ODE $\dot{\tilde{X}}(t) = -\tilde{A}(\tilde{X}(t)) - \frac{\gamma}{t^p} (\tilde{X}(t) - X_0)$, by taking limit $t \rightarrow 0+$ we have

$$\lim_{t \rightarrow 0+} \dot{\tilde{X}}(t) = -\tilde{A}(X_0) - \lim_{t \rightarrow 0+} \left(\frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right) = \begin{cases} -\tilde{A}(X_0) & 0 < p < 1 \\ -\frac{1}{\gamma+1} \tilde{A}(X_0) & p = 1. \end{cases}$$

Therefore $\dot{\tilde{X}}(t)$ is continuous at $t = 0$.

(ii) $p > 1$

For $p > 1$, we don't know the value of $\lim_{t \rightarrow 0} \frac{\gamma}{t^p} (\tilde{X}(t) - X_0)$. Thus we first find the limit.

Let's go back to

$$C(t)(\tilde{X}(t) - X_0) = - \int_0^t C(s) \tilde{A}(\tilde{X}(s)) ds.$$

Recall $\dot{C}(t) = \frac{\gamma}{t^p} C(t)$. By taking integration by parts for the right hand side, we have

$$\int_0^t C(s) \tilde{A}(\tilde{X}(s)) ds = \frac{t^p}{\gamma} C(t) \tilde{A}(\tilde{X}(t)) - \int_0^t p \frac{s^{p-1}}{\gamma} C(s) \tilde{A}(\tilde{X}(s)) ds - \int_0^t \frac{s^p}{\gamma} C(s) \frac{d}{ds} \tilde{A}(\tilde{X}(s)) ds.$$

Where we know $\tilde{A}(\tilde{X}(s))$ is differentiable almost everywhere from Lemma B.4. Now, divide both sides by $\frac{t^p C(t)}{\gamma}$. Then for $s = tv$ we have

$$\begin{aligned} \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) &= -\tilde{A}(\tilde{X}(t)) + \int_0^t \frac{ps^{p-1}}{t^p} \frac{C(s)}{C(t)} \tilde{A}(\tilde{X}(s)) ds + \int_0^t \frac{s^p}{t^p} \frac{C(s)}{C(t)} \left(\frac{d}{ds} \tilde{A}(\tilde{X}(s)) \right) ds \\ &= -\tilde{A}(\tilde{X}(t)) + \int_0^1 pv^{p-1} \frac{C(tv)}{C(t)} \tilde{A}(\tilde{X}(s)) dv + \int_0^1 v^p \frac{C(tv)}{C(t)} \left(\frac{d}{ds} \tilde{A}(\tilde{X}(s)) \right) t dv. \end{aligned}$$

From Proposition B.7 and since \tilde{X} satisfies (12) for $t > 0$, for $s \in (0, t]$ we have

$$\dot{\tilde{X}}(s) = \tilde{A}(\tilde{X}(s)) + \frac{\gamma}{t^p} (\tilde{X}(s) - X_0) = \lim_{m \rightarrow \infty} \left(\tilde{A}(\tilde{X}_{\delta_m}(s)) + \frac{\gamma}{t^p} (\tilde{X}_{\delta_m}(s) - X_0) \right) = \lim_{m \rightarrow \infty} \dot{\tilde{X}}_{\delta_m}(s).$$

From Lemma B.6 we know $\|\dot{\tilde{X}}_{\delta_m}(s)\| \leq \tilde{M}_{dot}(t)$ for $s \in [0, t]$, taking limit $m \rightarrow \infty$ we have $\|\dot{\tilde{X}}(s)\| \leq \tilde{M}_{dot}(t)$ for $s \in [0, t]$. Thus $\tilde{X}(s)$ becomes $\tilde{M}_{dot}(t)$ -Lipschitz continuous in $s \in [0, t]$. And so $(\tilde{A} \circ \tilde{X})(s)$ becomes $L\tilde{M}_{dot}(t)$ -Lipschitz continuous in $s \in [0, t]$, we have $\left\| \frac{d}{ds} \tilde{A}(\tilde{X}(s)) \right\| \leq L\tilde{M}_{dot}(t)$ for almost all $s \in [0, t]$. Moreover $C(tv)/C(t)$ is bounded for $v \in [0, 1]$ since C is a nonnegative nondecreasing function. Therefore we can again apply dominated convergence theorem. Reminding $\lim_{t \rightarrow 0+} \frac{C(tv)}{C(t)} = 0$ for $p > 1$, taking limit $t \rightarrow 0+$ we have

$$\lim_{t \rightarrow 0+} \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) = -\tilde{A}(X_0) + 0 = -\tilde{A}(X_0). \quad (17)$$

Finally we have

$$\lim_{t \rightarrow 0+} \dot{\tilde{X}}(t) = -\tilde{A}(X_0) - \lim_{t \rightarrow 0+} \left(\frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right) = -\tilde{A}(X_0) - (-\tilde{A}(X_0)) = 0 = \dot{\tilde{X}}(0). \quad \square$$

Before we move on to original inclusion, we prove important corollaries that bound $\|\dot{\tilde{X}}(t)\|$ and $\|\tilde{A}(\tilde{X}(t))\|$ uniformly on $[0, T]$. Note the main difference between following corollary and Lemma B.6 is that the dependency on Lipschitz constant L is dropped for the case $p > 1$, which will be crucial in the next section.

Corollary B.8. *Denote \tilde{X} as the solution of (12). Then for $T > 0$, following inequality is true for $t \in [0, T]$.*

$$\|\dot{\tilde{X}}(t)\| \leq \begin{cases} \|\tilde{A}(X_0)\| & 0 < p < 1 \\ \frac{1}{\sqrt{\gamma+1}} \|\tilde{A}(X_0)\| & p = 1 \\ \sqrt{\frac{p}{\gamma} T^{p-1}} \|\tilde{A}(X_0)\| & p > 1. \end{cases}$$

Proof. (i) $0 < p \leq 1$

As \tilde{X} is the solution for (12), from Lemma B.5 we know

$$\tilde{U}_1(t) = \|\dot{\tilde{X}}(t)\|^2 + \frac{\gamma p}{t^{p+1}} \|\tilde{X}(t) - X_0\|^2$$

is a nonincreasing function. From (15), (16) and continuity of $\dot{\tilde{X}}(t)$ at $t = 0$, we have

$$\lim_{\epsilon \rightarrow 0+} \tilde{U}_1(\epsilon) = \begin{cases} \|\tilde{A}(X_0)\|^2 & 0 < p < 1 \\ \frac{1}{\gamma+1} \|\tilde{A}(X_0)\|^2 & p = 1. \end{cases} \quad (18)$$

Therefore $\|\dot{\tilde{X}}(t)\|^2 \leq \tilde{U}_1(t) \leq \lim_{\epsilon \rightarrow 0+} \tilde{U}_1(\epsilon)$ for $t > 0$, we get the desired result.

(ii) $p > 1$

As \tilde{X} is the solution for (12), from Lemma B.5 we know

$$\tilde{U}_2(t) = \frac{1}{t^{p-1}} \|\dot{\tilde{X}}(t)\|^2 + \frac{\gamma p}{t^{2p}} \|\tilde{X}(t) - X_0\|^2$$

is a nonincreasing function. However, as we don't know the value of $\lim_{t \rightarrow 0+} \frac{1}{t^{p-1}} \|\dot{\tilde{X}}(t)\|^2$, we first calculate it. To do so, we consider

$$\tilde{U}_1(t) = \|\dot{\tilde{X}}(t)\|^2 + \frac{\gamma p}{t^{p+1}} \|\tilde{X}(t) - X_0\|^2.$$

In the proof of Lemma B.5, we have observed its derivative becomes

$$\begin{aligned}\dot{\tilde{U}}_1(t) &= -2 \left\langle \dot{\tilde{X}}(t), \frac{d}{dt} \tilde{A}(\tilde{X}(t)) \right\rangle - \frac{2\gamma}{t^p} \left\| \dot{\tilde{X}}(t) - \frac{p}{t} (\tilde{X}(t) - X_0) \right\|^2 - \frac{\gamma p(1-p)}{t^{p+2}} \left\| \tilde{X}(t) - X_0 \right\|^2 \\ &\leq \frac{\gamma p(p-1)}{t^{p+2}} \left\| \tilde{X}(t) - X_0 \right\|^2.\end{aligned}$$

For $\epsilon > 0$, integrating from ϵ to t we have

$$\tilde{U}_1(t) \leq \tilde{U}_1(\epsilon) + \int_{\epsilon}^t \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X}(s) - X_0 \right\|^2 ds.$$

We consider taking limit $\epsilon \rightarrow 0+$. Observe from (17) and (15) we know $\lim_{t \rightarrow 0+} \frac{\tilde{X} - X_0}{t^p} = -\frac{\tilde{A}(X_0)}{\gamma}$ and $\left\| \dot{\tilde{X}}(0) \right\| = 0$, and therefore we have

$$\lim_{\epsilon \rightarrow 0+} \tilde{U}_1(\epsilon) = 0 + \lim_{\epsilon \rightarrow 0+} \frac{\gamma p}{\epsilon^{2p}} \left\| \tilde{X}(\epsilon) - X_0 \right\|^2 \cdot \epsilon^{p-1} = \frac{p}{\gamma} \left\| \tilde{A}(X_0) \right\|^2 \lim_{\epsilon \rightarrow 0+} \epsilon^{p-1} = 0.$$

Moreover as $\lim_{t \rightarrow 0+} \frac{\left\| \tilde{X}(t) - X_0 \right\|^2}{t^{2p}} = \frac{\left\| \tilde{A}(X_0) \right\|^2}{\gamma^2}$, there is $\delta > 0$ such that $0 < s < \delta$ implies $\frac{\left\| \tilde{X}(t) - X_0 \right\|^2}{s^{2p}} \leq \frac{2\left\| \tilde{A}(X_0) \right\|^2}{\gamma^2}$. Recalling $p > 1$, for $0 < \epsilon < \delta$ we have

$$\begin{aligned}\int_0^{\epsilon} \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X}(s) - X_0 \right\|^2 ds &\leq \int_0^{\epsilon} \frac{\gamma p(p-1)}{s^{2-p}} \frac{2\left\| \tilde{A}(X_0) \right\|^2}{\gamma^2} ds \\ &= \frac{2p(p-1)}{\gamma} \frac{\left\| \tilde{A}(X_0) \right\|^2}{\left[\frac{1}{p-1} s^{p-1} \right]_0^{\epsilon}} = \frac{2p}{\gamma} \frac{\left\| \tilde{A}(X_0) \right\|^2}{\epsilon^{p-1}},\end{aligned}$$

therefore

$$\int_0^t \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X}(s) - X_0 \right\|^2 ds \leq \frac{2p}{\gamma} \frac{\left\| \tilde{A}(X_0) \right\|^2}{\epsilon^{p-1}} + \int_{\epsilon}^t \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X}(s) - X_0 \right\|^2 ds < \infty.$$

Thus the integral is well-defined when $\epsilon \rightarrow 0+$. Taking limit $\epsilon \rightarrow 0+$ we have

$$\tilde{U}_1(t) \leq \int_0^t \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X}(s) - X_0 \right\|^2 ds.$$

Moving $\frac{\gamma p}{t^{p+1}} \left\| \tilde{X}(t) - X_0 \right\|^2$ to the right hand side, we get a inequality for $\left\| \dot{\tilde{X}}(t) \right\|^2$,

$$\left\| \dot{\tilde{X}}(t) \right\|^2 = \tilde{U}_1(t) - \frac{\gamma p}{t^{p+1}} \left\| \tilde{X}(t) - X_0 \right\|^2 \leq \int_0^t \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X}(s) - X_0 \right\|^2 ds - \frac{\gamma p}{t^{p+1}} \left\| \tilde{X}(t) - X_0 \right\|^2.$$

Observe, by L'Hôpital's rule we have

$$\lim_{t \rightarrow 0+} \frac{\int_0^t \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X}(s) - X_0 \right\|^2 ds}{t^{p-1}} = \lim_{t \rightarrow 0+} \frac{\frac{\gamma p(p-1)}{t^{p+2}} \left\| \tilde{X}(t) - X_0 \right\|^2}{(p-1)t^{p-2}} = \lim_{t \rightarrow 0+} \gamma p \frac{\left\| \tilde{X}(t) - X_0 \right\|^2}{t^{2p}} = \frac{p}{\gamma} \left\| \tilde{A}(X_0) \right\|^2.$$

Now dividing t^{p-1} to previous inequality and taking limit, we conclude

$$\begin{aligned}\lim_{t \rightarrow 0+} \frac{1}{t^{p-1}} \left\| \dot{\tilde{X}} \right\|^2 &\leq \lim_{t \rightarrow 0+} \frac{1}{t^{p-1}} \int_0^t \frac{\gamma p(p-1)}{s^{p+2}} \left\| \tilde{X} - X_0 \right\|^2 ds - \lim_{t \rightarrow 0+} \frac{\gamma p}{t^{2p}} \left\| \tilde{X} - X_0 \right\|^2 \\ &= \frac{p}{\gamma} \left\| \tilde{A}(X_0) \right\|^2 - \frac{p}{\gamma} \left\| \tilde{A}(X_0) \right\|^2 \\ &= 0.\end{aligned}$$

Thus we have $\lim_{t \rightarrow 0^+} \frac{1}{t^{p-1}} \|\dot{\tilde{X}}\|^2 = 0$. Therefore,

$$\lim_{\epsilon \rightarrow 0^+} \tilde{U}_2(\epsilon) = \lim_{\epsilon \rightarrow 0^+} \left(\frac{1}{\epsilon^{p-1}} \|\dot{\tilde{X}}(\epsilon)\|^2 + \frac{\gamma p}{\epsilon^{2p}} \|\tilde{X}(\epsilon) - X_0\|^2 \right) = \frac{p}{\gamma} \|\tilde{A}(X_0)\|^2.$$

From $\frac{\|\dot{\tilde{X}}\|^2}{t^{p-1}} \leq \lim_{\epsilon \rightarrow 0^+} \tilde{U}_2(\epsilon) = \frac{p}{\gamma} \|\tilde{A}(X_0)\|^2$, we conclude the desired result. □

Corollary B.9. Denote \tilde{X} as the solution of (12). Then for $T > 0$, following inequality is true for $t \in [0, T]$.

$$\|\tilde{A}(\tilde{X}(t))\| \leq \begin{cases} \sqrt{(\gamma+1) \left(1 + \frac{T^{1-p}}{p}\right)} \|\tilde{A}(X_0)\| & 0 < p < 1 \\ \|\tilde{A}(X_0)\| & p = 1 \\ \sqrt{\frac{p(\gamma+1)}{\gamma} \left(T^{p-1} + \frac{1}{p}\right)} \|\tilde{A}(X_0)\| & p > 1. \end{cases}$$

Proof. First observe,

$$\begin{aligned} \|\tilde{A}(\tilde{X}(t))\|^2 &= \left\| \dot{\tilde{X}}(t) + \frac{\gamma}{t^p} (\tilde{X}(t) - X_0) \right\|^2 \\ &= \|\dot{\tilde{X}}(t)\|^2 + 2\gamma \left\langle \dot{\tilde{X}}(t), \frac{\tilde{X}(t) - X_0}{t^p} \right\rangle + \frac{\gamma^2}{t^{2p}} \|\tilde{X}(t) - X_0\|^2 \\ &\leq \|\dot{\tilde{X}}(t)\|^2 + \gamma \left(\|\dot{\tilde{X}}(t)\|^2 + \frac{\|\tilde{X}(t) - X_0\|^2}{t^{2p}} \right) + \frac{\gamma^2}{t^{2p}} \|\tilde{X}(t) - X_0\|^2 \\ &= (\gamma+1) \left(\|\dot{\tilde{X}}(t)\|^2 + \frac{\gamma}{t^{2p}} \|\tilde{X}(t) - X_0\|^2 \right) \end{aligned}$$

The inequality comes from Young's inequality. Now we consider each case.

(i) $p = 1$

For this case, the terms on the right hand side exactly become $\tilde{U}_1(t)$. Therefore from (18)

$$\|\tilde{A}(\tilde{X}(t))\|^2 \leq (\gamma+1)\tilde{U}_1(t) \leq (\gamma+1) \lim_{\epsilon \rightarrow 0^+} \tilde{U}_1(\epsilon) = \|\tilde{A}(X_0)\|^2.$$

(ii) $0 < p < 1$

Recall from the proof of Corollary B.8 we know $\tilde{U}_1(t) \leq \lim_{\epsilon \rightarrow 0^+} \tilde{U}_1(\epsilon) = \|\tilde{A}(X_0)\|^2$. Therefore we have

$\frac{\gamma p}{t^{p+1}} \|\tilde{X}(t) - X_0\|^2 \leq \tilde{U}_1(t) \leq \|\tilde{A}(X_0)\|^2$, applying Corollary B.8 we get

$$\begin{aligned} \|\tilde{A}(\tilde{X}(t))\|^2 &\leq (\gamma+1) \left(\|\dot{\tilde{X}}(t)\|^2 + \frac{\gamma p}{t^{p+1}} \|\tilde{X}(t) - X_0\|^2 \cdot \frac{t^{1-p}}{p} \right) \\ &\leq (\gamma+1) \|\tilde{A}(X_0)\|^2 \left(1 + \frac{T^{1-p}}{p} \right). \end{aligned}$$

(iii) $p > 1$

Recall from the proof of Corollary B.8 we know $\tilde{U}_2(t) \leq \lim_{\epsilon \rightarrow 0^+} \tilde{U}_2(\epsilon) = \frac{p}{\gamma} \|\tilde{A}(X_0)\|^2$. Therefore we have

$\frac{\gamma p}{t^{2p}} \left\| \tilde{X}(t) - X_0 \right\|^2 \leq \tilde{U}_2(t) \leq \frac{p}{\gamma} \left\| \tilde{A}(X_0) \right\|^2$, applying Corollary B.8 we get

$$\begin{aligned} \left\| \tilde{A}(\tilde{X}(t)) \right\|^2 &\leq (\gamma + 1) \left(\left\| \dot{\tilde{X}}(t) \right\|^2 + \frac{\gamma p}{t^{2p}} \left\| \tilde{X}(t) - X_0 \right\|^2 \cdot \frac{1}{p} \right) \\ &\leq \frac{p(\gamma + 1)}{\gamma} \left(T^{p-1} + \frac{1}{p} \right) \left\| \tilde{A}(X_0) \right\|^2. \end{aligned}$$

□

B.2.2 Existence proof for general \mathbf{A}

Now we move on to the original inclusion. As noticed before, we will approximate \mathbf{A} with a Lipschitz function \mathbf{A}_λ called Yosida approximation. We define and state some facts about \mathbf{A}_λ as a lemma, and use it without proof. For the ones who are interested in proofs, see [11, Chapter 3.1, Theorem 2].

Lemma B.10. Define $\mathbf{A}_\lambda : \mathbb{R}^n \rightarrow \mathbb{R}^n$ as

$$\mathbf{A}_\lambda = \frac{1}{\lambda} (\mathbf{I} - \mathbf{J}_{\lambda \mathbf{A}}) = \frac{1}{\lambda} \left(\mathbf{I} - (\mathbf{I} + \lambda \mathbf{A})^{-1} \right)$$

This is so called Yosida approximation of \mathbf{A} . Followings are true.

- (i) \mathbf{A}_λ is $\frac{1}{\lambda}$ -Lipschitz continuous and maximal monotone.
- (ii) $\lim_{\lambda \rightarrow 0^+} \mathbf{A}_\lambda x = m(\mathbf{A}(x))$.
Here $m(\mathbf{A}(x))$ is defined as $m(\mathbf{A}(x)) = \Pi_{\mathbf{A}(x)}(0)$, the element of $\mathbf{A}(x)$ with minimal norm.
- (iii) $\|\mathbf{A}_\lambda(x)\| \leq \|m(\mathbf{A}(x))\|$.
- (iv) $\forall x \in \mathbb{R}^n, \mathbf{A}_\lambda(x) \in \mathbf{A}(\mathbf{J}_{\lambda \mathbf{A}}x)$.

Now we can state the proposition that proves existence of Theorem 2.2

Proposition B.11. For Yosida approximation \mathbf{A}_λ , consider the ODE

$$\dot{X}_\lambda(t) = -\mathbf{A}_\lambda(X_\lambda(t)) - \frac{\gamma}{t^p} (X_\lambda(t) - X_0) \quad (19)$$

with initial value condition $X_\lambda(0) = X_0 \in \text{dom}(\mathbf{A})$. The solution uniquely exists by Theorem B.2, denote the solution as X_λ . Now for $T > 0$ and a positive sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ such that $\lim_{n \rightarrow \infty} \lambda_n = 0$, define a sequence of solutions as $\mathcal{F}_T = \{X_{\lambda_n} : [0, T] \rightarrow \mathbb{R}^n \mid m \in \mathbb{N}\}$. Then there is a subsequence $\{\lambda_{n_k}\}_{k \in \mathbb{N}}$ such that $X_{\lambda_{n_k}}$ converges to the solution of (6) uniformly on $[0, T]$.

From Corollary B.8, Corollary B.9 and Lemma B.10 (iii), following lemma is immediate.

Lemma B.12. Let X_λ be the solution of (19). Then following is true for $t \in [0, T]$ for all $T > 0$.

$$\left\| \dot{X}_\lambda(t) \right\| \leq M_{dot}(T) = \begin{cases} \|m(\mathbf{A}(X_0))\| & 0 < p < 1 \\ \frac{1}{\sqrt{\gamma+1}} \|m(\mathbf{A}(X_0))\| & p = 1 \\ \sqrt{\frac{p}{\gamma}} T^{p-1} \|m(\mathbf{A}(X_0))\| & p > 1 \end{cases}$$

and

$$\|\mathbf{A}_\lambda(X_\lambda(t))\| \leq M_{\mathbf{A}}(T) = \begin{cases} \sqrt{(\gamma + 1) \left(1 + \frac{T^{1-p}}{p} \right)} \|m(\mathbf{A}(X_0))\| & 0 < p < 1 \\ \|m(\mathbf{A}(X_0))\| & p = 1 \\ \sqrt{\frac{p(\gamma+1)}{\gamma} \left(T^{p-1} + \frac{1}{p} \right)} \|m(\mathbf{A}(X_0))\| & p > 1. \end{cases}$$

Proof. Replace \tilde{A} with \mathbf{A}_λ and \tilde{X} with X_λ in Corollary B.8 and Corollary B.9. Applying $\|\mathbf{A}_\lambda(X_0)\| \leq \|m(\mathbf{A}(X_0))\|$ from Lemma B.10 (iii), we're done. □

While we're concluding the existence of converging sequence, we will exploit following lemma from [11, Chapter 0.3, Theorem 4]. For convenience, we restate the lemma here.

Lemma B.13. *Let us consider a sequence of absolutely continuous functions $x_k(\cdot)$ from an interval I of \mathbb{R} to a Banach space X satisfying*

- (i) $\forall t \in I, \{x_k(t)\}_{k \in \mathbb{N}}$ is a relatively compact subset of X
- (ii) *there exists a positive function $c(\cdot) \in L^1(I, [0, \infty))$ such that, for almost all $t \in I, \|\dot{x}_k(t)\| \leq c(t)$*

Then there exists a subsequence (again denoted by) $x_k(\cdot)$ converging to an absolutely continuous function $x(\cdot)$ from I to X in the sense that

- (i) $x_k(\cdot)$ converges to $x(\cdot)$ over compact subsets of I
- (ii) $\dot{x}_k(\cdot)$ converges weakly to $\dot{x}(\cdot)$ in $L^1(I, X)$

Proof. See [11, Chapter 0.3, Theorem 4]. □

From Lemma B.12 we can immediately check norm of all derivatives \dot{X}_{λ_m} are bounded by $M_{dot}(T)$. So condition (ii) holds with $M_{dot}(T)$. For condition (i), we prove \mathcal{F}_T is convergent in $C([0, T], \mathbb{R}^n)$.

Lemma B.14. \mathcal{F}_T is convergent sequence in $C([0, T], \mathbb{R}^n)$. In other words, $\forall \epsilon > 0$, there is $N > 0$ such that $n, m > N$ implies $\sup_{t \in [0, T]} \|X_{\lambda_n}(t) - X_{\lambda_m}(t)\| < \epsilon$

Proof. We will show X_{λ_n} is Cauchy sequence in $C([0, T], \mathbb{R}^n)$. Let $\nu, \lambda > 0$. From (19), we see for $t \in (0, T]$

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \|X_\nu(t) - X_\lambda(t)\|^2 &= \langle \dot{X}_\nu(t) - \dot{X}_\lambda(t), X_\nu(t) - X_\lambda(t) \rangle \\ &= -\langle \mathbf{A}_\nu(X_\nu(t)) - \mathbf{A}_\lambda(X_\lambda(t)), X_\nu(t) - X_\lambda(t) \rangle - \frac{\gamma}{t^p} \|X_\nu(t) - X_\lambda(t)\|^2 \\ &\leq -\langle \mathbf{A}_\nu(X_\nu(t)) - \mathbf{A}_\lambda(X_\lambda(t)), X_\nu(t) - X_\lambda(t) \rangle. \end{aligned}$$

From definition of resolvent we know $\mathbf{I} - \mathbf{J}_{\lambda\mathbf{A}} = \lambda\mathbf{A}_\lambda$. And from Lemma B.10 (iv) we know $\mathbf{A}_\lambda(X_\lambda(t)) \in \mathbf{A}(\mathbf{J}_{\lambda\mathbf{A}}(X_\lambda(t)))$. Thus from monotone inequality we see

$$\begin{aligned} &-\langle \mathbf{A}_\nu(X_\nu(t)) - \mathbf{A}_\lambda(X_\lambda(t)), X_\nu(t) - X_\lambda(t) \rangle \\ &= -\langle \mathbf{A}_\nu(X_\nu(t)) - \mathbf{A}_\lambda(X_\lambda(t)), \nu\mathbf{A}_\nu(X_\nu(t)) - \lambda\mathbf{A}_\lambda(X_\lambda(t)) \rangle \\ &\quad - \langle \mathbf{A}_\nu(X_\nu(t)) - \mathbf{A}_\lambda(X_\lambda(t)), \mathbf{J}_{\nu\mathbf{A}}(X_\nu(t)) - \mathbf{J}_{\lambda\mathbf{A}}(X_\lambda(t)) \rangle \\ &\leq -\langle \mathbf{A}_\nu(X_\nu(t)) - \mathbf{A}_\lambda(X_\lambda(t)), \nu\mathbf{A}_\nu X_\nu(t) - \lambda\mathbf{A}_\lambda(X_\lambda(t)) \rangle \\ &= (\nu + \lambda) \langle \mathbf{A}_\nu(X_\nu(t)), \mathbf{A}_\lambda(X_\lambda(t)) \rangle - \left(\nu \|\mathbf{A}_\nu X_\nu(t)\|^2 + \lambda \|\mathbf{A}_\lambda X_\lambda(t)\|^2 \right). \end{aligned}$$

By Young's inequality

$$\begin{aligned} &(\nu + \lambda) \langle \mathbf{A}_\nu(X_\nu(t)), \mathbf{A}_\lambda(X_\lambda(t)) \rangle - \left(\nu \|\mathbf{A}_\nu X_\nu(t)\|^2 + \lambda \|\mathbf{A}_\lambda X_\lambda(t)\|^2 \right) \\ &\leq \nu \left(\|\mathbf{A}_\nu(X_\nu(t))\|^2 + \frac{1}{4} \|\mathbf{A}_\lambda(X_\lambda(t))\|^2 \right) + \lambda \left(\|\mathbf{A}_\lambda(X_\lambda(t))\|^2 + \frac{1}{4} \|\mathbf{A}_\nu(X_\nu(t))\|^2 \right) \\ &\quad - \left(\nu \|\mathbf{A}_\nu X_\nu(t)\|^2 + \lambda \|\mathbf{A}_\lambda X_\lambda(t)\|^2 \right) \\ &= \frac{1}{4} \left(\nu \|\mathbf{A}_\lambda(X_\lambda(t))\|^2 + \lambda \|\mathbf{A}_\nu(X_\nu(t))\|^2 \right). \end{aligned}$$

Now applying Lemma B.12 we have

$$\frac{d}{dt} \frac{1}{2} \|X_\nu(t) - X_\lambda(t)\|^2 \leq \frac{1}{4} (\nu + \lambda) M_{\mathbf{A}}(T)^2.$$

Then integrating both sides from 0 to t we have

$$\|X_\nu(t) - X_\lambda(t)\|^2 \leq \frac{t}{2} (\nu + \lambda) M_{\mathbf{A}}(T)^2. \quad (20)$$

Now take $\epsilon > 0$. Then there is $N > 0$ such that for $n > N$, $\lambda_n < \frac{\epsilon}{TM_{\mathbf{A}}(T)^2}$ holds. Then for $t \in [0, T]$, $n, m > N$, we have

$$\|X_{\lambda_n}(t) - X_{\lambda_m}(t)\|^2 \leq \frac{t}{2}(\lambda_n + \lambda_m)M_{\mathbf{A}}(T)^2 < \epsilon.$$

Therefore we get the desired result. \square

Finally, we are ready to prove Proposition B.11, which implies the main theorem.

Proof of Proposition B.11. Take $T > 0$. We know X_{λ_n} uniformly converges on $[0, T]$ by Lemma B.14. Name the limit as X , i.e. define $X: [0, T] \rightarrow \mathbb{R}^n$ as $X(t) = \lim_{n \rightarrow \infty} X_{\lambda_n}(t)$. Then as $X_{\lambda_n}(0) = X_0$ for all $n \in \mathbb{N}$, we see X satisfies the initial condition. It remains to show X satisfies (6) almost everywhere.

Recall $\{\dot{X}_{\lambda_n}\}$ is bounded in $L^\infty([0, T], \mathbb{R}^n)$ by Lemma B.12. Thus we can apply Lemma B.13, there is a subsequence $\{\dot{X}_{\lambda_{n_k}}\}$ converges weakly to \dot{X} in $L^1([0, T], \mathbb{R}^n)$. Furthermore we have $\dot{X} \in L^\infty([0, T], \mathbb{R}^n)$ and so $\{\dot{X}_{\lambda_{n_k}}\}$ also converges weakly to \dot{X} in $L^2([0, T], \mathbb{R}^n)$ as well.

For $\lambda > 0$, define $f_\lambda: [0, T] \rightarrow \mathbb{R}^n$ as

$$f_\lambda(t) = \begin{cases} \frac{\gamma}{t^p}(X_\lambda(t) - X_0) & \text{if } t > 0 \\ 0 & \text{if } t = 0. \end{cases}$$

Then for $f: [0, T] \rightarrow \mathbb{R}^n$ defined as $f(t) = \frac{\gamma}{t^p}(X(t) - X_0)$ for $t > 0$ and $f(0) = 0$, we have $\lim_{k \rightarrow \infty} f_{\lambda_{n_k}}(t) = f(t)$. As $\|f_\lambda(t)\| = \|\dot{X}_\lambda(t) + \mathbf{A}_{\lambda_n}(X_\lambda)(t)\| \leq M_{\text{dot}}(T) + M_{\mathbf{A}}(T)$ for $t \in (0, T]$ by Lemma B.12, we have

$$\|f_\lambda(t) - f(t)\|^2 \leq (\|f_\lambda(t)\| + \|f(t)\|)^2 \leq 4(M_{\text{dot}}(T) + M_{\mathbf{A}}(T))^2.$$

Therefore by dominated convergence theorem we have

$$\lim_{k \rightarrow \infty} \int_0^T \|f_{\lambda_{n_k}}(t) - f(t)\|^2 dt = \int_0^T \lim_{k \rightarrow \infty} \|f_{\lambda_{n_k}}(t) - f(t)\|^2 dt = 0,$$

we conclude $f_{\lambda_{n_k}}$ strongly converges to f in $L^2([0, T], \mathbb{R}^n)$.

Now consider $F_\lambda: [0, T] \rightarrow \mathbb{R}^n$ defined as

$$F_\lambda(t) = \begin{cases} \dot{X}_\lambda(t) + \frac{\gamma}{t^p}(X_\lambda(t) - X_0) & \text{if } t > 0 \\ -\mathbf{A}_\lambda(X_0) & \text{if } t = 0 \end{cases}$$

Note since X_λ are solution to ODE (19), we have

$$F_\lambda(t) = -\mathbf{A}_\lambda(X_\lambda(t)).$$

Then for $F: [0, T] \rightarrow \mathbb{R}^n$ defined as

$$F(t) = \begin{cases} \dot{X}(t) + \frac{\gamma}{t^p}(X(t) - X_0) & \text{if } t > 0 \\ -m(\mathbf{A}(X_0)) & \text{if } t = 0, \end{cases}$$

we see $\{F_{\lambda_{n_k}}\}_{k \in \mathbb{N}}$ converges weakly to F in $L^2([0, T], \mathbb{R}^n)$.

On the other hand, by Lemma B.10 (iv) and the fact $-F_{\lambda_{n_k}}(t) = \mathbf{A}_{\lambda_{n_k}}(X_{\lambda_{n_k}}(t))$, we see

$$-F_{\lambda_{n_k}}(t) \in \mathbf{A}(\mathbf{J}_{\lambda_{n_k} \mathbf{A}}(X_{\lambda_{n_k}}(t))).$$

Observe, from the definition of \mathbf{A}_{λ_n} and Lemma B.12, we see

$$\|X_{\lambda_n}(t) - \mathbf{J}_{\lambda_n \mathbf{A}}(X_{\lambda_n}(t))\| = \lambda_n \|\mathbf{A}_{\lambda_n}(X_{\lambda_n}(t))\| \leq \lambda_n M_{\mathbf{A}}(T).$$

Since X_{λ_n} converges to X in $\mathcal{C}([0, T], \mathbb{R}^n)$, by taking $n \rightarrow \infty$ above inequality we see $\mathbf{J}_{\lambda_n \mathbf{A}}(X_{\lambda_n})$ also converges to X in $\mathcal{C}([0, T], \mathbb{R}^n)$.

Now for $\mathcal{A}: L^2([0, T], \mathbb{R}^n) \rightarrow L^2([0, T], \mathbb{R}^n)$ defined as $(\mathcal{A}(x))(t) = \mathbf{A}(x(t))$ almost everywhere, by [11, Chapter 3.1, Proposition 4], \mathcal{A} is maximal monotone since \mathbf{A} is maximal monotone. Since $-F_{\lambda_k}$ weakly converges to $-F$ in $L^2([0, T], \mathbb{R}^n)$

and $\mathbf{J}_{\lambda_k \mathbf{A}}(X_{\lambda_k})$ strongly converges to X in $L^2([0, T], \mathbb{R}^n)$, by [11, Chapter 3.1, Proposition 2] we have $-F \in \mathcal{A}(X)$ in $L^2([0, T], \mathbb{R}^n)$. Therefore for almost all $t \in (0, T]$ we have

$$-\left(\dot{X}(t) + \frac{\gamma}{t^p}(X(t) - X_0)\right) \in \mathbf{A}(X(t)).$$

Reorganizing the result with respect to \dot{X} , we have following is true for almost all $t \in (0, T]$

$$\dot{X}(t) \in -\mathbf{A}(X(t)) - \frac{\gamma}{t^p}(X(t) - X_0).$$

Since $T > 0$ was arbitrary, we conclude X satisfies above inclusion for almost all $t \in (0, \infty)$. \square

C Proof of existence and uniqueness of the solution of (10)

As uniqueness comes from Theorem B.1, we only need to show the existence. What we need for the existence proof are

- (i) Nonincreasing function $\tilde{U}(t)$ which contains $\|\dot{X}\|^2$ as in Lemma B.5.
- (ii) Uniform boundedness of $\dot{X}_\delta(t)$ for $t \in [0, T]$ as shown in Lemma B.6.
- (iii) $\dot{X}(0) = \lim_{t \rightarrow 0^+} \frac{\tilde{X}(t) - X_0}{t} = \lim_{t \rightarrow 0^+} \dot{\tilde{X}}(t)$ as shown in the existence proof of Lipschitz case.
- (iv) Uniform boundedness of $\|\dot{X}_\lambda(t)\|$ and $\|\mathbf{A}_\lambda(X_\lambda(t))\|$ for $t \in [0, T]$ as shown in Lemma B.12

We now show these steps can be also done to the $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1}$.

(i) Nonincreasing function $\tilde{U}(t)$ which contains $\|\dot{X}\|^2$

From

$$\dot{X}(t) = -\tilde{A}(\tilde{X}(t)) - \frac{2\mu}{e^{2\mu t} - 1}(\tilde{X}(t) - X_0)$$

for almost all t we have

$$\ddot{X}(t) = -\frac{d}{dt}\tilde{A}(\tilde{X}(t)) + \left(\frac{2\mu}{e^{2\mu t} - 1}\right)^2 e^{2\mu t}(\tilde{X}(t) - X_0) - \frac{2\mu}{e^{2\mu t} - 1}\dot{X}(t)$$

Define

$$\tilde{U}(t) = e^{-2\mu t} \|\dot{X}(t)\|^2 + \left(\frac{2\mu}{e^{2\mu t} - 1}\right)^2 \|\tilde{X}(t) - X_0\|^2$$

Therefore for almost all $t > 0$,

$$\begin{aligned} \dot{\tilde{U}}(t) &= 2e^{-2\mu t} \left\langle \dot{X}(t), \ddot{X}(t) \right\rangle - 2\mu e^{-2\mu t} \|\dot{X}(t)\|^2 \\ &\quad + 2 \left(\frac{2\mu}{e^{2\mu t} - 1}\right)^2 \left\langle \dot{X}(t), \tilde{X}(t) - X_0 \right\rangle - \left(\frac{2\mu}{e^{2\mu t} - 1}\right)^3 2e^{2\mu t} \|\tilde{X}(t) - X_0\|^2 \\ &= -2e^{-2\mu t} \left\langle \dot{X}(t), \frac{d}{dt}\tilde{A}(\tilde{X}(t)) \right\rangle + 2 \left(\frac{2\mu}{e^{2\mu t} - 1}\right)^2 \left\langle \dot{X}(t), \tilde{X}(t) - X_0 \right\rangle - 2e^{-2\mu t} \frac{2\mu}{e^{2\mu t} - 1} \|\dot{X}(t)\|^2 \\ &\quad - 2\mu e^{-2\mu t} \|\dot{X}(t)\|^2 + 2 \left(\frac{2\mu}{e^{2\mu t} - 1}\right)^2 \left\langle \dot{X}(t), \tilde{X}(t) - X_0 \right\rangle - \left(\frac{2\mu}{e^{2\mu t} - 1}\right)^3 2e^{2\mu t} \|\tilde{X}(t) - X_0\|^2 \\ &= -2\mu e^{-2\mu t} \|\dot{X}(t)\|^2 - 2e^{-2\mu t} \left\langle \dot{X}(t), \frac{d}{dt}\tilde{A}(\tilde{X}(t)) \right\rangle - \frac{2\mu e^{-2\mu t}}{e^{2\mu t} - 1} \left\| \dot{X}(t) - \frac{2\mu e^{2\mu t}}{e^{2\mu t} - 1} (\tilde{X}(t) - X_0) \right\|^2 \\ &\leq 0. \end{aligned}$$

(ii) Uniform boundedness of $\dot{\tilde{X}}_\delta(t)$

As (13), we define $\dot{\tilde{X}}_\delta(t)$ as the solution of

$$\dot{\tilde{X}}_\delta(t) = \begin{cases} -\tilde{A}(\tilde{X}_\delta(t)) - \frac{2\mu}{e^{2\mu\delta}-1}(\tilde{X}_\delta(t) - X_0) & 0 \leq t \leq \delta \\ -\tilde{A}(\tilde{X}_\delta(t)) - \frac{2\mu}{e^{2\mu t}-1}(\tilde{X}_\delta(t) - X_0) & t \geq \delta \end{cases}$$

Again with same arguments of Lemma B.6 we have $\|\tilde{X}_\delta(\delta) - X_0\| \leq \delta \|\tilde{A}(X_0)\|$ and $\|\dot{\tilde{X}}_\delta(\delta)\| \leq \|\tilde{A}(X_0)\|$. Now for $t \in [0, T]$

$$\begin{aligned} \|\dot{\tilde{X}}_\delta(t)\| &\leq \sqrt{e^{2\mu t} \tilde{U}(t)} \leq \sqrt{e^{2\mu t} \tilde{U}(\delta)} = e^{\mu t} \sqrt{e^{-2\mu\delta} \|\dot{\tilde{X}}_\delta(\delta)\|^2 + \left(\frac{2\mu}{e^{2\mu\delta}-1}\right)^2 \|\tilde{X}_\delta(\delta) - X_0\|^2} \\ &\leq e^{\mu t} \sqrt{e^{-2\mu\delta} \|\tilde{A}(X_0)\|^2 + \left(\frac{2\mu\delta}{e^{2\mu\delta}-1}\right)^2 \|\tilde{A}(X_0)\|^2} \\ &\leq e^{\mu T} \sqrt{2} \|\tilde{A}(X_0)\|. \end{aligned}$$

(iii) $\dot{\tilde{X}}(0) = \lim_{t \rightarrow 0+} \frac{\tilde{X}(t) - X_0}{t} = \lim_{t \rightarrow 0+} \dot{\tilde{X}}(t)$

Define $C(t) := 1 - e^{-2\mu t}$. Then $\dot{\tilde{X}}(t) = C(t)\beta(t)$ for $\beta(t) = \frac{2\mu}{e^{2\mu t}-1}$. And since

$$\lim_{t \rightarrow 0+} \frac{C(tv)}{C(t)} = \lim_{t \rightarrow 0+} \frac{1 - e^{-2\mu tv}}{1 - e^{-2\mu t}} = v,$$

with same argument done to arrive (15), we have

$$\dot{\tilde{X}}(0) = \lim_{t \rightarrow 0+} \frac{\tilde{X}(t) - X_0}{t} = - \int_0^1 \lim_{t \rightarrow 0+} \left(\frac{C(tv)}{C(t)} \tilde{A}(\tilde{X}(tv)) \right) dv = -\frac{1}{2} \tilde{A}(X_0)$$

Now from ODE $\dot{\tilde{X}}(t) = -\tilde{A}(\tilde{X}(t)) - \frac{2\mu}{e^{2\mu t}-1}(\tilde{X}(t) - X_0)$, by taking limit both sides by $t \rightarrow 0+$ we have

$$\lim_{t \rightarrow 0+} \dot{\tilde{X}}(t) = -\tilde{A}(\tilde{X}(0)) - \lim_{t \rightarrow 0+} \frac{2\mu t}{e^{2\mu t}-1} \frac{\tilde{X}(t) - X_0}{t} = -\frac{1}{2} \tilde{A}(X_0)$$

Therefore, $\lim_{t \rightarrow 0+} \frac{\tilde{X}(t) - X_0}{t} = \lim_{t \rightarrow 0+} \dot{\tilde{X}}(t)$.

(iv) Uniform boundedness of $\|\dot{X}_\lambda(t)\|$ and $\|\mathbf{A}_\lambda(t)\|$ for $t \in [0, T]$

Recall $U(t) = e^{-2\mu t} \left\| \dot{X}_\lambda(t) \right\|^2 + \left(\frac{2\mu}{e^{2\mu t}-1} \right)^2 \|X_\lambda(t) - X_0\|^2$ is nonincreasing. So from (iii), we have

$$\begin{aligned} e^{-2\mu t} \left\| \dot{X}_\lambda(t) \right\|^2 + \left(\frac{2\mu}{e^{2\mu t}-1} \right)^2 \|X_\lambda(t) - X_0\|^2 &\leq \lim_{t \rightarrow 0+} \left(e^{-2\mu t} \left\| \dot{X}_\lambda(t) \right\|^2 + \left(\frac{2\mu}{e^{2\mu t}-1} \right)^2 \|X_\lambda(t) - X_0\|^2 \right) \\ &= \frac{1}{4} \|\mathbf{A}_\lambda(X_0)\|^2 + \frac{1}{4} \|\mathbf{A}_\lambda(X_0)\|^2 = \frac{1}{2} \|\mathbf{A}_\lambda(X_0)\|^2 \end{aligned}$$

Therefore we have from Lemma B.10 (iii)

$$e^{-2\mu t} \left\| \dot{X}_\lambda(t) \right\|^2 \leq \frac{1}{2} \|\mathbf{A}_\lambda(X_0)\|^2 \leq \frac{1}{2} \|m(\mathbf{A}(X_0))\|^2 \implies \left\| \dot{X}_\lambda(t) \right\| \leq \frac{e^{\mu T}}{\sqrt{2}} \|m(\mathbf{A}(X_0))\|, \quad (21)$$

and by Young's inequality

$$\begin{aligned}
\|\mathbf{A}_\lambda(X_\lambda(t))\|^2 &\leq \left\| \dot{X}_\lambda(t) + \frac{2\mu}{e^{2\mu t} - 1} (X_\lambda(t) - X_0) \right\|^2 \\
&= 2 \left(\left\| \dot{X}_\lambda(t) \right\|^2 + \left(\frac{2\mu}{e^{2\mu t} - 1} \right)^2 \|X_\lambda(t) - X_0\|^2 \right) \\
&\leq 2 \left(\frac{e^{2\mu T}}{2} \|\mathbf{A}_\lambda(X_0)\|^2 + \frac{1}{2} \|\mathbf{A}_\lambda(X_0)\|^2 \right) = (e^{2\mu T} + 1) \|\mathbf{A}_\lambda(X_0)\|^2 \leq (e^{2\mu T} + 1) \|m(\mathbf{A}(X_0))\|^2.
\end{aligned}$$

Therefore

$$\|\mathbf{A}_\lambda(X_\lambda(t))\| \leq \sqrt{e^{2\mu T} + 1} \|m(\mathbf{A}(X_0))\|. \quad (22)$$

D Omitted proofs for derivation of anchor ODE (4)

D.1 Preparation for the proof of Theorem 2.1

We first provide the boundedness of trajectories as a lemma. As mentioned in the discussion after Lemma 2.3, boundedness of trajectories is an immediate corollary of Lemma 2.3. However, to address cases that are slightly more generalized, we present a proof using an argument similar to the one used in the proof of Lemma 2.3. Note the proof argument of following lemma is valid for the solution of differential equation (11) with satisfying the assumptions in Theorem 7.1 as well.

Lemma D.1. (*Boundedness of solutions*) Suppose $X(\cdot)$ is the solution of the differential inclusion (3). Then for all $X_\star \in \text{Zer}\mathbf{A}$, $t \in [0, \infty)$, following holds.

$$\|X(t) - X_\star\| \leq \|X_0 - X_\star\|.$$

And so, $\|X(t) - X_0\| \leq 2\|X_0 - X_\star\|$.

Proof. It is trivial for $t = 0$, so we may assume $t > 0$.

Take $X_\star \in \text{Zer}\mathbf{A}$. By monotonicity of \mathbf{A} and Young's inequality, we get the following inequality.

$$\begin{aligned}
\frac{d}{dt} \|X(t) - X_\star\|^2 &= 2 \left\langle \dot{X}(t), X(t) - X_\star \right\rangle = -2 \left\langle \tilde{\mathbf{A}}(X(t)) + \beta(t)(X(t) - X_0), X(t) - X_\star \right\rangle \\
&= -2 \left\langle \tilde{\mathbf{A}}(X(t)) + \beta(t)(X(t) - X_\star) - \beta(t)(X_0 - X_\star), X(t) - X_\star \right\rangle \\
&\leq -2\beta(t) \|X(t) - X_\star\|^2 + 2\beta(t) \langle X_0 - X_\star, X(t) - X_\star \rangle \\
&\leq -2\beta(t) \|X(t) - X_\star\|^2 + \beta(t) \left(\|X_0 - X_\star\|^2 + \|X(t) - X_\star\|^2 \right) \\
&= -\beta(t) \|X(t) - X_\star\|^2 + \beta(t) \|X_0 - X_\star\|^2.
\end{aligned}$$

Now again define $C(t) = e^{\int_0^t \beta(s) ds}$ for some $v > 0$, then we see $\dot{C}(t) = C(t)\beta(t)$. Moving $-\beta(t) \|X(t) - X_\star\|^2$ to the left hand side and multiplying both sides by $C(t)$, we have

$$\begin{aligned}
\frac{d}{dt} \left(C(t) \|X(t) - X_\star\|^2 \right) &= C(t) \frac{d}{dt} \|X(t) - X_\star\|^2 + C(t)\beta(t) \|X(t) - X_\star\|^2 \\
&\leq C(t)\beta(t) \|X_0 - X_\star\|^2 = \frac{d}{dt} C(t) \|X_0 - X_\star\|^2.
\end{aligned}$$

Integrating both sides from $\epsilon > 0$ to t we have

$$C(t) \|X(t) - X_\star\|^2 - C(\epsilon) \|X(\epsilon) - X_\star\|^2 \leq C(t) \|X_0 - X_\star\|^2 - C(\epsilon) \|X_0 - X_\star\|^2.$$

As $\beta > 0$, we have C is nonnegative and nondecreasing, $\lim_{\epsilon \rightarrow 0+} C(\epsilon)$ exists. Taking limit $\epsilon \rightarrow 0+$ both sides we have and dividing both sides by $C(t)$ we conclude

$$\|X(t) - X_\star\|^2 \leq \|X_0 - X_\star\|^2.$$

The latter statement holds directly from triangular inequality,

$$\|X(t) - X_0\| \leq \|X(t) - X_\star\| + \|X_\star - X_0\| \leq 2\|X_0 - X_\star\|.$$

□

Following lemma shows APPM is an instance of Halpern method. It is immediate from induction, but we state it as a lemma due as its importance.

Lemma D.2. *Consider a method defined as*

$$\begin{aligned}x^{k+1} &= \mathbf{J}_{\mathbf{A}}y^k \\y^{k+1} &= (1 - \beta_k)(2x^{k+1} - y^k) + \beta_kx^0,\end{aligned}$$

for $k = 0, 1, \dots$, with initial condition $y^0 = x^0$. Then for reflected resolvent $\mathbf{R}_{\mathbf{A}}$ defined as $\mathbf{R}_{\mathbf{A}} = 2\mathbf{J}_{\mathbf{A}} - \mathbf{I}$, above method is equivalent to

$$\tilde{y}^{k+1} = \beta_k\tilde{y}^0 + (1 - \beta_k)\mathbf{T}\tilde{y}^k$$

when $\mathbf{T} = \mathbf{R}_{\mathbf{A}}$, $\tilde{y}^0 = y^0$. Here equivalence means $\tilde{y}^k = y^k$ holds for $k = 0, 1, \dots$.

Proof. Proof by induction. As $\tilde{y}^0 = y^0$ by assumption, the statement is true for $k = 0$. Now suppose $y^k = \tilde{y}^k$ holds for $k \in \mathbb{N}$, then

$$\begin{aligned}\tilde{y}^{k+1} &= (1 - \beta_k)\mathbf{R}_{\mathbf{A}}\tilde{y}^k + \beta_0\tilde{y}^0 = (1 - \beta_k)(2\mathbf{J}_{\mathbf{A}} - \mathbf{I})\tilde{y}^k + \beta_0\tilde{y}^0 \\ &= (1 - \beta_k)(2\mathbf{J}_{\mathbf{A}} - \mathbf{I})y^k + \beta_0y^0 = (1 - \beta_k)(2x^{k+1} - y^k) + \beta_0y^0 = y^{k+1}.\end{aligned}$$

Thus $\tilde{y}^{k+1} = y^{k+1}$, the statement is true for $k + 1$. By induction, we get the desired result. \square

D.2 Proof of Theorem 2.1

Let $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a maximal monotone operator, and $h, \lambda, \delta > 0$. Again, denote $\mathbf{A}_\lambda = \frac{1}{\lambda}(\mathbf{I} - \mathbf{A}_{\lambda\mathbf{A}}) = \frac{1}{\lambda}(\mathbf{I} - (\mathbf{I} + \lambda\mathbf{A})^{-1})$. Since various kind of terms appear in the proof, we first organize the terms and notations.

- X : Solution of differential inclusion,

$$\dot{X} \in -\mathbf{A}(X) - \frac{1}{t}(X - X_0).$$

- X_λ : Solution of differential equation,

$$\dot{X}_\lambda = -\mathbf{A}_\lambda(X_\lambda) - \frac{1}{t}(X_\lambda - X_0).$$

- $X_{\lambda,\delta}$: Solution of approximated differential equation,

$$\dot{X}_{\lambda,\delta}(t) = \begin{cases} -\mathbf{A}_\lambda(X_{\lambda,\delta})(t) - \frac{1}{\delta}(X_{\lambda,\delta}(t) - X_0) & 0 \leq t < \delta \\ -\mathbf{A}_\lambda(X_{\lambda,\delta})(t) - \frac{1}{t}(X_{\lambda,\delta}(t) - X_0) & t \geq \delta. \end{cases} \quad (23)$$

- $X_{\lambda,\delta}^k$: Sequence obtained by taking Euler discretization of ODE (23),

$$X_{\lambda,\delta}^{k+1} = \begin{cases} X_\delta^k - \left(2h\mathbf{A}_\lambda(X_{\lambda,\delta}^k) + \frac{2h}{\delta}(X_{\lambda,\delta}^k - X_0)\right) & 0 \leq k < \frac{\delta}{2h} \\ X_\delta^k - \left(2h\mathbf{A}_\lambda(X_{\lambda,\delta}^k) + \frac{1}{k}(X_{\lambda,\delta}^k - X_0)\right) & k \geq \frac{\delta}{2h}. \end{cases}$$

- $x_{h,\lambda}^k$: Sequence obtained from APPM with operator $h\mathbf{A}_\lambda$, i.e.

$$\begin{aligned}x_{h,\lambda}^k &= \mathbf{J}_{h\mathbf{A}_\lambda}y_{h,\lambda}^{k-1} \\ y_{h,\lambda}^k &= \frac{k}{k+1}(2x_{h,\lambda}^k - y_{h,\lambda}^{k-1}) + \frac{1}{k+1}X_0.\end{aligned}$$

- x_h^k : Sequence obtained from APPM with operator $h\mathbf{A}$, i.e.

$$\begin{aligned}x_h^k &= \mathbf{J}_{h\mathbf{A}}y_h^{k-1} \\ y_h^k &= \frac{k}{k+1}(2x_h^k - y_h^{k-1}) + \frac{1}{k+1}X_0.\end{aligned}$$

We want to show for fixed $T > 0$,

$$\lim_{h \rightarrow 0^+} \sup_{0 \leq k < \frac{T}{2h}} \|x_h^k - X(2hk)\| = 0.$$

Equivalently we may show for fixed $T > 0$, for every $\{h_n\}_{n \in \mathbb{N}}$ such that $h_n > 0$ and converges to 0,

$$\lim_{n \rightarrow \infty} \sup_{0 \leq k < \frac{T}{2h_n}} \|x_{h_n}^k - X(2h_n k)\| = 0.$$

We will show this by considering inequality

$$\begin{aligned} \|x_h^k - X(2hk)\| &\leq \|X(2hk) - X_\lambda(2hk)\| + \|X_\lambda(2hk) - X_{\lambda,\delta}(2hk)\| \\ &\quad + \|X_{\lambda,\delta}(2hk) - X_{\lambda,\delta}^k\| + \|X_{\lambda,\delta}^k - x_h^k\| + \|x_{h,\lambda}^k - x_h^k\| =: S(h, \lambda, \delta, k). \end{aligned} \quad (24)$$

Our goal is to show, for every $\{h_n\}_{n \in \mathbb{N}}$ such that $h_n > 0$ and converges to 0, there is a sequence $\{(\delta_n, \lambda_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \rightarrow \infty} \sup_{0 \leq k < \frac{T}{2h_n}} S(h_n, \lambda_n, \delta_n, k) = 0$, and thus

$$\lim_{n \rightarrow \infty} \sup_{0 \leq k < \frac{T}{2h_n}} \|x_{h_n}^k - X(2h_n k)\| \leq \lim_{n \rightarrow \infty} \sup_{0 \leq k < \frac{T}{2h_n}} S(h_n, \lambda_n, \delta_n, k) = 0.$$

To clarify our goal, we need to find proper $\{(\delta_n, \lambda_n)\}_{n \in \mathbb{N}}$ in terms of $\{h_n\}_{n \in \mathbb{N}}$. As $\{h_n\}_{n \in \mathbb{N}}$ is determined, $\|x_{h_n}^k - X(2h_n k)\|$ is fixed and doesn't change by the choice of $\{(\delta_n, \lambda_n)\}_{n \in \mathbb{N}}$. But if we find $\{(\delta_n, \lambda_n)\}_{n \in \mathbb{N}}$ that makes $S(h_n, \lambda_n, \delta_n, k)$ small, since (24) holds for any choice of δ, λ , right choice of $\{(\delta_n, \lambda_n)\}_{n \in \mathbb{N}}$ can guarantee $\|x_{h_n}^k - X(2h_n k)\|$ is small. Thus to find such $\{(\delta_n, \lambda_n)\}_{n \in \mathbb{N}}$, we will observe each terms in S to find the required conditions.

Lemma D.3. For $h, \lambda, \delta > 0$ following is true.

$$(i) \|X(2hk) - X_\lambda(2hk)\| = \mathcal{O}(\sqrt{\lambda})$$

$$(ii) \|X_\lambda(2hk) - X_{\lambda,\delta}(2hk)\| = \mathcal{O}(\delta)$$

$$(iii) \|x_{h,\lambda}^k - x_h^k\| = \mathcal{O}(\lambda).$$

For $L_{\lambda,\delta} = \max\left\{\frac{1}{\lambda}, \frac{\sqrt{2}}{\lambda} \|m(\mathbf{A}(X_0))\|, \frac{1}{\delta}, \frac{4\sqrt{2}}{\delta} \|m(\mathbf{A}(X_0))\|\right\}$,

$$(iv) \|X_{\lambda,\delta}(2hk) - X_{\lambda,\delta}^k\| = \mathcal{O}(he^{2L_{\lambda,\delta}T}).$$

Further more if $0 < \frac{h}{\lambda} < \frac{1}{2}$,

$$\begin{aligned} (v) \|X_{\lambda,\delta}^k - x_h^k\| &= \mathcal{O}(h) + \mathcal{O}(h^2 L_{\lambda,\delta} e^{2L_{\lambda,\delta}T}) \\ &\quad + 3^{\frac{T}{\lambda}} \left(\mathcal{O}\left(\frac{h}{\lambda}\right) + \mathcal{O}\left(\frac{h}{\lambda} e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}(h) + \mathcal{O}\left(\frac{h^2}{\lambda} e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}\left(\frac{e^{2\delta L_{\lambda,\delta}}}{L_{\lambda,\delta}}\right) + \mathcal{O}(\delta) \right). \end{aligned}$$

We prove this lemma in next subsection, here we assume the lemma is true and prove Theorem 2.1. Suppose above lemma is true. The calculations are messy but the strategy is simple; balancing the speed of the terms h, δ, λ going zero to make above terms reach to zero. Above lemma motivate to take sequences as

$$\begin{aligned} \delta_n &= \min \left\{ \frac{h_n}{2}, \frac{8MT}{\log_3\left(\frac{1}{h_n}\right)} \right\} \\ \lambda_n &= \max \left\{ \delta_n, \frac{\sqrt{2} \|m(\mathbf{A}(X_0))\| \delta_n}{M}, \frac{2T}{\log_3\left(\frac{1}{\delta_n}\right)} \right\}. \end{aligned}$$

where $M = \max\{1, 4\sqrt{2} \|m(\mathbf{A}(X_0))\|\}$. When $\lim_{n \rightarrow \infty} h_n \rightarrow 0$ with $h_n > 0$, we can easily check $\lim_{n \rightarrow \infty} \delta_n = 0$ and $\lim_{n \rightarrow \infty} \lambda_n = 0$. So the cases (i), (ii), (iii) go to zero.

Now observe from the definition of λ_n we have,

$$\lambda_n \geq \max \left\{ \frac{\sqrt{2} \|m(\mathbf{A}(X_0))\| \delta_n}{M}, \delta_n \right\} \implies \frac{M}{\delta_n} \geq \max \left\{ \frac{\sqrt{2} \|m(\mathbf{A}(X_0))\|}{\lambda_n}, \frac{1}{\lambda_n} \right\} \implies L_{\lambda_n, \delta_n} = \frac{M}{\delta_n}.$$

Thus

$$\begin{aligned} e^{2L_{\lambda, \delta} T} &\leq e^{\frac{2MT}{\delta_n}} \leq 3^{\frac{2MT}{\delta_n}} \leq 3^{\frac{1}{4} \log_3(\frac{1}{h_n})} = \frac{1}{h_n^{1/4}} \\ 3^{\frac{T}{\lambda_n}} &\leq 3^{\frac{MT}{\delta_n}} \leq 3^{MT \frac{\log_3(\frac{1}{h_n})}{8MT}} = \frac{1}{h_n^{1/8}} \\ \frac{e^{2\delta_n L_{\lambda_n, \delta_n}}}{L_{\lambda_n, \delta_n}} &= e^{2M} \frac{\delta_n}{M} = \mathcal{O}(\delta_n). \end{aligned}$$

Therefore when $\lim_{n \rightarrow \infty} h_n \rightarrow 0$,

$$\begin{aligned} h_n e^{2L_{\lambda_n, \delta_n} T} &\leq h_n^{3/4} \rightarrow 0 \\ h_n^2 L_{\lambda_n, \delta_n} e^{2L_{\lambda_n, \delta_n} T} &= \frac{h_n^2}{2T} (2L_{\lambda_n, \delta_n} T) e^{2L_{\lambda_n, \delta_n} T} \leq \frac{h_n^2}{2T} e^{4L_{\lambda_n, \delta_n} T} \leq \frac{h_n^{3/2}}{2T} \rightarrow 0 \\ 3^{\frac{T}{\lambda_n}} h_n &= h_n^{7/8} \rightarrow 0 \\ 3^{\frac{T}{\lambda_n}} \frac{h_n}{\lambda_n} &= 3^{\frac{T}{\lambda_n}} \frac{T}{\lambda_n} \frac{h_n}{T} \leq 3^{\frac{2T}{\lambda_n}} \frac{h_n}{T} \leq \frac{h_n^{3/4}}{T} \rightarrow 0 \\ 3^{\frac{T}{\lambda_n}} \frac{h_n}{\lambda_n} e^{2L_{\lambda_n, \delta_n} T} &\leq \frac{h_n^{3/4}}{T} e^{2L_{\lambda_n, \delta_n} T} \leq \frac{h_n^{1/2}}{T} \rightarrow 0 \\ 3^{\frac{T}{\lambda_n}} \frac{h_n^2}{\lambda_n} e^{2L_{\lambda_n, \delta_n} T} &\leq \frac{h_n^{3/2}}{T} \rightarrow 0 \\ 3^{\frac{T}{\lambda_n}} \delta_n &\leq 3^{T \frac{\log_3(\delta_n)}{2T}} \delta_n = \delta_n^{1/2} \rightarrow 0. \end{aligned}$$

As $\lim_{n \rightarrow \infty} h_n = 0$, without loss of generality we may assume $h_n < 1$. Since $h_n > 0$ we have λ_n is well-defined and satisfies the condition for Lemma D.3. Thus terms for the case (iv) and (v) go to zero as well. Therefore we have $\lim_{n \rightarrow \infty} \sup_{0 \leq k < \frac{T}{2h_n}} S(h_n, \lambda_n, \delta_n, k) = 0$, as $\{h_n\}_{n \in \mathbb{N}}$ is arbitrary, we get the desired result.

D.2.1 Proof for case (i), (ii), (iii), (iv) of Lemma D.3

As case proof for (v) need lot of work, we provide it in a different subsection and here we provide the proofs for the cases from (i) to (iv).

(i) $\|X(2hk) - X_\lambda(2hk)\| = \mathcal{O}(\sqrt{\lambda})$

This is result of Lemma B.14. Considering (20) with $p = 1$, taking limit $\nu \rightarrow 0$ we know

$$\sup_{t \in [0, T]} \|X(t) - X_\lambda(t)\| \leq \sqrt{\frac{\lambda T}{2}} \|m(\mathbf{A}(X_0))\| = \mathcal{O}(\sqrt{\lambda}).$$

(ii) $\|X_\lambda(2hk) - X_{\lambda, \delta}(2hk)\| = \mathcal{O}(\delta)$

This is result of Proposition B.7. Consider (14) with $\gamma = p = 1$ for Lemma B.6 and taking limit $\nu \rightarrow 0$. Then applying Lemma B.10 (iii) we have

$$\sup_{t \in [0, T]} \|X_\lambda(t) - X_{\lambda, \delta}(t)\| \leq 2\sqrt{2}\delta \|\mathbf{A}_\lambda(X_0)\| \leq 2\sqrt{2}\delta \|m(\mathbf{A}(X_0))\| = \mathcal{O}(\delta).$$

$$(iii) \left\| x_{h,\lambda}^k - x_h^k \right\| = O(\lambda)$$

We first show a general fact about Yosida approximation and resolvent. From [15, Proposition 23.7 (iv)] we have

$$\begin{aligned} \|\mathbf{J}_{h\mathbf{A}}(x) - \mathbf{J}_{h\mathbf{A}_\lambda}(x)\| &= \left\| \mathbf{J}_{h\mathbf{A}}(x) - \left(\mathbf{I} + \frac{1}{1+\lambda} (\mathbf{J}_{(1+\lambda)h\mathbf{A}} - \mathbf{I}) \right) (x) \right\| \\ &= \left\| \mathbf{J}_{h\mathbf{A}}(x) - \left(\frac{\lambda}{1+\lambda} x + \frac{1}{1+\lambda} \mathbf{J}_{(1+\lambda)h\mathbf{A}}(x) \right) \right\| \\ &\leq \frac{1}{1+\lambda} \|\mathbf{J}_{h\mathbf{A}}(x) - \mathbf{J}_{(1+\lambda)h\mathbf{A}}(x)\| + \frac{\lambda}{1+\lambda} \|x - \mathbf{J}_{h\mathbf{A}}(x)\| \end{aligned}$$

From [15, Proposition 23.31 (iii)], we have

$$\|\mathbf{J}_{h\mathbf{A}}(x) - \mathbf{J}_{(1+\lambda)h\mathbf{A}}(x)\| \leq \lambda \|\mathbf{J}_{h\mathbf{A}}(x) - x\|.$$

Combining two facts we get

$$\|\mathbf{J}_{h\mathbf{A}}(x) - \mathbf{J}_{h\mathbf{A}_\lambda}(x)\| \leq \frac{2\lambda}{1+\lambda} \|x - \mathbf{J}_{h\mathbf{A}}(x)\|.$$

From Lemma D.2, we know the iteration of y^k sequence in (5) is equivalent to below sequence

$$y^{k+1} = \frac{1}{k+1} X_0 + \frac{k}{k+1} (2\mathbf{J}_{h\mathbf{A}} - \mathbf{I})(y^k).$$

Using this alternating form we have

$$\begin{aligned} \|y_h^{k+1} - y_{h,\lambda}^{k+1}\| &= \frac{k}{k+1} \|(2\mathbf{J}_{h\mathbf{A}} - \mathbf{I})(y_h^k) - (2\mathbf{J}_{h\mathbf{A}_\lambda} - \mathbf{I})(y_{h,\lambda}^k)\| \\ &\leq \frac{k}{k+1} (\|(2\mathbf{J}_{h\mathbf{A}} - \mathbf{I})(y_h^k) - (2\mathbf{J}_{h\mathbf{A}_\lambda} - \mathbf{I})(y_h^k)\| + \|\mathbf{R}_{h\mathbf{A}_\lambda}(y_h^k) - \mathbf{R}_{h\mathbf{A}_\lambda}(y_{h,\lambda}^k)\|) \\ &\leq \frac{k}{k+1} (2\|\mathbf{J}_{h\mathbf{A}}(y_h^k) - \mathbf{J}_{h\mathbf{A}_\lambda}(y_h^k)\| + \|y_h^k - y_{h,\lambda}^k\|) \\ &\leq \frac{k}{k+1} \left(\frac{4\lambda}{1+\lambda} \|y_h^k - \mathbf{J}_{h\mathbf{A}}(y_h^k)\| + \|y_h^k - y_{h,\lambda}^k\| \right) \\ &\leq \frac{k}{k+1} \left(\frac{4\lambda}{1+\lambda} \frac{\|X_0 - X_\star\|}{k} + \|y_h^k - y_{h,\lambda}^k\| \right). \end{aligned}$$

The first inequality comes from triangular inequality. The second inequality is from nonexpansiveness of reflected resolvent $\mathbf{R}_{\mathbf{A}} = 2\mathbf{J}_{\mathbf{A}} - \mathbf{I}$, [15, Corollary 23.11]. The third inequality is from the inequality shown previously. The last inequality comes from the convergence rate of APPM[35, Theorem 4.1], $\|y_h^k - \mathbf{J}_{h\mathbf{A}}(y_h^k)\| \leq \frac{\|X_0 - X_\star\|}{k}$.

Now multiplying both sides by $k+1$ and summing up from 0 to k we get

$$(k+1) \|y_h^{k+1} - y_{h,\lambda}^{k+1}\| \leq \sum_{i=0}^k \frac{4\lambda}{1+\lambda} \|X_0 - X_\star\| = (k+1) \frac{4\lambda}{1+\lambda} \|X_0 - X_\star\|.$$

Finally, from the relation between x^k and y^k in APPM we have

$$\begin{aligned} \|x_h^{k+1} - x_{h,\lambda}^{k+1}\| &= \|\mathbf{J}_{h\mathbf{A}}(y_h^{k+1}) - \mathbf{J}_{h\mathbf{A}_\lambda}(y_{h,\lambda}^{k+1})\| \\ &\leq \|\mathbf{J}_{h\mathbf{A}_\lambda}(y_h^{k+1}) - \mathbf{J}_{h\mathbf{A}_\lambda}(y_{h,\lambda}^{k+1})\| + \|\mathbf{J}_{h\mathbf{A}}(y_h^{k+1}) - \mathbf{J}_{h\mathbf{A}_\lambda}(y_h^{k+1})\| \\ &\leq \|y_h^{k+1} - y_{h,\lambda}^{k+1}\| + \frac{2\lambda}{1+\lambda} \|y_h^{k+1} - \mathbf{J}_{h\mathbf{A}}(y_h^{k+1})\| \\ &\leq \frac{4\lambda}{1+\lambda} \|X_0 - X_\star\| + \frac{2\lambda}{1+\lambda} \frac{\|X_0 - X_\star\|}{k+1} \\ &= \left(2 + \frac{1}{k+1} \right) \frac{2\lambda}{1+\lambda} \|X_0 - X_\star\| = O(\lambda). \end{aligned}$$

$$(iv) \left\| X_{\lambda,\delta}(2hk) - X_{\lambda,\delta}^k \right\| = \mathcal{O}(he^{2L\lambda,\delta T})$$

From (23), we can consider $\dot{X}_{\lambda,\delta}$ as of function $F: \mathbb{R}^n \times [0, \infty) \rightarrow \mathbb{R}^n$ defined as below

$$F(X, t) = \begin{cases} -\mathbf{A}_\lambda(X) - \frac{1}{\delta}(X - X_0) & 0 \leq t < \delta \\ -\mathbf{A}_\lambda(X) - \frac{1}{t}(X - X_0) & t > \delta. \end{cases} \quad (25)$$

Note F is $2 \max\{\frac{1}{\lambda}, \frac{1}{\delta}\}$ -Lipschitz with respect to X .

For convenience name $\alpha = 2h$. Define $\epsilon_k := X_{\lambda,\delta}(\alpha k) - X_{\lambda,\delta}^k$. By definition of Euler discretization and from fundamental theorem of calculus, we have the following

$$\begin{aligned} X_{\lambda,\delta}^{k+1} &= X_{\lambda,\delta}^k + \alpha F(X_{\lambda,\delta}^k, \alpha k) \\ X_{\lambda,\delta}(\alpha(k+1)) &= X_{\lambda,\delta}(\alpha k) + \int_{\alpha k}^{\alpha(k+1)} \dot{X}_{\lambda,\delta}(t) dt \\ &= X_{\lambda,\delta}(\alpha k) + \int_{\alpha k}^{\alpha(k+1)} \left(\dot{X}_{\lambda,\delta}(\alpha k) + \int_{\alpha k}^t \ddot{X}_{\lambda,\delta}(s) ds \right) dt \\ &= X_{\lambda,\delta}(\alpha k) + \alpha F(X_{\lambda,\delta}(\alpha k), \alpha k) + \int_{\alpha k}^{\alpha(k+1)} \int_{\alpha k}^t \ddot{X}_{\lambda,\delta}(s) ds dt. \end{aligned}$$

From Lemma B.4 we have $F(X_{\lambda,\delta}, t) = \dot{X}_{\lambda,\delta}(t)$ is Lipschitz continuous respect to t , so $\ddot{X}_{\lambda,\delta}$ is defined almost everywhere and fundamental theorem of calculus is valid. Therefore we have

$$\begin{aligned} \epsilon_{k+1} &= X_{\lambda,\delta}(\alpha(k+1)) - X_{\lambda,\delta}^{k+1} \\ &= X_{\lambda,\delta}(\alpha k) - X_{\lambda,\delta}^k + \alpha (F(X_{\lambda,\delta}(\alpha k), \alpha k) - F(X_{\lambda,\delta}^k, \alpha k)) + \int_{\alpha k}^{\alpha(k+1)} \int_{\alpha k}^t \ddot{X}_{\lambda,\delta}(s) ds dt \end{aligned}$$

As F is $2 \max\{\frac{1}{\lambda}, \frac{1}{\delta}\}$ -Lipschitz with respect to first variable, we have

$$\begin{aligned} \|\epsilon_{k+1}\| &\leq \|X_{\lambda,\delta}(\alpha k) - X_{\lambda,\delta}^k\| + \alpha \|F(X_{\lambda,\delta}(\alpha k), \alpha k) - F(X_{\lambda,\delta}^k, \alpha k)\| + \int_{\alpha k}^{\alpha(k+1)} \int_{\alpha k}^t \|\ddot{X}_{\lambda,\delta}(s)\| ds dt \\ &\leq \left(1 + 2\alpha \max\left\{\frac{1}{\lambda}, \frac{1}{\delta}\right\}\right) \|\epsilon_k\| + \int_{\alpha k}^{\alpha(k+1)} \int_{\alpha k}^t \|\ddot{X}_{\lambda,\delta}(s)\| ds dt. \end{aligned}$$

Now we observe $\|\ddot{X}_{\lambda,\delta}\|$ is bounded. By differentiating $\dot{X}_{\lambda,\delta}$, as

$$\ddot{X}_{\lambda,\delta} = -\frac{d}{dt} \mathbf{A}_\lambda(X_{\lambda,\delta}) + \frac{1}{t^2} (\dot{X}_{\lambda,\delta} - X_0) - \frac{1}{t} \dot{X}_{\lambda,\delta} = -\frac{d}{dt} \mathbf{A}_\lambda(X_{\lambda,\delta}) + \frac{1}{t} (-\mathbf{A}_\lambda(X_{\lambda,\delta}) - \dot{X}_{\lambda,\delta}) - \frac{1}{t} \dot{X}_{\lambda,\delta}$$

for $t > \delta$, we have for almost every t

$$\ddot{X}_{\lambda,\delta} = \begin{cases} -\frac{d}{dt} \mathbf{A}_\lambda(X_{\lambda,\delta}) + \frac{1}{\delta} \dot{X}_{\lambda,\delta} & 0 \leq t < \delta \\ -\frac{d}{dt} \mathbf{A}_\lambda(X_{\lambda,\delta}) - \frac{2}{t} \dot{X}_{\lambda,\delta} & t \geq \delta \end{cases}.$$

Considering $\gamma = 1, p = 1$ to Lemma B.6 and from Lemma B.10 (iv) we have

$$\|\dot{X}_{\lambda,\delta}(t)\| \leq \sqrt{2} \|\mathbf{A}_\lambda(X_0)\| \leq \sqrt{2} \|m(\mathbf{A}(X_0))\|,$$

and thus

$$\begin{aligned} \|\mathbf{A}_\lambda(X_{\lambda,\delta}(t))\| &= \left\| \dot{X}_{\lambda,\delta}(t) + \frac{1}{\max\{\delta, t\}} (X_{\lambda,\delta}(t) - X_0) \right\| \\ &\leq \|\dot{X}_{\lambda,\delta}(t)\| + \frac{1}{\max\{\delta, t\}} \|X_{\lambda,\delta}(t) - X_0\| \\ &\leq \|\dot{X}_{\lambda,\delta}(t)\| + \frac{1}{\max\{\delta, t\}} \int_0^t \|\dot{X}_{\lambda,\delta}(s)\| ds \leq 2\sqrt{2} \|m(\mathbf{A}(X_0))\|. \end{aligned}$$

And since \mathbf{A}_λ is $\frac{1}{\lambda}$ -Lipschitz, we know $\mathbf{A}_\lambda \circ X_{\lambda,\delta}$ is $\frac{1}{\lambda}\sqrt{2}\|m(\mathbf{A}(X_0))\|$ -Lipschitz, thus we have for almost all t ,

$$\left\| \frac{d}{dt} \mathbf{A}_\lambda(X_{\lambda,\delta}(t)) \right\| \leq \frac{\sqrt{2}\|m(\mathbf{A}(X_0))\|}{\lambda}.$$

Applying these facts we have

$$\begin{aligned} \left\| \ddot{X}_{\lambda,\delta}(t) \right\| &\leq \left\| \frac{d}{dt} \mathbf{A}_\lambda(X_{\lambda,\delta}(t)) \right\| + \frac{1}{\delta} \|\mathbf{A}_\lambda(X_{\lambda,\delta}(t))\| + \frac{2}{\delta} \|\dot{X}_{\lambda,\delta}(t)\| \\ &\leq \frac{\sqrt{2}\|m(\mathbf{A}(X_0))\|}{\lambda} + \frac{4\sqrt{2}}{\delta} \|m(\mathbf{A}(X_0))\| \leq 2 \max \left\{ \frac{\sqrt{2}\|m(\mathbf{A}(X_0))\|}{\lambda}, \frac{4\sqrt{2}}{\delta} \|m(\mathbf{A}(X_0))\| \right\}. \end{aligned}$$

Therefore

$$\|\epsilon_{k+1}\| \leq \left(1 + 2\alpha \max \left\{ \frac{1}{\lambda}, \frac{1}{\delta} \right\} \right) \|\epsilon_k\| + 2 \max \left\{ \frac{\sqrt{2}\|m(\mathbf{A}(X_0))\|}{\lambda}, \frac{4\sqrt{2}}{\delta} \|m(\mathbf{A}(X_0))\| \right\} \alpha^2 \quad (26)$$

Now for $L_{\delta,\lambda} = \max \left\{ \frac{1}{\lambda}, \frac{\sqrt{2}\|m(\mathbf{A}(X_0))\|}{\lambda}, \frac{1}{\delta}, \frac{4\sqrt{2}}{\delta} \|m(\mathbf{A}(X_0))\| \right\}$, we show

$$\|\epsilon_k\| \leq h e^{2L_{\lambda,\delta}T}.$$

Multiplying $(1 + 2\alpha L_{\lambda,\delta})^{-(k+1)}$ to (26) we have

$$(1 + 2\alpha L_{\lambda,\delta})^{-(k+1)} \|\epsilon_{k+1}\| \leq (1 + 2\alpha L_{\lambda,\delta})^{-k} \|\epsilon_k\| + (1 + 2\alpha L_{\lambda,\delta})^{-(k+1)} L_{\lambda,\delta} \alpha^2$$

As $\|\epsilon_0\| = \|X_0 - X(0)\| = 0$, summing up from 0 to $k - 1$ we have

$$\begin{aligned} (1 + 2\alpha L_{\lambda,\delta})^{-k} \|\epsilon_k\| &\leq \sum_{i=1}^k (1 + 2\alpha L_{\lambda,\delta})^{-i} L_{\lambda,\delta} \alpha^2 \\ &= \frac{(1 + 2\alpha L_{\lambda,\delta})^{-1} \left(1 - (1 + 2\alpha L_{\lambda,\delta})^{-k} \right)}{1 - (1 + 2\alpha L_{\lambda,\delta})^{-1}} L_{\lambda,\delta} \alpha^2 \\ &= \frac{1}{2} \left(1 - (1 + 2\alpha L_{\lambda,\delta})^{-k} \right) \alpha. \end{aligned}$$

Multiplying $(1 + 2\alpha L_{\lambda,\delta})^k$ to both sides and applying $\alpha = 2h$ we have

$$\|\epsilon_k\| \leq \frac{\alpha}{2} \left((1 + 2\alpha L_{\lambda,\delta})^k - 1 \right) = h \left((1 + 4hL_{\lambda,\delta})^k - 1 \right). \quad (27)$$

Now from

$$(1 + 4hL_{\lambda,\delta})^k \leq \left((1 + 4hL_{\lambda,\delta})^{\frac{1}{4hL_{\lambda,\delta}}} \right)^{4hL_{\lambda,\delta}k} \leq e^{4hL_{\lambda,\delta}k},$$

applying $k \leq \frac{T}{2h}$

$$\|\epsilon_k\| \leq h \left(e^{4hL_{\lambda,\delta}k} - 1 \right) \leq h e^{2L_{\lambda,\delta}T}.$$

Therefore

$$\|X_{\lambda,\delta}(2hk) - X_{\lambda,\delta}^k\| \leq h e^{2L_{\lambda,\delta}T} = \mathcal{O}(h e^{2L_{\lambda,\delta}T}). \quad (28)$$

D.3 Proof for case (v) of Lemma D.3

As APPM has coefficient $\frac{1}{k+1}$, we consider $X_{\lambda,\delta}^{k+1}$ instead of $X_{\lambda,\delta}^k$ due to calculation simplicity. From triangular inequality, we have

$$\|X_{\lambda,\delta}^k - x_\lambda^k\| \leq \|X_{\lambda,\delta}^k - X_{\lambda,\delta}^{k+1}\| + \|X_{\lambda,\delta}^{k+1} - x_\lambda^k\|.$$

We will show

$$\begin{aligned}\|X_{\lambda,\delta}^k - X_{\lambda,\delta}^{k+1}\| &= \mathcal{O}(h) + \mathcal{O}(h^2 L_{\lambda,\delta} e^{2L_{\lambda,\delta} T}) \\ \|X_{\lambda,\delta}^{k+1} - x_{\lambda}^k\| &= 3^{\frac{T}{\delta}} \left(\mathcal{O}\left(\frac{h}{\lambda}\right) + \mathcal{O}\left(\frac{h}{\lambda} e^{2L_{\lambda,\delta} T}\right) + \mathcal{O}(h) + \mathcal{O}\left(\frac{h^2}{\lambda} e^{2L_{\lambda,\delta} T}\right) + \mathcal{O}\left(\frac{e^{2\delta L_{\lambda,\delta}}}{L_{\lambda,\delta}}\right) + \mathcal{O}(\delta) \right).\end{aligned}$$

First one is simple. Since $X_{\lambda,\delta}^{k+1} = X_{\lambda,\delta}^k + 2hF(X_{\lambda,\delta}^k, 2hk)$ and F is $2 \max\{\frac{1}{\lambda}, \frac{1}{\delta}\}$ -Lipschitz with respect to the first variable, we have

$$\begin{aligned}\|X_{\lambda,\delta}^k - X_{\lambda,\delta}^{k+1}\| &= 2h \|F(X_{\lambda,\delta}^k, 2hk)\| \\ &\leq 2h \|F(X_{\lambda,\delta}(2hk), 2hk)\| + 2h \|F(X_{\lambda,\delta}(2hk), 2hk) - F(X_{\lambda,\delta}^k, 2hk)\| \\ &\leq 2h \|\dot{X}_{\lambda,\delta}(2hk)\| + 4h \max\left\{\frac{1}{\lambda}, \frac{1}{\delta}\right\} \|X_{\lambda,\delta}(2hk) - X_{\lambda,\delta}^k\| \\ &\leq 2\sqrt{2}h \|m(\mathbf{A}_{\lambda}(X_0))\| + 4h^2 L_{\lambda,\delta} e^{2L_{\lambda,\delta} T} \\ &= \mathcal{O}(h) + \mathcal{O}(h^2 L_{\lambda,\delta} e^{2L_{\lambda,\delta} T}).\end{aligned}$$

Second one is complicated, we present our proof with dividing steps to subsections.

D.3.1 Recursive inequality for $\epsilon_k = \|X_{\lambda,\delta}^{k+1} - x_{h,\lambda}^k\|$

Define $\epsilon_k = \|X_{\lambda,\delta}^{k+1} - x_{h,\lambda}^k\|$. Recall, $X_{\lambda,\delta}$ was solution of approximated ODE

$$\dot{X}_{\lambda,\delta}(t) = F(X_{\lambda,\delta}, t) = \begin{cases} -\mathbf{A}_{\lambda}(X_{\lambda,\delta})(t) - \frac{1}{\delta}(X(t) - X_0) & 0 \leq t < \delta \\ -\mathbf{A}_{\lambda}(X_{\lambda,\delta})(t) - \frac{1}{t}(X(t) - X_0) & t \geq \delta \end{cases}.$$

We now wish to write ϵ_{k+1} in terms of ϵ_k . As ϵ_{k+1} involves $X_{\lambda,\delta}^{k+2}$, we first write it explicitly.

$$\begin{aligned}X_{\lambda,\delta}^{k+2} &= X_{\lambda,\delta}^{k+1} + 2hF\left(X_{\lambda,\delta}^{k+1}, 2h(k+1)\right) \\ &= \begin{cases} X_{\lambda,\delta}^{k+1} - \left(2h\mathbf{A}_{\lambda}(X_{\lambda,\delta}^{k+1}) + \frac{2h}{\delta}(X_{\lambda,\delta}^{k+1} - X_0)\right) & 0 \leq k+1 < \frac{\delta}{2h} \\ X_{\lambda,\delta}^{k+1} - \left(2h\mathbf{A}_{\lambda}(X_{\lambda,\delta}^{k+1}) + \frac{1}{k+1}(X_{\lambda,\delta}^{k+1} - X_0)\right) & k+1 \geq \frac{\delta}{2h}. \end{cases}\end{aligned}$$

Now we find recursive inequality considering two cases.

- (i) $k+1 \geq \frac{\delta}{2h}$
Recall APPM (5) was defined as

$$\begin{aligned}x_{h,\lambda}^k &= \mathbf{J}_{h\mathbf{A}_{\lambda}} y_{h,\lambda}^{k-1} \\ y_{h,\lambda}^k &= \frac{k}{k+1}(2x_{h,\lambda}^k - y_{h,\lambda}^{k-1}) + \frac{1}{k+1}X_0,\end{aligned}$$

and substituting $y_{h,\lambda}^{k-1} = x_{h,\lambda}^k + h\mathbf{A}_{\lambda}(x_{h,\lambda}^k)$, we get a one line expression

$$x_{h,\lambda}^{k+1} = \frac{k}{k+1}x_{h,\lambda}^k - h \left(\mathbf{A}_{\lambda}(x_{h,\lambda}^{k+1}) + \frac{k}{k+1}\mathbf{A}_{\lambda}(x_{h,\lambda}^k) \right) + \frac{1}{k+1}X_0.$$

Rewriting $X_{\lambda,\delta}^{k+2}$ to make easier to compare with above,

$$\begin{aligned}X_{\lambda,\delta}^{k+2} &= \frac{k}{k+1}X_{\lambda,\delta}^{k+1} - h \left(\mathbf{A}_{\lambda}(X_{\lambda,\delta}^{k+2}) + \frac{k}{k+1}\mathbf{A}_{\lambda}(X_{\lambda,\delta}^{k+1}) \right) + \frac{1}{k+1}X_0 \\ &\quad + h \left(\mathbf{A}_{\lambda}(X_{\lambda,\delta}^{k+2}) - \mathbf{A}_{\lambda}(X_{\lambda,\delta}^{k+1}) \right) - \frac{h}{k+1}\mathbf{A}_{\lambda}(X_{\lambda,\delta}^{k+1}).\end{aligned}$$

As \mathbf{A}_λ is $\frac{1}{\lambda}$ -Lipschitz

$$\epsilon_{k+1} \leq \frac{k}{k+1} \epsilon_k + \frac{h}{\lambda} \left(\epsilon_{k+1} + \frac{k}{k+1} \epsilon_k \right) + \frac{h}{\lambda} \left\| X_{\lambda,\delta}^{k+2} - X_{\lambda,\delta}^{k+1} \right\| + \frac{h}{k+1} \left\| \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) \right\|.$$

Organizing,

$$\epsilon_{k+1} \leq \frac{1}{1 - \frac{h}{\lambda}} \left(\frac{k}{k+1} \left(1 + \frac{h}{\lambda} \right) \epsilon_k + h \underbrace{\left(\frac{1}{\lambda} \left\| X_{\lambda,\delta}^{k+2} - X_{\lambda,\delta}^{k+1} \right\| + \frac{1}{k+1} \left\| \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) \right\| \right)}_{:=e_k} \right).$$

(ii) $k+1 < \frac{\delta}{2h}$

Rewriting $X_{\lambda,\delta}^{k+2}$ we have

$$\begin{aligned} X_{\lambda,\delta}^{k+2} &= \frac{k}{k+1} X_{\lambda,\delta}^{k+1} - h \left(\mathbf{A}_\lambda(X_{\lambda,\delta}^{k+2}) + \frac{k}{k+1} \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) \right) + \frac{1}{k+1} X_0 \\ &\quad + h \left(\mathbf{A}_\lambda(X_{\lambda,\delta}^{k+2}) - \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) \right) + \frac{h}{k+1} \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) + \left(\frac{1}{k+1} - \frac{2h}{\delta} \right) (X_{\lambda,\delta}^{k+1} - X_0). \end{aligned}$$

The only difference between the case (i) is the last term. Therefore with same calculation, we get below inequality.

$$\epsilon_{k+1} \leq \frac{1}{1 - \frac{h}{\lambda}} \left(\frac{k}{k+1} \left(1 + \frac{h}{\lambda} \right) \epsilon_k + h \underbrace{\left(\frac{1}{\lambda} \left\| X_{\lambda,\delta}^{k+2} - X_{\lambda,\delta}^{k+1} \right\| + \frac{1}{k+1} \left\| \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) \right\| \right)}_{:=e_k} + \left| \frac{1}{k+1} - \frac{2h}{\delta} \right| \left\| X_{\lambda,\delta}^{k+1} - X_0 \right\| \right).$$

From case (i) and (ii), by defining

$$e_k = \begin{cases} h \left(\frac{1}{\lambda} \left\| X_{\lambda,\delta}^{k+2} - X_{\lambda,\delta}^{k+1} \right\| + \frac{1}{k+1} \left\| \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) \right\| \right) + \left| \frac{1}{k+1} - \frac{2h}{\delta} \right| \left\| X_{\lambda,\delta}^{k+1} - X_0 \right\| & 0 \leq k+1 < \frac{\delta}{2h} \\ h \left(\frac{1}{\lambda} \left\| X_{\lambda,\delta}^{k+2} - X_{\lambda,\delta}^{k+1} \right\| + \frac{1}{k+1} \left\| \mathbf{A}_\lambda(X_{\lambda,\delta}^{k+1}) \right\| \right) & k+1 \geq \frac{\delta}{2h}, \end{cases}$$

we can write a recursive inequality for all k as below

$$\epsilon_{k+1} \leq \frac{1}{1 - \frac{h}{\lambda}} \left(\frac{k}{k+1} \left(1 + \frac{h}{\lambda} \right) \epsilon_k + e_k \right).$$

$$\mathbf{D.3.2} \quad \sup_{0 \leq i \leq \frac{T}{2h}} \epsilon_i = 3 \frac{T}{\lambda} \left(\mathcal{O}\left(\frac{h}{\lambda}\right) + \mathcal{O}\left(\frac{h}{\lambda} e^{2L\lambda,\delta T}\right) + \mathcal{O}(h) + \mathcal{O}\left(\frac{h^2}{\lambda} e^{2L\lambda,\delta T}\right) + \mathcal{O}\left(\frac{e^{2\delta L\lambda,\delta}}{L\lambda,\delta}\right) + \mathcal{O}(\delta) \right)$$

We will now sum up above inequality. Multiplying $(k+1) \left(\frac{1 - \frac{h}{\lambda}}{1 + \frac{h}{\lambda}} \right)^{k+1}$ both sides we have

$$\left(\frac{1 - \frac{h}{\lambda}}{1 + \frac{h}{\lambda}} \right)^{k+1} (k+1) \epsilon_{k+1} \leq \left(\frac{1 - \frac{h}{\lambda}}{1 + \frac{h}{\lambda}} \right)^k k \epsilon_k + \frac{\left(1 - \frac{h}{\lambda} \right)^{k+1}}{\left(1 + \frac{h}{\lambda} \right)^k} (k+1) e_k.$$

By summing up above inequality from 0 to $k-1$, we have

$$\left(\frac{1 - \frac{h}{\lambda}}{1 + \frac{h}{\lambda}} \right)^k k \epsilon_k \leq \sum_{i=0}^{k-1} \frac{\left(1 - \frac{h}{\lambda} \right)^{i+1}}{\left(1 + \frac{h}{\lambda} \right)^i} (i+1) e_i.$$

Note ϵ_0 vanished as it is multiplied with 0. Reorganizing with respect to ϵ_k we have

$$\begin{aligned} \epsilon_k &\leq \left(\frac{1 + \frac{h}{\lambda}}{1 - \frac{h}{\lambda}} \right)^k \frac{1}{k} \sum_{i=0}^{k-1} \frac{\left(1 - \frac{h}{\lambda} \right)^{i+1}}{\left(1 + \frac{h}{\lambda} \right)^i} (i+1) e_i \\ &\leq \left(\frac{1 + \frac{h}{\lambda}}{1 - \frac{h}{\lambda}} \right)^{\frac{T}{2h}} \frac{1}{k} \sum_{i=0}^{k-1} (i+1) e_i = \left(\left(\frac{1 + \frac{h}{\lambda}}{1 - \frac{h}{\lambda}} \right)^{\frac{\lambda}{h}} \right)^{\frac{T}{2\lambda}} \frac{1}{k} \sum_{i=0}^{k-1} (i+1) e_i \end{aligned}$$

For second inequality follows from the fact $0 < \frac{h}{\lambda} < \frac{1}{2}$, which implies $0 < \frac{(1-\frac{h}{\lambda})^{i+1}}{(1+\frac{h}{\lambda})^i} \leq 1$ and $1 < \frac{1+\frac{h}{\lambda}}{1-\frac{h}{\lambda}}$. Observe

$f(x) = \left(\frac{1+x}{1-x}\right)^{\frac{1}{x}}$ is nondecreasing in $x \in (0, 1)$ since

$$f'(x) = -\frac{\left(\frac{1+x}{1-x}\right)^{\frac{1}{x}} \left((x^2-1) \log\left(\frac{1+x}{1-x}\right) + 2x\right)}{x^2(x^2-1)}.$$

Therefore, from $f\left(\frac{1}{2}\right) = 9$ we have

$$\epsilon_k \leq f\left(\frac{1}{2}\right)^{\frac{T}{2\lambda}} \frac{1}{k} \sum_{i=0}^{k-1} (i+1)e_i = 3^{\frac{T}{\lambda}} \frac{1}{k} \sum_{i=0}^{k-1} (i+1)e_i.$$

Now we show

$$\frac{1}{k} \sum_{i=0}^{k-1} (i+1)e_i = \mathcal{O}\left(\frac{h}{\lambda}\right) + \mathcal{O}\left(\frac{h}{\lambda} e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}(h) + \mathcal{O}\left(\frac{h^2}{\lambda} e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}(\delta) + \mathcal{O}\left(\frac{e^{2\delta L_{\lambda,\delta}}}{L_{\lambda,\delta}}\right)$$

for $0 < T < \infty$, $0 \leq k < \frac{T}{2h}$. Name $N_k = \min\{k, \lfloor \frac{\delta}{2h} \rfloor\}$. From the definition of e_i we have

$$\begin{aligned} & \frac{1}{k} \sum_{i=0}^{k-1} (i+1)e_i \\ &= \frac{1}{k} \sum_{i=0}^{k-1} h(i+1) \left(\frac{1}{\lambda} \|X_{\lambda,\delta}^{i+2} - X_{\lambda,\delta}^{i+1}\| + \frac{1}{i+1} \|\mathbf{A}_\lambda(X_{\lambda,\delta}^{i+1})\| \right) + \frac{1}{k} \sum_{i=0}^{N_k-1} (i+1) \left| \frac{1}{i+1} - \frac{2h}{\delta} \right| \|X_{\lambda,\delta}^{i+1} - X_0\| \\ &\leq \sum_{i=0}^{k-1} h \left(\frac{1}{\lambda} \|X_{\lambda,\delta}^{i+2} - X_{\lambda,\delta}^{i+1}\| + \frac{1}{k} \|\mathbf{A}_\lambda(X_{\lambda,\delta}^{i+1})\| \right) + \frac{1}{k} \sum_{i=0}^{N_k-1} \|X_{\lambda,\delta}^{i+1} - X_0\|, \end{aligned}$$

where inequality follows from the fact $\frac{i+1}{k} \leq 1$ for $0 \leq i \leq k-1$ and $\frac{1}{i+1} \geq \frac{2h}{\delta}$ for $0 \leq i \leq N_k-1$.

Now let's observe each term. From Lemma B.12 and Lemma D.1 we know

$$\begin{aligned} \|\dot{X}_{\lambda,\delta}(t)\| &\leq \sqrt{2} \|\mathbf{A}_\lambda(X_0)\| \leq \sqrt{2} \|m(\mathbf{A}(X_0))\| \\ \|\mathbf{A}_\lambda(X(t))\| &\leq \|\mathbf{A}_\lambda(X_0)\| \leq \|m(\mathbf{A}(X_0))\| \\ \|X_{\lambda,\delta}(t) - X_0\| &\leq 2 \|X_0 - X_\star\|. \end{aligned}$$

Name $M = \max\{\sqrt{2} \|m(\mathbf{A}(X_0))\|, 2 \|X_0 - X_\star\|\}$.

(i) $\frac{h}{\lambda} \|X_{\lambda,\delta}^{i+2} - X_{\lambda,\delta}^{i+1}\|$
First observe

$$\begin{aligned} & h \sum_{i=0}^{k-1} \frac{1}{\lambda} \|X_{\lambda,\delta}(2h(i+2)) - X_{\lambda,\delta}(2h(i+1))\| \\ &= \frac{h}{\lambda} \sum_{i=0}^{k-1} \left\| \int_{2h(i+1)}^{2h(i+2)} \dot{X}_{\lambda,\delta}(t) dt \right\| \leq \frac{h}{\lambda} \sum_{i=0}^{k-1} \int_{2h(i+1)}^{2h(i+2)} \|\dot{X}_{\lambda,\delta}(t)\| dt \leq \frac{h}{\lambda} k(2h)M \leq \frac{h}{\lambda} TM. \end{aligned}$$

Thus,

$$\begin{aligned}
& h \sum_{i=0}^{k-1} \frac{1}{\lambda} \left\| X_{\lambda,\delta}^{i+2} - X_{\lambda,\delta}^{i+1} \right\| \\
& \leq \frac{h}{\lambda} \sum_{i=0}^{k-1} \left(\left\| X_{\lambda,\delta}^{i+2} - X_{\lambda,\delta}(2h(i+2)) \right\| + \left\| X_{\lambda,\delta}(2h(i+2)) - X_{\lambda,\delta}(2h(i+1)) \right\| + \left\| X_{\lambda,\delta}(2h(i+1)) - X_{\lambda,\delta}^{i+1} \right\| \right) \\
& \leq \frac{h}{\lambda} \left(TM + 2 \sum_{i=0}^{k-1} h e^{2L_{\lambda,\delta}T} \right) \\
& \leq \frac{h}{\lambda} (TM + T e^{2L_{\lambda,\delta}T}).
\end{aligned}$$

Therefore

$$\frac{h}{\lambda} \sum_{i=0}^{k-1} \left\| X_{\lambda,\delta}^{i+2} - X_{\lambda,\delta}^{i+1} \right\| = \mathcal{O}\left(\frac{h}{\lambda}\right) + \mathcal{O}\left(\frac{h}{\lambda} e^{2L_{\lambda,\delta}T}\right).$$

(ii) $\frac{h}{k} \left\| \mathbf{A}_{\lambda}(X_{\lambda,\delta}^{i+1}) \right\|$

Since \mathbf{A}_{λ} is $\frac{1}{\lambda}$ -Lipschitz continuous and from (28)

$$\begin{aligned}
\sum_{i=0}^{k-1} \frac{h}{k} \left\| \mathbf{A}_{\lambda}(X_{\lambda,\delta}^{i+1}) \right\| & \leq \frac{h}{k} \sum_{i=0}^{k-1} \left(\left\| \mathbf{A}_{\lambda}(X_{\lambda,\delta}^i) - \mathbf{A}_{\lambda}(X_{\lambda,\delta}(2hi)) \right\| + \left\| \mathbf{A}_{\lambda}(X_{\lambda,\delta}(2hi)) \right\| \right) \\
& \leq \frac{h}{k} \sum_{i=0}^{k-1} \left(\frac{h}{\lambda} e^{2L_{\lambda,\delta}T} + M \right) = h \left(\frac{h}{\lambda} e^{2L_{\lambda,\delta}T} + M \right).
\end{aligned}$$

Therefore

$$\sum_{i=0}^{k-1} \frac{h}{k} \left\| \mathbf{A}_{\lambda}(X_{\lambda,\delta}^{i+1}) \right\| = \mathcal{O}(h) + \mathcal{O}\left(\frac{h^2}{\lambda} e^{2L_{\lambda,\delta}T}\right).$$

(iii) $\frac{1}{k} \sum_{i=0}^{N_k-1} \left\| X_{\lambda,\delta}^{i+1} - X_0 \right\|$ for $N_k = \min \{k, \lfloor \frac{\delta}{2h} \rfloor\}$.

First, observe

$$\left\| X_0 - X_{\lambda,\delta}(t) \right\| = \left\| \int_0^t \dot{X}_{\lambda,\delta}(s) ds \right\| \leq \int_0^t \left\| \dot{X}_{\lambda,\delta}(s) \right\| ds \leq tM.$$

From (27)

$$\begin{aligned}
\left\| X_{\lambda,\delta}^{i+1} - X_0 \right\| & \leq \left\| X_{\lambda,\delta}^{i+1} - X_{\lambda,\delta}(2h(i+1)) \right\| + \left\| X_{\lambda,\delta}(2h(i+1)) - X_0 \right\| \\
& \leq h(1 + 4hL_{\lambda,\delta})^{i+1} + 2h(i+1)M.
\end{aligned}$$

Now as $i+1 \leq N_k = \min \{k, \lfloor \frac{\delta}{2h} \rfloor\}$,

$$\begin{aligned}
\frac{1}{k} \sum_{i=0}^{N_k-1} \left\| X_{\lambda,\delta}^{i+1} - X_0 \right\| & \leq \frac{1}{k} \sum_{i=0}^{N_k-1} h \left((1 + 4hL_{\lambda,\delta})^{i+1} + 2(i+1)M \right) \\
& \leq \frac{h}{k} \frac{(1 + 4hL_{\lambda,\delta})^{N_k} - 1}{4hL_{\lambda,\delta}} + 2hM \sum_{i=0}^{N_k-1} \frac{i+1}{k} \\
& \leq \frac{e^{4hL_{\lambda,\delta}N_k}}{kL_{\lambda,\delta}} + 2hN_kM \\
& \leq \frac{e^{2\delta L_{\lambda,\delta}}}{L_{\lambda,\delta}} + 2\delta M = \mathcal{O}\left(\frac{e^{2\delta L_{\lambda,\delta}}}{L_{\lambda,\delta}}\right) + \mathcal{O}(\delta).
\end{aligned}$$

From (i), (ii), (iii) we have

$$\frac{1}{k} \sum_{i=0}^{k-1} (i+1)e_i = \mathcal{O}\left(\frac{h}{\lambda}\right) + \mathcal{O}\left(\frac{h}{\lambda}e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}(h) + \mathcal{O}\left(\frac{h^2}{\lambda}e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}\left(\frac{e^{2\delta L_{\lambda,\delta}}}{L_{\lambda,\delta}}\right) + \mathcal{O}(\delta).$$

Therefore,

$$\epsilon_k = 3^{\frac{T}{\lambda}} \left(\mathcal{O}\left(\frac{h}{\lambda}\right) + \mathcal{O}\left(\frac{h}{\lambda}e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}(h) + \mathcal{O}\left(\frac{h^2}{\lambda}e^{2L_{\lambda,\delta}T}\right) + \mathcal{O}\left(\frac{e^{2\delta L_{\lambda,\delta}}}{L_{\lambda,\delta}}\right) + \mathcal{O}(\delta) \right).$$

D.4 Derivation of ODE (4) from EAG and FEG

D.4.1 Derivation from EAG

For L -Lipschitz continuous monotone operator \mathbf{A} and stepsize $h > 0$, EAG-C [75] is defined as

$$\begin{aligned} z^{k+\frac{1}{2}} &= z^k - \frac{1}{k+2} (z^k - z^0) - h\mathbf{A}(z^k) \\ z^{k+1} &= z^k - \frac{1}{k+2} (z^k - z^0) - h\mathbf{A}(z^{k+\frac{1}{2}}). \end{aligned}$$

Dividing the second line by h and reorganizing we have

$$\begin{aligned} \frac{z^{k+1} - z^k}{h} &= -\mathbf{A}(z^{k+\frac{1}{2}}) - \frac{1}{h(k+2)} (z^k - z^0) \\ &= -\mathbf{A}(z^k) - \frac{1}{h(k+2)} (z^k - z^0) - \left(\mathbf{A}(z^{k+\frac{1}{2}}) - \mathbf{A}(z^k) \right). \end{aligned}$$

Identify $hk = t$, $z^k = X(t)$, $z^0 = X_0$. As z^k is a converging sequence [76], it is bounded. Thus we see

$$\begin{aligned} \left\| \mathbf{A}(z^{k+\frac{1}{2}}) - \mathbf{A}(z^k) \right\| &\leq L \left\| z^{k+\frac{1}{2}} - z^k \right\| \\ &= L \left\| \frac{h}{h(k+2)} (z^k - z^0) + h\mathbf{A}(z^k) \right\| \\ &= Lh \left\| \frac{1}{t+2h} (X(t) - X_0) + \mathbf{A}(X(t)) \right\| = \mathcal{O}(h). \end{aligned}$$

Therefore taking limit $h \rightarrow 0+$ we have

$$\dot{X}(t) = -\mathbf{A}(X(t)) - \frac{1}{t}(X(t) - X_0).$$

D.4.2 Derivation from FEG

For L -Lipschitz continuous monotone operator \mathbf{A} and stepsize $h > 0$, FEG [42] is defined as

$$\begin{aligned} z^{k+\frac{1}{2}} &= z^k - \frac{1}{k+1} (z^k - z^0) - \frac{k}{k+1} h\mathbf{A}(z^k) \\ z^{k+1} &= z^k - \frac{1}{k+1} (z^k - z^0) - h\mathbf{A}(z^{k+\frac{1}{2}}). \end{aligned}$$

Dividing the second line by h and reorganizing we have

$$\begin{aligned} \frac{z^{k+1} - z^k}{h} &= -\mathbf{A}(z^{k+\frac{1}{2}}) - \frac{1}{h(k+1)} (z^k - z^0) \\ &= -\mathbf{A}(z^k) - \frac{1}{h(k+1)} (z^k - z^0) - \left(\mathbf{A}(z^{k+\frac{1}{2}}) - \mathbf{A}(z^k) \right). \end{aligned}$$

Identify $hk = t$, $z^k = X(t)$, $z^0 = X_0$. As z^k is a converging sequence [76], it is bounded. Thus we see

$$\begin{aligned} \left\| \mathbf{A}(z^{k+\frac{1}{2}}) - \mathbf{A}(z^k) \right\| &\leq L \left\| z^{k+\frac{1}{2}} - z^k \right\| \\ &= L \left\| \frac{h}{h(k+1)}(z^k - z^0) + h \frac{hk}{h(k+1)} \mathbf{A}(z^k) \right\| \\ &= Lh \left\| \frac{1}{t+h}(X(t) - X_0) + \frac{t}{t+h} \mathbf{A}(X(t)) \right\| = \mathcal{O}(h). \end{aligned}$$

Therefore taking limit $h \rightarrow 0+$ we have

$$\dot{X}(t) = -\mathbf{A}(X(t)) - \frac{1}{t}(X(t) - X_0).$$

E Proof of convergence analysis for monotone \mathbf{A}

E.1 Extending $\tilde{\mathbf{A}}$ to $[0, \infty)$

Lemma E.1. *Suppose \mathbf{A} is a maximal monotone operator. Let X be the solution for (3). Define S as*

$$S = \left\{ t \in [0, \infty) \mid \dot{X}(t) \in -\mathbf{A}(X(t)) - \beta(t)(X(t) - X_0) \text{ is true} \right\}.$$

Then as X is the solution for (3), we know $[0, \infty) \setminus S$ is of measure zero.

Define $\tilde{\mathbf{A}}(X): S \rightarrow \mathbb{R}^n$ as

$$\tilde{\mathbf{A}}(X)(t) = -\dot{X}(t) - \beta(t)(X(t) - X_0) \quad (29)$$

and denote $\tilde{\mathbf{A}}(X(t)) = \tilde{\mathbf{A}}(X)(t)$. Assume for every $T > 0$, there is $M > 0$ such that for all $t \in S \cap [0, T]$

$$\|\tilde{\mathbf{A}}(X(t))\| \leq M.$$

Then $\tilde{\mathbf{A}}(X)$ can be extended to $t \in [0, \infty)$ with satisfying following properties.

(i) $\tilde{\mathbf{A}}(X(t)) \in \mathbf{A}(X(t))$ for all $t \in [0, \infty)$.

(ii) For $t \in [0, \infty) \setminus S$, there is a sequence $\{t_k\}_{k \in \mathbb{N}}$ such that $t_k \in S$, $\lim_{k \rightarrow \infty} t_k = t$ and $\tilde{\mathbf{A}}(X(t_k))$ converges to $\tilde{\mathbf{A}}(X(t))$.

Proof. Take $t \in [0, \infty)$, and take a sequence $\{t_n\}_{n \in \mathbb{N}}$ such that $t_n \in S$ and $\lim_{n \rightarrow \infty} t_n = t$. As a converging sequence t_n is bounded, there is $T > 0$ such that $t_n \in [0, T]$. For that T , we have $\|\tilde{\mathbf{A}}(X(t_n))\| \leq M$ by the assumption. As $n \mapsto \tilde{\mathbf{A}}(X(t_n))$ is a bounded sequence, there is a subsequence $\{t_{n_k}\}_{k \in \mathbb{N}}$ such that $k \mapsto \tilde{\mathbf{A}}(X(t_{n_k}))$ converges. Name the limit as $u = \lim_{k \rightarrow \infty} \tilde{\mathbf{A}}(X(t_{n_k}))$. On the other hand, as X is a continuous curve we have $\lim_{k \rightarrow \infty} X(t_{n_k}) = X(t)$. Then since \mathbf{A} is maximally monotone, from [15, Proposition 20.38] we have $(X(t), u) \in \text{Gra } \mathbf{A}$. Defining $\tilde{\mathbf{A}}(X(t))$ as u , we get the desired result. \square

Corollary E.2. *Let X be the solution for (6). Then $\tilde{\mathbf{A}}(X)$ defined as (29) has extension with properties stated in Lemma E.1.*

Proof. First consider the case $\beta(t) = \frac{\gamma}{t^p}$, $p > 0$, $\gamma > 0$. Take $T > 0$. It is enough to show there is $M > 0$ such that for $t \in S \cap [0, T]$

$$\|\tilde{\mathbf{A}}(X(t))\| \leq M.$$

Recall from Proposition B.11 for Yosida approximation \mathbf{A}_λ , we denoted X_λ as the solution of

$$\dot{X}_\lambda = -\mathbf{A}_\lambda(X_\lambda) - \frac{\gamma}{t^p}(X_\lambda - X_0),$$

and we have shown there is a sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ such that X_{λ_n} uniformly converges to X on $[0, T]$.

From Lemma B.12, we see for $h \neq 0$

$$\left\| \frac{X_\lambda(t+h) - X_\lambda(t)}{h} \right\| \leq \frac{\int_t^{t+h} \|\dot{X}_\lambda(s)\| ds}{h} \leq \frac{\int_t^{t+h} M_{dot}(T) ds}{h} = M_{dot}(T),$$

thus

$$\left\| \frac{X(t+h) - X(t)}{h} \right\| = \lim_{\lambda \rightarrow 0^+} \left\| \frac{X_\lambda(t+h) - X_\lambda(t)}{h} \right\| \leq M_{dot}(T).$$

Therefore for $t \in S \cap [0, T]$, $\dot{X}(t) = \lim_{h \rightarrow 0} \frac{X(t+h) - X(t)}{h}$ holds, we conclude

$$\|\dot{X}(t)\| = \lim_{h \rightarrow 0} \left\| \frac{X(t+h) - X(t)}{h} \right\| \leq M_{dot}(T).$$

And also from Lemma B.12, for $t \in [0, T]$ we have

$$\left\| \frac{\gamma}{t^p} (X_\lambda(t) - X_0) \right\| = \left\| \dot{X}_\lambda(t) + \mathbf{A}_\lambda(t) \right\| \leq \|\dot{X}_\lambda(t)\| + \|\mathbf{A}_\lambda(t)\| \leq M_{dot}(T) + M_{\mathbf{A}}(T),$$

therefore

$$\left\| \frac{\gamma}{t^p} (X(t) - X_0) \right\| = \lim_{\lambda \rightarrow 0^+} \left\| \frac{\gamma}{t^p} (X_\lambda(t) - X_0) \right\| \leq M_{dot}(T) + M_{\mathbf{A}}(T).$$

Gathering the result, for $t \in S \cap [0, T]$ we have

$$\|\tilde{\mathbf{A}}(X(t))\| = \left\| \dot{X}(t) + \frac{\gamma}{t^p} (X(t) - X_0) \right\| \leq \|\dot{X}(t)\| + \left\| \frac{\gamma}{t^p} (X(t) - X_0) \right\| \leq 2M_{dot}(T) + M_{\mathbf{A}}(T).$$

□

E.2 Proof of Proposition 3.2

We prove this theorem by deriving the energy with dilated coordinate $W(t) = C(t)(X(t) - X_0)$ and conservation law from [67]. Think of dilated coordinate $W(t) = C(t)(X(t) - X_0)$. As we're considering the case $\tilde{\mathbf{A}}$ is Lipschitz continuous, the differential inclusion (3) becomes ODE

$$\dot{X}(t) = -\tilde{\mathbf{A}}(X(t)) - \beta(t)(X(t) - X_0).$$

Rewriting the ODE in terms of W , we have

$$\frac{1}{C(t)} \left(\dot{W}(t) - \beta(t)W(t) \right) = -\tilde{\mathbf{A}}(X(W(t), t)) - \frac{\beta(t)}{C(t)} W(t)$$

where $X(W(t), t) = X(t) = \frac{W(t)}{C(t)} + X_0$. Organizing, we have

$$0 = \dot{W}(t) + C(t)\tilde{\mathbf{A}}(X(W(t), t)).$$

From Lemma B.4 we know $\tilde{\mathbf{A}}(X(W, t))$ is differentiable almost everywhere, by differentiating we obtain second order ODE which holds almost everywhere

$$\begin{aligned} 0 &= \ddot{W} + C(t)\beta(t)\tilde{\mathbf{A}}(X(W(t), t)) + C(t)\frac{d}{dt}\tilde{\mathbf{A}}(X(W(t), t)) \\ &= \ddot{W} - \beta(t)\dot{W} + C(t)\frac{d}{dt}\tilde{\mathbf{A}}(X(W(t), t)). \end{aligned} \tag{30}$$

Now by taking inner product with \dot{W} and integrating, we get equality which was referred as conservation law in [67]

$$E_1 \equiv \frac{1}{2} \|\dot{W}(t)\|^2 - \int_{t_0}^t \beta(s) \|\dot{W}(s)\|^2 ds + \int_{t_0}^t C(s) \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{W}(s) \right\rangle ds.$$

Since $\dot{W}(t) = C(t) \left(\dot{X}(t) + \beta(t)(X(t) - X_0) \right)$, we can rewrite the integrand in the last term as

$$\int_{t_0}^t C(s) \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{W}(s) \right\rangle ds = \int_{t_0}^t C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds + \int_{t_0}^t \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), C(s)\beta(s)W(s) \right\rangle ds.$$

Note the purpose was to obtain the first term, which is nonnegative due to monotonicity of $\tilde{\mathbf{A}}$. Now taking integration by parts to the second term we have

$$\begin{aligned} & \int_{t_0}^t \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), C(s)\beta(s)W(s) \right\rangle ds \\ &= [\langle \tilde{\mathbf{A}}(X(W, s)), C(s)\beta(s)W(s) \rangle]_{t_0}^t - \int_{t_0}^t \left\langle \tilde{\mathbf{A}}(X(W, s)), \left(C(s)\beta(s)^2 + C(s)\dot{\beta}(s) \right) W(s) + C(s)\beta(s)\dot{W}(s) \right\rangle ds \\ &= [\langle \tilde{\mathbf{A}}(X(W, s)), C(s)\beta(s)W(s) \rangle]_{t_0}^t + \int_{t_0}^t \beta(s) \|\dot{W}(s)\|^2 ds + \int_{t_0}^t \left(\beta(s)^2 + \dot{\beta}(s) \right) \langle \dot{W}(s), W(s) \rangle ds \\ &= [\langle \tilde{\mathbf{A}}(X(W, s)), C(s)\beta(s)W(s) \rangle]_{t_0}^t + \int_{t_0}^t \beta(s) \|\dot{W}(s)\|^2 ds \\ &+ \left[\left(\beta(s)^2 + \dot{\beta}(s) \right) \frac{1}{2} \|W(s)\|^2 \right]_{t_0}^t - \frac{1}{2} \int_{t_0}^t \left(2\beta(s)\dot{\beta}(s) + \ddot{\beta}(s) \right) \|W(s)\|^2 ds \end{aligned}$$

On the second equality, we used the fact $\dot{W}(t) = -C(t)\tilde{\mathbf{A}}(X(W(t), t))$. Note the fundamental theorem of calculus for $\langle \tilde{\mathbf{A}}(X(W, s)), C(s)\beta(s)W(s) \rangle$ is valid since $\tilde{\mathbf{A}}(X(W, s))$ is Lipschitz continuous and $C(s)\beta(s)W(s)$ is continuously differentiable in $[t_0, t]$, so their inner product is absolutely continuous in $[t_0, t]$.

Observe the integrand in the last term can be rewritten as

$$\left(2\beta(s)\dot{\beta}(s) + \ddot{\beta}(s) \right) \|W(s)\|^2 = C(s)^2 \left(2\beta(s)\dot{\beta}(s) + \ddot{\beta}(s) \right) \|X(s) - X_0\|^2 = \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2.$$

Now gathering the results, we conclude

$$\begin{aligned} E_1 &\equiv \frac{1}{2} \|\dot{W}(t)\|^2 - \int_{t_0}^t \beta(s) \|\dot{W}(s)\|^2 ds + \int_{t_0}^t C(s) \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{W}(s) \right\rangle ds \\ &= \frac{C(t)^2}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 + [\langle \tilde{\mathbf{A}}(X(W(s), s)), C(s)\beta(s)W(s) \rangle]_{t_0}^t + \left[\left(\beta(s)^2 + \dot{\beta}(s) \right) \frac{1}{2} \|W(s)\|^2 \right]_{t_0}^t \\ &+ \int_{t_0}^t C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds - \frac{1}{2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds \\ &= \frac{C(t)^2}{2} \left(\|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \left(\beta(t)^2 + \dot{\beta}(t) \right) \|X(t) - X_0\|^2 \right) \\ &+ \int_{t_0}^t C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds - \frac{1}{2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds \\ &- \underbrace{\frac{C(t_0)^2}{2} \left(2\beta(t_0) \langle \tilde{\mathbf{A}}(X(t_0)), X(t_0) - X_0 \rangle + \left(\beta(t_0)^2 + \dot{\beta}(t_0) \right) \|X(t_0) - X_0\|^2 \right)}_{=\text{constant}} \end{aligned}$$

Moving the constant terms to left hand side and naming $E = E_1 - \text{constant}$, we get the desired result.

E.3 Proof of Corollary 3.3

E.3.1 V is nonincreasing when $\tilde{\mathbf{A}}$ is Lipschitz continuous monotone

We first check it is true for the case $\tilde{\mathbf{A}}$ is Lipschitz continuous. Then from Proposition 3.2 we can write $V(t)$ as

$$V(t) = E - 2 \int_{t_0}^t C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds$$

with E in Proposition 3.2. As $\tilde{\mathbf{A}}$ is monotone, from (1) we know $C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle \geq 0$ holds almost everywhere. Therefore for $h > 0$

$$V(t+h) - V(t) = -2 \int_t^{t+h} C(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds \leq 0,$$

we see V is a nonincreasing function. Therefore for all $t > 0$, $V(t) \leq \lim_{\epsilon \rightarrow 0^+} V(\epsilon)$. It remains to show the limit $\lim_{\epsilon \rightarrow 0^+} V(\epsilon)$ exists.

E.3.2 Calculation of $V(0) = \lim_{t \rightarrow 0^+} V(t)$ for Lipschitz continuous monotone $\tilde{\mathbf{A}}$

In this section, we calculate $\lim_{t \rightarrow 0^+} V(t)$ when $\tilde{\mathbf{A}}$ is Lipschitz continuous. Recall V was defined as

$$\begin{aligned} V(t) &= \frac{C(t)^2}{2} \left(\|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \right) \\ &\quad - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds. \end{aligned}$$

From now we denote V for the case $t_0 = 0$ as V^0 .

We first check

$$C(t) = \begin{cases} t^\gamma & p = 1 \\ e^{\frac{\gamma}{1-p} t^{1-p}} & p > 0, p \neq 1. \end{cases} \quad (31)$$

Observing

$$\begin{aligned} \int_1^t \frac{\gamma}{s} ds &= \gamma \log t \\ \int_0^t \frac{\gamma}{s^p} ds &= \frac{\gamma}{1-p} t^{1-p} \quad \text{for } 0 < p < 1 \\ \int_\infty^t \frac{\gamma}{s^p} ds &= \frac{\gamma}{1-p} t^{1-p} \quad \text{for } p > 1, \end{aligned}$$

we see $C(t)$ defined above agrees with the definition of $C(t)$ for each case. Note

$$\lim_{t \rightarrow 0^+} C(t) = \begin{cases} 0 & p \geq 1 \\ 1 & 0 < p < 1. \end{cases}$$

Now we will show

$$\lim_{t \rightarrow 0^+} V^0(t) = \lim_{t \rightarrow 0^+} \frac{C(t)^2}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 = \begin{cases} 0 & \text{if } p \geq 1 \\ \frac{\|\tilde{\mathbf{A}}(X_0)\|^2}{2} & \text{if } 0 < p < 1. \end{cases} \quad (32)$$

To do so, we first show for $t > 0$

$$\lim_{\epsilon \rightarrow 0^+} \int_\epsilon^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds < \infty.$$

We first provide an elementary fact as a lemma.

Lemma E.3. *Let $f: (0, \infty) \rightarrow \mathbb{R}$ is a continuous function. Suppose there is $q < 1$, $0 < l < \infty$ such that*

$$\limsup_{t \rightarrow 0^+} |f(t)t^q| < l.$$

Then for $t > 0$, $f \in L^1([0, t], \mathbb{R})$.

Proof. Since $\limsup_{s \rightarrow 0^+} f(s)s^q = l$, there is $\epsilon \in (0, t)$ such that

$$0 < s < \epsilon \implies |f(s)s^q| < 2|l|.$$

Then since $q < 1$

$$\int_0^\epsilon |f(s)| ds = \int_0^\epsilon |f(s)s^q| \frac{1}{s^q} ds \leq \int_0^\epsilon \frac{2|l|}{s^q} ds = \left[\frac{2|l|}{1-q} s^{1-q} \right]_0^\epsilon = \frac{2|l|}{1-q} \epsilon^{1-q} < \infty.$$

By the way as $f(t)$ is continuous on $[\epsilon, t]$, $M = \max_{s \in [\epsilon, t]} |f(s)|$ exists. Therefore,

$$\int_0^t |f(s)| ds = \int_0^\epsilon |f(s)| ds + \int_\epsilon^t |f(s)| ds \leq \frac{2|l|}{1-q} \epsilon^{1-q} + M(t - \epsilon) < \infty.$$

□

Applying Lemma E.3, we will show for $t > 0$

$$\left| s^2 \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \right| \in L^1([0, t], \mathbb{R}). \quad (33)$$

Observe

$$\frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) = 2C(s)^2 \beta(s) \dot{\beta}(s) + C(s)^2 \ddot{\beta}(s) = C(s)^2 \left(-\frac{2p\gamma^2}{s^{2p+1}} + \frac{p(1+p)\gamma}{s^{p+2}} \right). \quad (34)$$

(i) $p > 1$

We first show $\lim_{s \rightarrow 0^+} C(s)^2 \frac{1}{s^n} = 0$ for all $n > 0$. Take $n > 0$. Then there is some $k \in \mathbb{N}$ such that $k(p-1) > n$. With change of variable $u = \frac{1}{s}$ and L'Hôpital's rule we see

$$\begin{aligned} \lim_{s \rightarrow 0^+} C(s)^2 \frac{1}{s^n} &= \lim_{u \rightarrow \infty} \frac{u^n}{e^{\frac{2\gamma}{p-1} u^{p-1}}} \\ &= \lim_{u \rightarrow \infty} \frac{nu^{n-1}}{2\gamma u^{p-2} e^{\frac{2\gamma}{p-1} u^{p-1}}} = \frac{n}{2\gamma} \lim_{u \rightarrow \infty} \frac{u^{n+1-p}}{e^{\frac{2\gamma}{p-1} u^{p-1}}} = \dots = \frac{\prod_{m=0}^{k-1} (n - m(p-1))}{(2\gamma)^k} \lim_{u \rightarrow \infty} \frac{u^{n-k(p-1)}}{e^{\frac{2\gamma}{p-1} u^{p-1}}} = 0. \end{aligned} \quad (35)$$

And thus

$$\lim_{s \rightarrow 0^+} \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \right| = \lim_{s \rightarrow 0^+} \left| C(s)^2 \left(-\frac{2p\gamma^2}{s^{2p-1}} + \frac{p(1+p)\gamma}{s^p} \right) \right| = 0.$$

By Lemma E.3, we conclude $\int_0^t \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \right| ds < \infty$.

(ii) $p = 1$

Since $C(s) = s^\gamma$, we see

$$\lim_{s \rightarrow 0^+} \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \cdot s^{1-\gamma} \right| = \lim_{s \rightarrow 0^+} \left| s^2 \cdot \gamma(\gamma-1) s^{2\gamma-3} \cdot s^{1-\gamma} \right| = \gamma|\gamma-1| \lim_{s \rightarrow 0^+} s^\gamma = 0.$$

Since $1 - \gamma < 1$, by Lemma E.3 we conclude $\int_0^t \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \right| ds < \infty$.

(iii) $0 < p < 1$

Since $\lim_{s \rightarrow 0^+} C(s) = \lim_{s \rightarrow 0^+} e^{\frac{\gamma}{1-p} s^{1-p}} = 1$, we see

$$\limsup_{s \rightarrow 0^+} \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \cdot s^p \right| = \limsup_{s \rightarrow 0^+} C(s)^2 \left| -2p\gamma^2 s^{1-p} + p(1+p)\gamma \right| = p(1+p)\gamma.$$

Since $p < 1$, by Lemma E.3 we conclude $\int_0^t \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \right| ds < \infty$.

Naming the bound in Corollary B.8 as $M(T)$, we know for $0 < s < T$

$$\left\| \frac{X(s) - X_0}{s} \right\| \leq \frac{\int_0^s \|\dot{X}(u)\| du}{s} \leq \frac{\int_0^s M(T) du}{s} = M(T).$$

Therefore applying (33), for $0 < t < T$ we have

$$\int_0^t \left| \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 \right| ds \leq M(T)^2 \int_0^t \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \right| ds < \infty.$$

Hence we know $V^0(t)$ is well defined and since $\lim_{t \rightarrow 0+} \int_0^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds = 0$, we have

$$\lim_{t \rightarrow 0+} V^0(t) = \lim_{t \rightarrow 0+} \frac{C(t)^2}{2} \left(\|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \right).$$

Now we are ready to show the desired result.

(i) $p > 1$

As we know $X(t) - X_0$ and $\tilde{\mathbf{A}}(X(t))$ are bounded from Lemma D.1 and Corollary B.9. Therefore from (35) we have

$$0 = \lim_{t \rightarrow 0+} C(t)^2 = \lim_{t \rightarrow 0+} C(t)^2 \beta(t) = \lim_{t \rightarrow 0+} C(t)^2 (\beta(t)^2 + \dot{\beta}(t)),$$

therefore $\lim_{t \rightarrow 0+} V^0(t) = 0$.

(ii) $0 < p \leq 1$

As $\left\| \frac{X(t) - X_0}{t} \right\| \leq M(T) < \infty$ for $0 < t < T$, we see

$$\begin{aligned} \limsup_{t \rightarrow 0+} C(t)^2 \beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle &\leq \gamma \limsup_{t \rightarrow 0+} C(t)^2 t^{1-p} \|\tilde{\mathbf{A}}(X(t))\| M(T) = 0 \\ \limsup_{t \rightarrow 0+} C(t)^2 (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 &\leq \gamma \limsup_{t \rightarrow 0+} C(t)^2 (\gamma t^{2-2p} - t^{1-p}) M(T)^2 = 0. \end{aligned}$$

Therefore

$$\lim_{t \rightarrow 0+} V^0(t) = \lim_{t \rightarrow 0+} \frac{C(t)^2}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 = \begin{cases} 0 & \text{if } p = 1 \\ \frac{\|\tilde{\mathbf{A}}(X_0)\|^2}{2} & \text{if } 0 < p < 1. \end{cases}$$

From (i) and (ii) we get the desired conclusion

$$V^0(0) = \lim_{t \rightarrow 0+} V^0(t) = \begin{cases} 0 & \text{if } p \geq 1 \\ \frac{\|\tilde{\mathbf{A}}(X_0)\|^2}{2} & \text{if } 0 < p < 1, \end{cases}$$

and therefore for general $t_0 \geq 0$,

$$V(t) = \lim_{t \rightarrow 0+} V(t) = \lim_{t \rightarrow 0+} V^0(t) - \int_0^{t_0} \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds$$

is well-defined.

E.3.3 $V(t) \leq \lim_{n \rightarrow \infty} V_{\lambda_n}(0)$ holds for $t \in S$ and general maximal monotone \mathbf{A}

Define S as defined in Lemma E.1. Take $t \in S$, let $T > t$. Let $\{\lambda_n\}_{n \in \mathbb{N}}$ be a positive sequence λ_n that $\lim_{n \rightarrow \infty} \lambda_n = 0$, X_{λ_n} converges to X uniformly on $[0, T]$ and \dot{X}_{λ_n} converges weakly to \dot{X} in $L^2([0, T], \mathbb{R}^n)$. Recall existence of such sequence was guaranteed by Proposition B.11.

Recall we denoted X_λ as the solution of the ODE (19). Denote V_λ as V for the case $\mathbf{A} = \mathbf{A}_\lambda$, i.e.

$$V_\lambda(t) = \frac{C(t)^2}{2} \left(\|\mathbf{A}_\lambda(X_\lambda(t))\|^2 + 2\beta(t) \langle \mathbf{A}_\lambda(X_\lambda(t)), X_\lambda(t) - X_0 \rangle + (\beta(t)^2 + \dot{\beta}(t)) \|X_\lambda(t) - X_0\|^2 \right) - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X_\lambda(t) - X_0\|^2 ds.$$

Note equality $\dot{X}(t) = -\tilde{\mathbf{A}}(X(t)) - \beta(t)(X(t) - X_0)$ holds since $t \in S$, therefore we have

$$V(t) = \frac{C(t)^2}{2} \left(\|\dot{X}(t)\|^2 + \dot{\beta}(t) \|X(t) - X_0\|^2 \right) - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds.$$

The goal of this section is to show

$$V(t) \leq \limsup_{n \rightarrow \infty} V_{\lambda_n}(t) \leq \limsup_{n \rightarrow \infty} V_{\lambda_n}(0) = \lim_{n \rightarrow \infty} V_{\lambda_n}(0).$$

(1) $V(t) \leq \limsup_{n \rightarrow \infty} V_{\lambda_n}(t)$

First observe, from Lemma B.12 we know

$$\frac{\|X_\lambda(s) - X_0\|}{s} \leq \frac{\int_0^s \|\dot{X}_\lambda(s)\| ds}{s} \leq \frac{\int_0^s M_{dot}(T) ds}{s} = M_{dot}(T)$$

holds for $s \leq T$. Thus from (33) we have

$$\left| \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X_\lambda(s) - X_0\|^2 \right| \leq \left| s^2 \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \right| M_{dot}(T)^2 \in L^1([0, T], \mathbb{R}).$$

Therefore applying dominated convergence theorem, we have for $t_0 \in [0, T]$

$$\lim_{n \rightarrow \infty} \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X_{\lambda_n}(s) - X_0\|^2 ds = \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds. \quad (36)$$

From elementary analysis, we can easily check $\limsup_{n \rightarrow \infty} (a_n + b_n) = \limsup_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n$ holds when $\lim_{n \rightarrow \infty} b_n$ exists. Thus we have

$$\limsup_{n \rightarrow \infty} V_{\lambda_n}(t) - V(t) = \frac{C(t)^2}{2} \left(\limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(t)\|^2 - \|\dot{X}(t)\|^2 \right).$$

Therefore it suffices to show following lemma.

Lemma E.4. *Suppose for $T > 0$ and sequence $\{\lambda_n\}_{n \in \mathbb{N}}$, X_{λ_n} converges to X uniformly on $[0, T]$ and \dot{X}_{λ_n} converges weakly to \dot{X} in $L^2([0, T], \mathbb{R}^n)$. Let $t \in S \cap [0, T]$. Then following inequality is true*

$$\|\dot{X}(t)\| \leq \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(t)\|.$$

Proof. (i) $\|\dot{X}(s)\| \leq \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\|$ holds for almost every $s \in [0, T]$.

Let D be a measurable subset $D \subset [0, T]$. Since $\dot{X}_{\lambda_n} \rightharpoonup \dot{X}$ in $L^2([0, T], \mathbb{R}^n)$ and $\chi_D \dot{X} \in L^2([0, T], \mathbb{R}^n)$, we have

$$\begin{aligned} \int_D \|\dot{X}(s)\|^2 ds &= \int_0^T \langle \dot{X}(s), \chi_D(s) \dot{X}(s) \rangle ds = \lim_{n \rightarrow \infty} \int_0^T \langle \dot{X}_{\lambda_n}(s), \chi_D(s) \dot{X}(s) \rangle ds \\ &= \limsup_{n \rightarrow \infty} \int_D \langle \dot{X}_{\lambda_n}(s), \dot{X}(s) \rangle ds \leq \limsup_{n \rightarrow \infty} \int_D \|\dot{X}_{\lambda_n}(s)\| \|\dot{X}(s)\| ds. \end{aligned}$$

The inequality comes from the Cauchy-Schwarz inequality. Now from Reverse Fatou's Lemma we have

$$\limsup_{n \rightarrow \infty} \int_D \|\dot{X}_{\lambda_n}(s)\| \|\dot{X}(s)\| ds \leq \int_D \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\| \|\dot{X}(s)\| ds.$$

Therefore combining two inequalities we have

$$\int_D \left(\limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\| - \|\dot{X}(s)\| \right) \|\dot{X}(s)\| ds \geq 0.$$

As D was arbitrary measurable subset of $[0, T]$, we conclude for almost every $s \in [0, T]$

$$\left(\limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\| - \|\dot{X}(s)\| \right) \|\dot{X}(s)\| \geq 0 \implies \|\dot{X}(s)\| \leq \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\|.$$

(ii) $\|\dot{X}(t)\| \leq \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(t)\|$ for $t \in S \cap [0, T]$.

Let $t \in S \cap [0, T]$. Then for $h > 0$ such that $t + h < T$, since S is measure zero, from (i) we have

$$\int_t^{t+h} \|\dot{X}(s)\| ds \leq \int_t^{t+h} \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\| ds.$$

Now, for some $a > 0$ consider

$$U_{\lambda_n}(s) = \|\dot{X}_{\lambda_n}(s)\|^2 + \underbrace{\frac{\gamma p}{s^{p+1}} \|X_{\lambda_n}(s) - X_0\|^2 + \int_a^s \frac{\gamma p(1-p)}{u^{p+2}} \|X_{\lambda_n}(u) - X_0\|^2 du}_{=f_n(s)}.$$

Then from the proof Lemma B.5, we know

$$\dot{U}_{\lambda_n}(s) = -2 \left\langle \dot{X}_{\lambda_n}(s), \frac{d}{ds} \mathbf{A}_{\lambda_n}(X(s)) \right\rangle - \frac{2\gamma}{s^p} \left\| \dot{X}_{\lambda_n}(s) - \frac{p}{s} (X_{\lambda_n}(s) - X_0) \right\|^2 \leq 0$$

holds for almost every s , thus U_{λ_n} is nonincreasing.

By the way, as X_{λ_n} converges to X uniformly on $[0, T]$, using dominated convergence theorem we have

$$\lim_{n \rightarrow \infty} f_n(s) = \frac{\gamma p}{s^{p+1}} \|X(s) - X_0\|^2 + \int_a^s \frac{\gamma p(1-p)}{u^{p+2}} \|X(u) - X_0\|^2 du.$$

Denote $f(s) = \lim_{n \rightarrow \infty} f_n(s)$.

Using above facts, for $s \in [t, t+h]$ we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\| &= \limsup_{n \rightarrow \infty} \sqrt{U_{\lambda_n}(s) - f_n(s)} \\ &\leq \limsup_{n \rightarrow \infty} \sqrt{U_{\lambda_n}(t) - f_n(s)} = \sqrt{\limsup_{n \rightarrow \infty} U_{\lambda_n}(t) - f(s)}. \end{aligned} \quad (37)$$

Therefore,

$$\|X(t+h) - X(t)\| \leq \int_t^{t+h} \|\dot{X}(s)\| ds \leq \int_t^{t+h} \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(s)\| ds \leq \int_t^{t+h} \sqrt{\limsup_{n \rightarrow \infty} U_{\lambda_n}(t) - f(s)} ds.$$

Note $\sqrt{\limsup_{n \rightarrow \infty} U_{\lambda_n}(t) - f(s)}$ is a continuous function with respect to s . Thus we have

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \int_t^{t+h} \sqrt{\limsup_{n \rightarrow \infty} U_{\lambda_n}(t) - f(s)} ds &= \sqrt{\limsup_{n \rightarrow \infty} U_{\lambda_n}(t) - f(t)} \\ &= \sqrt{\limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(t)\|^2 + \lim_{n \rightarrow \infty} f_n(t) - f(t)} = \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(t)\|. \end{aligned}$$

Finally as $t \in S$, $\dot{X}(t) = \lim_{h \rightarrow 0} \frac{X(t+h) - X(t)}{h}$ exists. Therefore

$$\begin{aligned} \|\dot{X}(t)\| &= \lim_{h \rightarrow 0+} \frac{\|X(t+h) - X(t)\|}{h} \\ &\leq \lim_{h \rightarrow 0+} \frac{1}{h} \int_t^{t+h} \sqrt{\limsup_{n \rightarrow \infty} U_{\lambda_n}(t) - f(s)} ds = \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(t)\|, \end{aligned}$$

we conclude the desired result. \square

- (2) $\limsup_{n \rightarrow \infty} V_{\lambda_n}(t) \leq \limsup_{n \rightarrow \infty} V_{\lambda_n}(0)$
 As \mathbb{A}_{λ_n} is Lipschitz continuous, from Appendix E.3.1 we know V_{λ_n} is nonincreasing, and thus

$$V_{\lambda_n}(t) \leq \lim_{\epsilon \rightarrow 0^+} V_{\lambda_n}(\epsilon) = V_{\lambda_n}(0).$$

Taking limsup both sides we get the desired result.

- (3) $\limsup_{n \rightarrow \infty} V_{\lambda_n}(0) = \lim_{n \rightarrow \infty} V_{\lambda_n}(0)$
 From (32) and (ii) of Lemma B.10, we know

$$\lim_{n \rightarrow \infty} V_{\lambda_n}^0(0) = \begin{cases} 0 & \text{if } p \geq 1 \\ \frac{\|m(\mathbb{A}(X_0))\|^2}{2} & \text{if } 0 < p < 1. \end{cases}$$

And applying (36) we have

$$\lim_{n \rightarrow \infty} V_{\lambda_n}(0) = \lim_{n \rightarrow \infty} V_{\lambda_n}^0(0) - \int_0^{t_0} \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds.$$

As the limit $\lim_{n \rightarrow \infty} V_{\lambda_n}(0)$ exists, the limsup coincides with the limit.

E.3.4 Proof for general maximal monotone \mathbb{A}

First, we show $V(t) \leq \lim_{n \rightarrow \infty} V_{\lambda_n}(0)$ holds for $t \in [0, \infty)$. Next, we show $\lim_{n \rightarrow \infty} V_{\lambda_n}(0) = \lim_{t \rightarrow 0^+} V(t)$. Then we have

$$V(t) \leq \lim_{n \rightarrow \infty} V_{\lambda_n}(0) = \lim_{t \rightarrow 0^+} V(t) = V(0),$$

which is our desired result.

- (i) $V(t) \leq \lim_{n \rightarrow \infty} V_{\lambda_n}(0)$ holds for $t \in [0, \infty)$

Take $t \in [0, \infty)$. We know the inequality is true when $t \in S$, thus assume $t \notin S$. Then from Lemma E.1, we know there is a sequence $\{t_k\}_{k \in \mathbb{N}}$ such that $t_k \in S$, $\lim_{k \rightarrow \infty} t_k = t$ and $\tilde{\mathbb{A}}(X(t_k))$ converges to $\tilde{\mathbb{A}}(X(t))$. As $t_k \in S$, $V(t_k) \leq \lim_{n \rightarrow \infty} V_{\lambda_n}(0)$ holds. Since $\lim_{k \rightarrow \infty} V(t_k) = V(t)$, taking limit $k \rightarrow \infty$ to the inequality we get the desired result.

- (ii) $\lim_{n \rightarrow \infty} V_{\lambda_n}(0) = \lim_{t \rightarrow 0^+} V(t)$

Take a sequence $\{t_k\}_{k \in \mathbb{N}}$ such that $t_k > 0$ and $\lim_{k \rightarrow \infty} t_k = 0$. We wish to show $\lim_{k \rightarrow \infty} V(t_k) = \lim_{n \rightarrow \infty} V_{\lambda_n}(0)$. Note the arguments in Appendix E.3.2 are valid when $\|\tilde{\mathbb{A}}(X(t_k))\|$ and $\frac{\|X(t_k) - X_0\|}{t_k}$ are bounded for all $k \in \mathbb{N}$. The boundedness of $\|\tilde{\mathbb{A}}(X(t_k))\|$ comes from Corollary E.2. And from Lemma B.12 we have

$$\frac{\|X(t_k) - X_0\|}{t_k} = \lim_{\lambda \rightarrow 0^+} \frac{\|X_{\lambda}(t_k) - X_0\|}{t_k} \leq \lim_{\lambda \rightarrow 0^+} \frac{\int_0^{t_k} \|\dot{X}_{\lambda}(s)\| ds}{t_k} \leq M_{dot}(t_k) \leq \sup_{k \in \mathbb{N}} M_{dot}(t_k).$$

Therefore applying the arguments in Appendix E.3.2 we have

$$\lim_{k \rightarrow \infty} V^0(t_k) = \lim_{k \rightarrow \infty} \frac{C(t_k)^2}{2} \|\tilde{\mathbb{A}}(X(t_k))\|^2.$$

Thus it remains to show the limit on the right hand side exists and is equal to $\lim_{n \rightarrow \infty} V_{\lambda_n}^0(0)$.

For $p \geq 1$, we know $\lim_{k \rightarrow \infty} C(t_k)^2 = 0$, thus $\lim_{k \rightarrow \infty} \frac{C(t_k)^2}{2} \|\tilde{\mathbb{A}}(X(t_k))\|^2 = 0$ since $\|\tilde{\mathbb{A}}(X(t_k))\|$ is bounded. As $\lim_{n \rightarrow \infty} V_{\lambda_n}^0(0) = 0$ from (32), we're done.

Now consider the case $0 < p < 1$. Suppose $\tilde{\mathbb{A}}(X(t_{k_l}))$ is a convergent subsequence of $\tilde{\mathbb{A}}(X(t_k))$. First observe from (i) we have

$$V^0(t_{k_l}) \leq \lim_{n \rightarrow \infty} V_{\lambda_n}^0(0) = \frac{\|m(\mathbb{A}(X_0))\|^2}{2}.$$

From above inequality, recalling $\lim_{l \rightarrow \infty} C(t_{k_l})^2 = 1$ we have

$$\frac{\left\| \lim_{l \rightarrow \infty} \tilde{\mathbb{A}}(X(t_{k_l})) \right\|^2}{2} = \lim_{l \rightarrow \infty} V^0(t_{k_l}) \leq \frac{\|m(\mathbb{A}(X_0))\|^2}{2}.$$

By the way as $\lim_{l \rightarrow \infty} X(t_{k_l}) = X_0$ and $\tilde{\mathbf{A}}(X(t_{k_l})) \in \mathbf{A}(X(t_{k_l}))$, by closed graph theorem we have $\lim_{l \rightarrow \infty} \tilde{\mathbf{A}}(X(t_{k_l})) \in \mathbf{A}(X_0)$. As of $m(\mathbf{A}(X_0))$ is the element in $\mathbf{A}(X_0)$ with smallest norm, we have $\lim_{l \rightarrow \infty} \tilde{\mathbf{A}}(X(t_{k_l})) = m(\mathbf{A}(X_0))$. As all convergent subsequence converges to the same limit $m(\mathbf{A}(X_0))$, we conclude $\lim_{k \rightarrow \infty} \tilde{\mathbf{A}}(X(t_k)) = m(\mathbf{A}(X_0))$.

Therefore

$$\lim_{k \rightarrow \infty} V^0(t_k) = \lim_{k \rightarrow \infty} \frac{C(t_k)^2}{2} \|\tilde{\mathbf{A}}(X(t_k))\|^2 = \frac{\|m(\mathbf{A}(X_0))\|^2}{2} = \lim_{n \rightarrow \infty} V_{\lambda_n}^0(0).$$

As t_k was arbitrary positive sequence converges to 0, we conclude $\lim_{n \rightarrow \infty} V_{\lambda_n}^0(0) = \lim_{t \rightarrow 0^+} V^0(t) = V^0(0)$. Therefore

$$\lim_{n \rightarrow \infty} V_{\lambda_n}(0) = V^0(0) - \int_0^{t_0} \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds = \lim_{t \rightarrow 0^+} V(t).$$

E.4 Proof of Lemma 3.4

Most of the proof is done in the main text. Recall $\Phi(t)$ is defined as $\Phi(t) = \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle$, and from (8) we have

$$\Phi(t) \geq \frac{1}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 - 2\beta(t)^2 \|X_0 - X_\star\|^2.$$

On the other hand,

$$\frac{2V(t)}{C(t)^2} - \Phi(t) = (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 - \frac{2}{C(t)^2} \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds.$$

Note $C(t) \neq 0$ if $t > 0$, we can divide with $C(t)^2$. Now from Corollary 3.3 we have we have $V(t) \geq V(0)$ for $t > 0$, and so $\frac{V(t)}{C(t)^2} - \frac{V(0)}{C(t)^2} \geq 0$. Thus for $t > \epsilon$

$$\begin{aligned} & \frac{2V(0)}{C(t)^2} - (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 + \frac{2}{C(t)^2} \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds \\ &= \frac{2V(0)}{C(t)^2} - \left(\frac{2V(t)}{C(t)^2} - \Phi(t) \right) \\ &= \left(\frac{2V(0)}{C(t)^2} - \frac{2V(t)}{C(t)^2} \right) + \Phi(t) \geq \Phi(t) \geq \frac{1}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 - 2\beta(t)^2 \|X_0 - X_\star\|^2. \end{aligned}$$

Moving $2\beta(t)^2 \|X_0 - X_\star\|^2$ to the left hand side and multiplying with 2 we get the desired result

$$\begin{aligned} \|\tilde{\mathbf{A}}(X(t))\|^2 &\leq 4\beta(t)^2 \|X_0 - X_\star\|^2 + \frac{4V(0)}{C(t)^2} - 2(\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \\ &\quad + \frac{2}{C(t)^2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds. \end{aligned}$$

E.5 Proof of Theorem 3.1

Restating Lemma 3.4, we know for $t > 0$,

$$\begin{aligned} \|\tilde{\mathbf{A}}(X(t))\|^2 &\leq 4\beta(t)^2 \|X_0 - X_\star\|^2 + \frac{2V(0)}{C(t)^2} - 2(\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \\ &\quad + \frac{2}{C(t)^2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds. \end{aligned} \tag{7}$$

As we know $\|X(t) - X_0\| \leq 2\|X_0 - X_\star\|$ from Lemma D.1, it is clear that

$$\begin{aligned} 4\beta(t)^2 \|X_0 - X_\star\|^2 &= \mathcal{O}(\beta(t)^2) \\ \frac{2V(0)}{C(t)^2} &= \mathcal{O}\left(\frac{1}{C(t)^2}\right) \\ 2\left(\beta(t)^2 + \dot{\beta}(t)\right) \|X(t) - X_0\|^2 &= \mathcal{O}(\beta(t)^2) + \mathcal{O}(\dot{\beta}(t)). \end{aligned} \quad (38)$$

Therefore it remains to show

$$\frac{2}{C(t)^2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds = \mathcal{O}(\beta(t)^2) + \mathcal{O}\left(\frac{1}{C(t)^2}\right) + \mathcal{O}(\dot{\beta}(t)).$$

We can check there is $T > 0$ such that for $s > T$ the sign of $\frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right)$ does not change, for $\beta(t) = \frac{\gamma}{t^p}$ with $\gamma > 0$, $p > 0$. We first proceed our proof with assuming this condition. The point for this condition is that following equality holds for $t > T$.

$$\left| \int_T^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) ds \right| = \int_T^t \left| \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \right| ds.$$

Applying above equality and using $\|X(t) - X_0\| \leq 2\|X_0 - X_\star\|$, we see

$$\begin{aligned} & \left| \frac{2}{C(t)^2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds \right| \\ & \leq \frac{2}{C(t)^2} \underbrace{\left| \int_{t_0}^T \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds \right|}_{=M} + \frac{2}{C(t)^2} \int_T^t \left| \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 \right| ds \\ & \leq \frac{2M}{C(t)^2} + \frac{2}{C(t)^2} \int_T^t \left| \frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) 4\|X_0 - X_\star\|^2 \right| ds \\ & = \frac{2M}{C(t)^2} + \frac{8\|X_0 - X_\star\|^2}{C(t)^2} \left| C(t)^2 \dot{\beta}(t) - C(T)^2 \dot{\beta}(T) \right| \\ & \leq \frac{2}{C(t)^2} \left(M + 4\|X_0 - X_\star\|^2 C(T)^2 |\dot{\beta}(T)| \right) + 8\|X_0 - X_\star\|^2 |\dot{\beta}(t)| = \mathcal{O}\left(\frac{1}{C(t)^2}\right) + \mathcal{O}(\dot{\beta}(t)). \end{aligned}$$

Therefore from (7) and (38), we conclude

$$\|\tilde{\mathbf{A}}(X(t))\|^2 = \mathcal{O}(\beta(t)^2) + \mathcal{O}\left(\frac{1}{C(t)^2}\right) + \mathcal{O}(\dot{\beta}(t)).$$

It remains to show there is $T > 0$ such that for $s > T$ the sign of $\frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right)$ does not change. It can be shown by easy, but a little complicated calculations. Since $\dot{C}(t) = C(t)\beta(t)$, we see

$$\frac{d}{ds} \left(C(s)^2 \dot{\beta}(s) \right) = 2C(s)^2 \beta(s) \dot{\beta}(s) + C(s)^2 \ddot{\beta}(s) = C(s)^2 \left(2\beta(s) \dot{\beta}(s) + \ddot{\beta}(s) \right).$$

Therefore it is enough to check the sign of $2\beta(s) \dot{\beta}(s) + \ddot{\beta}(s)$. Observe

$$2\beta(s) \dot{\beta}(s) + \ddot{\beta}(s) = -\frac{2p\gamma^2}{s^{2p+1}} + \frac{p(1+p)\gamma}{s^{p+2}} = \begin{cases} \frac{2}{s^3} \gamma(1-\gamma) & \text{if } p = 1 \\ \frac{p(p+1)\gamma}{s^{2p+1}} \left(s^{p-1} - \frac{2\gamma}{p+1} \right) & \text{if } p \neq 1. \end{cases}$$

Thus when $p = 1$, for all $s > 0$ it is nonpositive if $\gamma \geq 1$ and nonnegative if $0 < \gamma < 1$.

When $p \neq 1$, we see

$$\lim_{s \rightarrow 0^+} s^{p-1} = \begin{cases} \infty & \text{if } 0 < p < 1 \\ 0 & \text{if } p > 1 \end{cases}, \quad \lim_{s \rightarrow \infty} s^{p-1} = \begin{cases} 0 & \text{if } 0 < p < 1 \\ \infty & \text{if } p > 1 \end{cases}$$

Therefore by intermediate value theorem there is $T > 0$ such that $T^{p-1} - \frac{2\gamma}{p+1} = 0$, and for that T we have $s^{p-1} - \frac{2\gamma}{p+1}$ is nonpositive for $s > T$ if $0 < p < 1$ and nonnegative for $s > T$ if $p > 1$. This concludes the proof.

E.5.1 Proof for the convergence rate in Table 1

Recall, from (31) we have

$$C(t) = \begin{cases} t^\gamma & p = 1 \\ e^{\frac{\gamma}{1-p}t^{1-p}} & p > 0, p \neq 1. \end{cases}$$

We now observe $\dot{\beta}(t)$ does not effect to convergence rate. In other words, we show $\dot{\beta}(t)$ is not the slowest one that goes to zero compared to $\beta(t)^2$ and $\frac{1}{C(t)^2}$ for every case.

(i) $0 < p \leq 1$

Comparing $\dot{\beta}(t) = -\frac{p\gamma}{t^{p+1}}$ with $\beta(t)^2 = \frac{\gamma^2}{t^{2p}}$, we see $\dot{\beta}(t) = \mathcal{O}(\beta(t)^2)$ when $0 < p \leq 1$.

(ii) $p > 1$

For $p > 1$, we see $\lim_{t \rightarrow \infty} \frac{1}{C(t)^2} = 1 \neq 0$. As $\lim_{t \rightarrow \infty} \dot{\beta}(t) = -\lim_{t \rightarrow \infty} \frac{p\gamma}{t^{p+1}} = 0$, we have $\dot{\beta}(t) = \mathcal{O}\left(\frac{1}{C(t)^2}\right)$.

From (i) and (ii), we conclude

$$\|\tilde{\mathbf{A}}(X(t))\|^2 = \mathcal{O}(\beta(t)^2) + \mathcal{O}\left(\frac{1}{C(t)^2}\right).$$

Now the results in Table 1 is straightforward.

• $p = 1$

When $p = 1$, we have $C(t) = t^\gamma$. Comparing $\frac{1}{C(t)^2} = \frac{1}{t^{2\gamma}}$ and $\beta(t)^2 = \frac{\gamma^2}{t^2}$,

$$(1) \|\tilde{\mathbf{A}}(X(t))\|^2 = \mathcal{O}(\beta(t)^2) = \mathcal{O}\left(\frac{1}{t^2}\right) \text{ if } \gamma \geq 1.$$

$$(2) \|\tilde{\mathbf{A}}(X(t))\|^2 = \mathcal{O}\left(\frac{1}{C(t)^2}\right) = \mathcal{O}\left(\frac{1}{t^{2\gamma}}\right) \text{ if } 0 < \gamma < 1.$$

• $0 < p < 1$

When $0 < p < 1$, we have

$$\lim_{t \rightarrow \infty} \frac{1}{\beta(t)^2} \cdot \frac{1}{C(t)^2} = \frac{1}{\gamma^2} \lim_{t \rightarrow \infty} \frac{t^2}{e^{\frac{2\gamma}{1-p}t^{1-p}}} = 0.$$

Therefore, $\|\tilde{\mathbf{A}}(X(t))\|^2 = \mathcal{O}(\beta(t)^2) = \mathcal{O}\left(\frac{1}{t^{2p}}\right)$.

• $p > 1$

We observed previously that $\lim_{t \rightarrow \infty} \frac{1}{C(t)^2} = 1 \neq 0$ when $p > 1$. Therefore $\|\tilde{\mathbf{A}}(X(t))\|^2 = \mathcal{O}\left(\frac{1}{C(t)^2}\right) = \mathcal{O}(1)$.

E.6 Proof of Theorem 3.5

Name

$$\bar{X}_* = \Pi_{\text{Zer } \mathbf{A}}(X_0) = \underset{z \in \text{Zer } \mathbf{A}}{\text{argmin}} \|z - X_0\|.$$

We first show

$$\limsup_{t \rightarrow \infty} \langle X(t) - \bar{X}_*, X_0 - \bar{X}_* \rangle \leq 0.$$

Proof by contradiction. Suppose not.

Then there is $\epsilon > 0$ such that for every $k \in \mathbb{N}$, there is $n_k \in [0, \infty)$ such that

$$\langle X(n_k) - \bar{X}_*, X_0 - \bar{X}_* \rangle > \epsilon$$

By the way, from Lemma D.1 we know $X(n_k) \in \bar{B}_{\|X_0 - \bar{X}_*\|}(\bar{X}_*)$. Thus by Bolzano-Weierstrass theorem, there is a converging subsequence $\{X(n_{k(l)})\}_{l \in \mathbb{N}}$. Name the limit as X_∞ . Since $\lim_{t \rightarrow \infty} \|\tilde{\mathbf{A}}(X(t))\| = 0$, $\{\tilde{\mathbf{A}}(X(n_{k(l)}))\}_{l \in \mathbb{N}}$ converges to 0. Then from [15, Proposition 20.38], $(X_\infty, 0) \in \text{Gra } \mathbf{A}$, i.e. $X_\infty \in \text{Zer } \mathbf{A}$.

On the other hand, as \mathbf{A} is maximal monotone, by [15, Proposition 23.39] $\text{Zer } \mathbf{A}$ is closed and convex. Since $\bar{X}_* = \Pi_{\text{Zer } \mathbf{A}}(X_0)$, by [15, Theorem 3.16] we have $\langle X_\infty - \bar{X}_*, X_0 - \bar{X}_* \rangle \leq 0$. Thus

$$0 < \epsilon \leq \lim_{l \rightarrow \infty} \langle X(n_{k(l)}) - \bar{X}_*, X_0 - \bar{X}_* \rangle = \langle X_\infty - \bar{X}_*, X_0 - \bar{X}_* \rangle \leq 0,$$

this is a contradiction, therefore we get the desired result.

Now we show $\lim_{t \rightarrow \infty} \|X(t) - \bar{X}_*\| = 0$. Recall, for $t > 0$

$$\dot{X} \in -\mathbf{A}(X(t)) - \beta(t)(X - X_0).$$

From $\dot{C}(t) = C(t)\beta(t)$, and monotonicity of \mathbf{A} , we observe

$$\begin{aligned} \frac{d}{dt} \left(C(t)^2 \|X(t) - \bar{X}_*\|^2 \right) &= 2C(t)^2 \langle \dot{X}(t), X(t) - \bar{X}_* \rangle + 2C(t)^2 \beta(t) \|X(t) - \bar{X}_*\|^2 \\ &= 2C(t)^2 \langle -\tilde{\mathbf{A}}(X(t)) - \beta(t)(X(t) - X_0), X(t) - \bar{X}_* \rangle + 2C(t)^2 \beta(t) \|X(t) - \bar{X}_*\|^2 \\ &\leq -2C(t)^2 \beta(t) \langle X(t) - X_0, X(t) - \bar{X}_* \rangle + 2C(t)^2 \beta(t) \|X(t) - \bar{X}_*\|^2 \\ &= 2C(t)^2 \beta(t) \langle X_0 - \bar{X}_*, X(t) - \bar{X}_* \rangle. \end{aligned}$$

Now take ϵ . From $\limsup_{t \rightarrow \infty} \langle X(t) - \bar{X}_*, X_0 - \bar{X}_* \rangle \leq 0$, there is $M > 0$ such that

$$t > M \implies \langle X_0 - \bar{X}_*, X(t) - \bar{X}_* \rangle < \epsilon.$$

For that M and $t > M$, integrating from M to t we get

$$\begin{aligned} \left[C(s)^2 \|X(s) - \bar{X}_*\|^2 \right]_M^t &= C(t)^2 \|X(t) - \bar{X}_*\|^2 - C(M)^2 \|X(M) - \bar{X}_*\|^2 \\ &\leq \int_M^t 2C(s)^2 \beta(s) \langle X_0 - \bar{X}_*, X(s) - \bar{X}_* \rangle ds \\ &\leq \int_M^t |2C(s)^2 \beta(s) \langle X_0 - \bar{X}_*, X(s) - \bar{X}_* \rangle| ds \\ &\leq \int_M^t 2C(s)^2 \beta(s) \epsilon ds = \epsilon [C(s)^2]_M^t = \epsilon (C(t)^2 - C(M)^2). \end{aligned}$$

By dividing $C(t)^2$ and organizing,

$$\|X(t) - \bar{X}_*\|^2 \leq \epsilon \left(1 - \left(\frac{C(M)}{C(t)} \right)^2 \right) + \left(\frac{C(M)}{C(t)} \right)^2 \|X(M) - \bar{X}_*\|^2.$$

By the way from assumption, we have $\lim_{t \rightarrow \infty} \frac{1}{C(t)} = 0$. Therefore we conclude

$$\lim_{t \rightarrow \infty} \|X(t) - \bar{X}_*\|^2 \leq \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we get the desired result.

F Proof for worst case examples

The explicit solution for linear A is crucially used in worst case examples. Therefore, we first provide proof for it.

F.1 Proof for explicit solution for linear A

F.1.1 Proof of Lemma 4.1

Observing

$$X(t) = \sum_{n=0}^{\infty} \frac{(-tA)^n}{\Gamma(n + \gamma + 1)} \Gamma(\gamma + 1) X_0 = X_0 + \sum_{n=1}^{\infty} \frac{(-tA)^n}{\Gamma(n + \gamma + 1)} \Gamma(\gamma + 1) X_0,$$

we can check $X(0) = X_0$. Now by using the property of Gamma function $\Gamma(x+1) = x\Gamma(x)$, and paying attention to the lower bound of the summation index we have

$$\begin{aligned}\dot{X}(t) &= \sum_{n=1}^{\infty} \frac{(-nA)(-tA)^{n-1}}{\Gamma(n+\gamma+1)} \Gamma(\gamma+1) X_0 = A \sum_{n=1}^{\infty} \frac{(-n-\gamma+\gamma)(-tA)^{n-1}}{\Gamma(n+\gamma+1)} \Gamma(\gamma+1) X_0 \\ &= -A \sum_{n=1}^{\infty} \frac{(-tA)^{n-1}}{\Gamma(n+\gamma)} \Gamma(\gamma+1) X_0 + \gamma \sum_{n=1}^{\infty} \frac{(-t)^{n-1} A^n}{\Gamma(n+\gamma+1)} \Gamma(\gamma+1) X_0 \\ &= -A \sum_{n=0}^{\infty} \frac{(-tA)^n}{\Gamma(n+\gamma+1)} \Gamma(\gamma+1) X_0 + \frac{\gamma}{(-t)} \sum_{n=1}^{\infty} \frac{(-t)^n A^n}{\Gamma(n+\gamma+1)} \Gamma(\gamma+1) X_0 = -AX(t) - \frac{\gamma}{t} (X(t) - X_0).\end{aligned}$$

The solution can be written in another form

$$X(t) = \gamma e^{-tA} t^{-\gamma} \int_0^t u^{\gamma-1} e^{uA} du X_0. \quad (39)$$

As this is a special case of (9), here we just briefly check this satisfies the ODE and check other details in the proof of Lemma 4.2. By product rule of differentiation,

$$\begin{aligned}\dot{X} &= \gamma \left(\frac{d}{dt} e^{-tA} \right) t^{-\gamma} \int_0^t u^{\gamma-1} e^{uA} du X_0 + \gamma e^{-tA} \left(\frac{d}{dt} t^{-\gamma} \right) \int_0^t u^{\gamma-1} e^{uA} du X_0 + \gamma e^{-tA} t^{-\gamma} \frac{d}{dt} \left(\int_0^t u^{\gamma-1} e^{uA} du X_0 \right) \\ &= -A(X(t)) - \frac{\gamma}{t} X(t) + \gamma e^{-tA} t^{-\gamma} (t^{\gamma-1} e^{tA}) X_0 \\ &= -A(X(t)) - \frac{\gamma}{t} (X(t) - X_0).\end{aligned}$$

F.1.2 Proof of Lemma 4.2

Define

$$X_*(t) = \left(I - \frac{Ae^{-tA}}{C(t)} \left(\int_0^t e^{sA} C(s) ds \right) \right) X_0.$$

We show X_* is a well-defined solution for (3) with linear \mathbf{A} , then show X_* is equal to $X(t)$ defined in (9).

We first check well-definedness. By definition, $C(t) = e^{\int_{t_0}^t \beta(s) ds}$ is nondecreasing. Also, $\|e^{tA}\|$ is also nondecreasing since from monotonicity of A we have for all $x \in \mathbb{R}^n$

$$\frac{d}{dt} \|e^{tA} x\|^2 = 2 \langle A(e^{tA} x), e^{tA} x \rangle \geq 0.$$

Now we see $X(t)$ is well-defined since

$$\begin{aligned}\left\| e^{-tA} \frac{1}{C(t)} \left(\int_0^t e^{sA} C(s) ds \right) \right\| &= \left\| \int_0^t e^{(s-t)A} \frac{C(s)}{C(t)} ds \right\| \\ &\leq \int_0^t \left\| e^{(s-t)A} \right\| \left\| \frac{C(s)}{C(t)} \right\| ds \leq \int_0^t (1 \cdot 1) ds = t < \infty.\end{aligned}$$

Also above inequality implies second term reaches to zero as $t \rightarrow 0+$, we have $\lim_{t \rightarrow 0+} X_*(t) = X_0$. Thus defining $X_*(0) = X_0$ if necessary, we see X_* satisfies the initial condition with $X_* \in \mathcal{C}([0, \infty), \mathbb{R}^n)$.

We then check $X_*(t)$ becomes the solution. This is immediate from product rule for differentiation. Since $\frac{d}{dt} \left(\frac{1}{C(t)} \right) = -\frac{1}{C(t)^2} C(t) \beta(t) = -\frac{1}{C(t)} \beta(t)$, we have

$$\begin{aligned}\dot{X}_*(t) &= - \left(\frac{d}{dt} e^{-tA} \right) \frac{1}{C(t)} \left(\int_0^t e^{sA} C(s) ds \right) X_0 \\ &\quad - A e^{-tA} \left(\frac{d}{dt} \frac{1}{C(t)} \right) \left(\int_0^t e^{sA} C(s) ds \right) X_0 \\ &\quad - A e^{-tA} \frac{1}{C(t)} \frac{d}{dt} \left(\int_0^t e^{sA} C(s) ds \right) X_0 \\ &= -A(X_*(t) - X_0) - \beta(t)(X_*(t) - X_0) - A(X_0) = -A(X(t)) - \beta(t)(X_*(t) - X_0).\end{aligned}$$

Finally we now show $X(t) = X_*(t)$, where $X(t)$ is defined as (9). From integral by parts we have

$$\begin{aligned} X(t) &= e^{-tA} \frac{1}{C(t)} \left(\int_0^t e^{sA} C(s) \beta(s) ds + C(0)I \right) X_0 \\ &= e^{-tA} \frac{1}{C(t)} \left([e^{sA} C(s)]_0^t - \int_0^t A e^{sA} C(s) ds + C(0)I \right) X_0 \\ &= \left(I - A e^{-tA} \frac{1}{C(t)} \left(\int_0^t e^{sA} C(s) ds \right) \right) X_0 = X_*(t). \end{aligned}$$

F.2 Proof of Theorem 4.3

If $\beta(t) \equiv 0$, for $A := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, we have $X(t) = e^{-tA}$, so

$$\lim_{t \rightarrow \infty} \|A(X(t))\| = 1 \neq 0.$$

So let's consider the case β is not $\beta(t) \equiv 0$ and $\beta(t) \geq 0$. For

$$A_\xi = 2\pi\xi \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

name the solution for ODE $\dot{X}_\xi = -A_\xi(X_\xi) - \beta(t)(X_\xi - X_0)$ as X_ξ . By Lemma 4.2 we have

$$X_\xi(t) = \frac{e^{-tA_\xi}}{C(t)} \left(\int_0^t e^{sA_\xi} C(s) \beta(s) ds + C(0)I \right) X_0.$$

We want to show,

$$\exists \xi \in \mathbb{R}, \quad \lim_{t \rightarrow \infty} \|A_\xi(X_\xi(t))\| \neq 0$$

From $\lim_{t \rightarrow \infty} \frac{1}{C(t)} \neq 0$, we see $\lim_{t \rightarrow \infty} \left\| \frac{e^{-tA_\xi}}{C(t)} \right\| = 1 \cdot \frac{1}{C(t)} \neq 0$. And since A_ξ is invertible except the case $\xi = 0$, we have

$$\begin{aligned} \lim_{t \rightarrow \infty} \|A_\xi(X_\xi(t))\| = 0 &\iff \lim_{t \rightarrow \infty} \|X_\xi(t)\| = 0 \\ &\iff \left\| \lim_{t \rightarrow \infty} \int_0^t e^{sA_\xi} C(s) \beta(s) ds + C(0)I \right\| = 0. \end{aligned}$$

Now let assume $\forall \xi \in \mathbb{R}, \lim_{t \rightarrow \infty} \|A_\xi(X_\xi(t))\| = 0$ and lead to contradiction. Define $f : \mathbb{R} \rightarrow \mathbb{R}$ as

$$f(s) = \begin{cases} C(s)\beta(s) & s > 0 \\ 0 & s \leq 0. \end{cases}$$

Then $\int_{-\infty}^{\infty} f(s) ds = [C(s)]_0^{\infty} = \lim_{t \rightarrow \infty} C(t) - C(0) < \infty$ we have $f \in L^1(\mathbb{R}, \mathbb{R})$. Note $\lim_{t \rightarrow \infty} C(t)$ exists, since C is nondecreasing and by the assumption is bounded.

Now setting $X_0 = (1, 0)^T$ and from $e^{sA_\xi} = \begin{pmatrix} \cos 2\pi\xi s & \sin 2\pi\xi s \\ -\sin 2\pi\xi s & \cos 2\pi\xi s \end{pmatrix}$, we see

$$\begin{aligned} \int_0^{\infty} e^{sA_\xi} C(s) \beta(s) ds X_0 &= \left(\int_0^{\infty} \cos(2\pi s \xi) f(s) ds, \int_0^{\infty} \sin(-2\pi s \xi) f(s) ds \right)^T \\ &= \left(\operatorname{Re}(\hat{f}(\xi)), \operatorname{Im}(\hat{f}(\xi)) \right)^T \end{aligned}$$

where $\hat{f}(\xi)$ is Fourier transform of f .

From assumption we have $\forall \xi \in \mathbb{R}, \left\| \int_0^{\infty} e^{sA_\xi} C(s) \beta(s) ds X_0 + C(0)X_0 \right\| = 0$, we have

$$\left(\operatorname{Re}(\hat{f}(\xi)), \operatorname{Im}(\hat{f}(\xi)) \right)^T \equiv -C(0)X_0$$

By the way, from Fourier theory we know \hat{f} vanishes at infinity, thus

$$\|C(0)X_0\| = \lim_{\xi \rightarrow \infty} |\hat{f}(\xi)| = 0.$$

Therefore we have $\hat{f}(\xi) = 0$ for all $\xi \in \mathbb{R}$.

Now $\hat{f} \equiv 0$ is clearly $L^1(\mathbb{R}, \mathbb{R})$, we can apply *Fourier inversion formula* and conclude $f \equiv 0$ almost everywhere.

However, $f(s) = C(s)\beta(s)$ is not zero almost everywhere since $\beta(s)$ is not constantly zero. Thus a contradiction, we get the desired result.

F.3 Proof of Theorem 4.4

F.3.1 Proof for the case $p = 1, \gamma \geq 1$

Recall from (39), we know

$$X(t) = \gamma e^{-tA} t^{-\gamma} \int_0^t e^{sA} s^{\gamma-1} ds X_0.$$

Without loss of generality, we consider the case $X_0 = (1, 0)^T$.

(i) $\gamma = 1$

Plugging $\gamma = 1$ gives

$$X(t) = \frac{e^{-tA}}{t} A^{-1} [e^{sA}]_0^t X_0 = \frac{1}{t} A^{-1} (I - e^{-tA}) (1, 0)^T = \frac{1}{t} A^{-1} (1 - \cos t, \sin t)^T$$

Therefore

$$\lim_{t \rightarrow \infty} \|tA(X(t))\| = \lim_{t \rightarrow \infty} \|(1 - \cos t, \sin t)^T\| \neq 0.$$

(ii) $\gamma > 1$

We will show $\lim_{t \rightarrow \infty} \|tA(X(t))\| = \gamma$. With change of variable $s = tv$ and integration by parts we have

$$\begin{aligned} \left(\frac{1}{\gamma}\right) tA(X(t)) &= e^{-tA} A t^{-(\gamma-1)} \int_0^t e^{sA} s^{\gamma-1} ds X_0 \\ &= e^{-tA} tA \int_0^1 e^{tvA} v^{\gamma-1} dv X_0 \\ &= e^{-tA} [e^{tvA} v^{\gamma-1}]_0^1 X_0 - (\gamma-1) e^{-tA} \int_0^1 e^{tvA} v^{\gamma-2} dv X_0 \\ &= X_0 - (\gamma-1) e^{-tA} \left(\int_0^1 \cos(tv) v^{\gamma-2} dv, \int_0^1 \sin(tv) v^{\gamma-2} dv \right)^T \end{aligned}$$

By the way, from $\gamma > 1$ we have $v^{\gamma-2} \in L^1[0, 1]$, by Riemann-Lebesgue lemma we have

$$\lim_{t \rightarrow \infty} \int_0^1 \cos(tv) v^{\gamma-2} dv = \lim_{t \rightarrow \infty} \int_0^1 \sin(tv) v^{\gamma-2} dv = 0.$$

Observe as e^{-tA} is a rotation, we know

$$\left\| e^{-tA} \left(\int_0^1 \cos(tv) v^{\gamma-2} dv, \int_0^1 \sin(tv) v^{\gamma-2} dv \right) \right\| = \left\| \left(\int_0^1 \cos(tv) v^{\gamma-2} dv, \int_0^1 \sin(tv) v^{\gamma-2} dv \right) \right\|.$$

Taking limit $t \rightarrow \infty$ we know right hand side converges to zero, we conclude

$$\lim_{t \rightarrow \infty} e^{-tA} \left(\int_0^1 \cos(tv) v^{\gamma-2} dv, \int_0^1 \sin(tv) v^{\gamma-2} dv \right) = 0.$$

Therefore,

$$\begin{aligned} \lim_{t \rightarrow \infty} \|tA(X(t))\| &= \left\| \gamma X_0 - \gamma(\gamma-1) \lim_{t \rightarrow \infty} e^{-tA} \left(\int_0^1 \cos(tv) v^{\gamma-2} dv, \int_0^1 \sin(tv) v^{\gamma-2} dv \right)^T \right\| \\ &= \gamma \|X_0\| = \gamma \end{aligned}$$

F.3.2 Proof for the case $p = 1, \gamma < 1$

Since A, e^{tA} are invertible, we see

$$\lim_{t \rightarrow \infty} t^{2\gamma} \|A(X(t))\|^2 \neq 0 \iff \lim_{t \rightarrow \infty} t^\gamma \|A(X(t))\| \neq 0 \iff \lim_{t \rightarrow \infty} t^\gamma \left\| \frac{1}{\gamma} e^{tA} X(t) \right\| \neq 0.$$

Therefore it is enough to observe $\frac{1}{\gamma} e^{tA} t^\gamma X(t)$. Again for $X_0 = (1, 0)^T$

$$\begin{aligned} \frac{1}{\gamma} e^{tA} t^\gamma X(t) &= \int_0^t e^{sA} s^{\gamma-1} ds X_0 \\ &= \left(\int_0^t \cos(s) s^{\gamma-1} ds, \int_0^t \sin(s) s^{\gamma-1} ds \right)^T. \end{aligned}$$

Since $s^{\gamma-1}$ decrease monotonically to zero, we may apply similar argument with alternative series test.

Define $a_n = \int_{(n-1)\pi}^{n\pi} \sin(s) s^{\gamma-1} ds$ and name $S_n = \sum_{k=1}^n a_k$. Then for $m \in \mathbb{N}, t > 2m\pi$ we see

$$S_{2m} \leq \int_0^t \sin(s) s^{\gamma-1} ds \leq S_{2m-1}.$$

By the way $S = \lim_{n \rightarrow \infty} S_n$ exists by alternative series test, thus we have

$$\lim_{t \rightarrow \infty} \int_0^t \sin(s) s^{\gamma-1} ds = S$$

by squeeze theorem. Note $S \geq S_2 > 0$. With similar argument, we can conclude $\lim_{t \rightarrow \infty} \int_0^t \cos(s) s^{\gamma-1} ds$ also exists.

Therefore we conclude $\lim_{t \rightarrow \infty} \|t^\gamma X(t)\| \neq 0$ since

$$\lim_{t \rightarrow \infty} \|t^\gamma X(t)\| = \lim_{t \rightarrow \infty} \gamma \left\| \left(\int_0^t \cos(s) s^{\gamma-1} ds, \int_0^t \sin(s) s^{\gamma-1} ds \right)^T \right\| \geq \gamma S > 0.$$

F.3.3 Proof for the case $0 < p < 1$

Recall from (9), we have

$$X(t) = \frac{e^{-tA}}{C(t)} \left(\int_0^t e^{sA} C(s) \beta(s) ds + C(0)I \right) X_0.$$

Our goal is to show

$$\lim_{t \rightarrow \infty} t^{2p} \|A(X(t))\|^2 \neq 0.$$

We first observe, it is enough to show

$$\lim_{t \rightarrow \infty} \frac{1}{\beta(t)C(t)} \int_0^t C(s) \beta(s) \sin s ds \neq 0.$$

This follows from below facts.

(1) Since $\beta(t) = \frac{\gamma}{t^p}$ and A is linear and invertible,

$$\lim_{t \rightarrow \infty} t^{2p} \|A(X(t))\|^2 = \lim_{t \rightarrow \infty} \frac{\gamma^2}{\beta(t)^2} \|A(X(t))\|^2 \neq 0 \iff \lim_{t \rightarrow \infty} \left\| \frac{1}{\beta(t)} X(t) \right\| \neq 0.$$

(2) Recall from (31), we have $C(t) = e^{\frac{\gamma}{1-p} t^{1-p}}$ for $\beta(t) = \frac{\gamma}{t^p}$, so $C(0) = \lim_{t \rightarrow 0^+} C(t) = 1$ and $\lim_{t \rightarrow \infty} \frac{1}{\beta(t)C(t)} = 0$.

Thus $\lim_{t \rightarrow \infty} \frac{1}{\beta(t)C(t)} \|C(0)I\| = 0$, second term is ignorable. Therefore the problem reduces to

$$\lim_{t \rightarrow \infty} \left\| \frac{e^{-tA}}{\beta(t)C(t)} \left(\int_0^t C(s) \beta(s) e^{-sA} ds \right) X_0 \right\| \neq 0.$$

(3) Since e^{tA} is linear and invertible, problem reduces to

$$\lim_{t \rightarrow \infty} \left\| \frac{1}{\beta(t)C(t)} \left(\int_0^t C(s)\beta(s)e^{-sA} ds \right) X_0 \right\| \neq 0.$$

(4) Again without loss of generality, let $X_0 = (1, 0)^T$. Recalling $e^{-sA} = \begin{pmatrix} \cos s & \sin s \\ -\sin s & \cos s \end{pmatrix}$, focusing on one component we see it is sufficient to show

$$\lim_{t \rightarrow \infty} \frac{1}{\beta(t)C(t)} \int_0^t C(s)\beta(s) \sin s ds \neq 0.$$

Now we prove the statement. For convenience, name $D(t) = \beta(t)C(t)$. We want to show

$$\lim_{t \rightarrow \infty} \underbrace{\int_0^t \frac{D(s)}{D(t)} \sin s ds}_{=S(t)} \neq 0.$$

We first observe

$$\lim_{t \rightarrow \infty} \sup_{\delta \in [0, \delta]} \left| \frac{D(t+\delta)}{D(t+\pi)} - 1 \right| = 0.$$

To do so, we first show $\lim_{t \rightarrow \infty} \frac{D(t+\delta)}{D(t+\pi)} = 1$ for $\delta \in [0, \pi]$. Take $\delta \in [0, \pi]$. Observe

$$\frac{D(t+\delta)}{D(t+\pi)} = \frac{\beta(t+\delta)C(t+\delta)}{\beta(t+\pi)C(t+\pi)} = \left(\frac{t+\pi}{t+\delta} \right)^p \frac{e^{\frac{\gamma}{1-p}(t+\delta)^{1-p}}}{e^{\frac{\gamma}{1-p}(t+\pi)^{1-p}}} = \left(\frac{t+\pi}{t+\delta} \right)^p e^{\frac{\gamma}{1-p}(t+\delta)^{1-p} \left(1 - \left(\frac{t+\pi}{t+\delta} \right)^{1-p} \right)}.$$

Considering L'Hôpital's rule for the exponent, we see

$$\lim_{t \rightarrow \infty} (t+\delta)^{1-p} \left(1 - \left(\frac{t+\pi}{t+\delta} \right)^{1-p} \right) = \lim_{t \rightarrow \infty} \frac{1 - \left(\frac{t+\pi}{t+\delta} \right)^{1-p}}{(t+\delta)^{p-1}} = \lim_{t \rightarrow \infty} \frac{-\pi(p-1) \left(\frac{t+\pi}{t+\delta} \right)^{-p}}{(t+\delta)^2} = \lim_{t \rightarrow \infty} \frac{-\pi}{(t+\pi)^p} = 0.$$

As the exponent reaches to zero, we conclude $\lim_{t \rightarrow \infty} \frac{D(t+\delta)}{D(t+\pi)} = 1$.

Now from

$$\dot{D}(t) = C(t) \left(\beta(t)^2 + \dot{\beta}(t) \right) = e^{\frac{\gamma}{1-p}t^{1-p}} \gamma t^{-2p-1} (\gamma t - pt^p),$$

we see $D(t)$ is nondecreasing for $t \geq \left(\frac{\gamma}{p} \right)^{\frac{1}{p-1}}$ since $0 < p < 1$. Therefore for $t \geq \left(\frac{\gamma}{p} \right)^{\frac{1}{p-1}}$ we have $\sup_{\delta \in [0, \delta]} \left| \frac{D(t+\delta)}{D(t+\pi)} - 1 \right| = \max \left\{ \left| \frac{D(t)}{D(t+\pi)} - 1 \right|, \left| \frac{D(t+\delta)}{D(t+\pi)} - 1 \right| \right\}$, so it reaches to zero as $t \rightarrow \infty$.

Now we prove desired statement. Proof by contradiction. Suppose $\lim_{t \rightarrow \infty} S(t) = 0$. Then for $0 < \epsilon < \frac{1}{2}$, there is $T_1 > 0$ such that $t > T_1$ implies $|S(t)| < \epsilon$. Now for $t > T_1$, observe

$$\begin{aligned} 2\epsilon &> |S(t+\pi) - S(t)| = \left| \int_0^{t+\pi} \frac{D(s)}{D(t+\pi)} \sin s ds - \int_0^t \frac{D(s)}{D(t)} \sin s ds \right| \\ &= \left| \frac{1}{D(t+\pi)} \left(\int_0^{t+\pi} D(s) \sin s ds - \int_0^t D(s) \sin s ds \right) + \left(\frac{1}{D(t+\pi)} - \frac{1}{D(t)} \right) \int_0^t D(s) \sin s ds \right| \\ &\geq \left| \int_t^{t+\pi} \frac{D(s)}{D(t+\pi)} \sin s ds \right| - \left| \left(\frac{D(t)}{D(t+\pi)} - 1 \right) \int_0^t \frac{D(s)}{D(t)} \sin s ds \right|. \end{aligned}$$

On the other hand, there is T_2 such that $t > T_2$ implies $\sup_{\delta \in [0, \pi]} \left| \frac{D(t+\delta)}{D(t+\pi)} - 1 \right| < \frac{\epsilon}{2}$. Now for $t = 2n\pi > \max\{T_1, T_2\}$,

$$2\epsilon > \left| \int_{2n\pi}^{(2n+1)\pi} \frac{D(s)}{D(t+\pi)} \sin s ds \right| - \left| \frac{D(t)}{D(t+\pi)} - 1 \right| \times |S(t)| > \int_{2n\pi}^{(2n+1)\pi} (1-\epsilon) |\sin s| ds - \epsilon = 2 - 2\epsilon.$$

This contradicts the fact $\epsilon < \frac{1}{2}$, we prove the assumption $\lim_{t \rightarrow \infty} \int_0^t S(t) = 0$ is not true. Therefore $\lim_{t \rightarrow \infty} \int_0^t S(t) \neq 0$.

G Proof of convergence analysis for discrete counterpart

G.1 Correspondence between discrete method in Theorem 5.1 and continuous model (6)

To check the correspondence of the method with (6), we provide a informal derivation. Assume operator \mathbf{A} is continuous. Then we have $y^k = x^{k+1} + \mathbf{A}x^{k+1}$, by substituting y^k and y^{k-1} , the method can be equivalently written as

$$x^{k+1} + \mathbf{A}x^{k+1} = \frac{k^p}{k^p + \gamma}(x^k - \mathbf{A}x^k) + \frac{\gamma}{k^p + \gamma}x^0. \quad (40)$$

This can be considered as a special case of below method, with $h = 1$.

$$x^{k+1} + h\mathbf{A}x^{k+1} = \frac{k^p}{k^p + h^{1-p}\gamma}(x^k - h\mathbf{A}x^k) + \frac{h^{1-p}\gamma}{k^p + h^{1-p}\gamma}x^0.$$

Dividing by h both sides and reorganizing, we have

$$\frac{x^{k+1} - x^k}{h} = -\mathbf{A}x^{k+1} - \frac{h^p k^p}{h^p k^p + h\gamma}\mathbf{A}x^k - \frac{\gamma}{h^p k^p + h\gamma}(x^k - x^0).$$

Identifying $hk = t$, $x^k = X(t)$ and taking $h \rightarrow 0$, we have

$$\dot{X}(t) = -2\mathbf{A}(X(t)) - \frac{\gamma}{t^p}(X - X_0).$$

As monotonicity is preserved for scalar multiple, rescaling $2\mathbf{A}$ to \mathbf{A} does not change the class of operators that the ODE covers. For notation simplicity, replacing $2\mathbf{A}$ to \mathbf{A} we have

$$\dot{X}(t) = -\mathbf{A}(X(t)) - \frac{\gamma}{t^p}(X - X_0).$$

G.2 Proof of boundedness of $\|x^k\|$

While proving Theorem 5.1, we need a upper bound for $\|x^k\|$. Therefore we first prove following lemma.

Lemma G.1. *Let \mathbf{A} be a maximal monotone operator. Consider a method*

$$\begin{aligned} x^k &= \mathbf{J}_{\mathbf{A}}y^{k-1} \\ y^k &= (1 - \beta_k)(2x^k - y^{k-1}) + \beta_k x^0 \end{aligned}$$

for $k = 1, 2, \dots$, with sequence $\{\beta_k\}_{k \in \mathbb{N}}$, $0 \leq \beta_k \leq 1$ and initial condition $y^0 = x^0 \in \mathbb{R}^n$. Then following holds

$$\|x^k - x^*\| \leq \|x^0 - x^*\|.$$

for $x^* \in \text{Zer}\mathbf{A}$. And so, $\|x^k - x^0\| \leq 2\|x^0 - x^*\|$.

Proof. Recall from Lemma D.2, we know for $\mathbf{T} = \mathbf{R}_{\mathbf{A}} = 2\mathbf{J}_{\mathbf{A}} - \mathbf{I}$ given method is equivalent to below method

$$y^{k+1} = \beta_k y^0 + (1 - \beta_k)\mathbf{T}y^k.$$

Note that $x^* \in \text{Zer}\mathbf{A} \Leftrightarrow 0 \in \mathbf{A}x^* \Leftrightarrow x^* = \mathbf{J}_{\mathbf{A}}x^* \Leftrightarrow x^* = \mathbf{T}x^* \Leftrightarrow x^* \in \text{Fix}\mathbf{T}$. We first prove $\|y^k - x^*\| \leq \|y^0 - x^*\|$ by induction. The statement is trivially true when $k = 0$. Now suppose the statement is true for $k \in \mathbb{N}$. Then

$$\begin{aligned} \|y^{k+1} - x^*\| &= \|\beta_k(y^0 - x^*) + (1 - \beta_k)(\mathbf{T}y^k - x^*)\| \\ &\leq \beta_k\|y^0 - x^*\| + (1 - \beta_k)\|\mathbf{T}y^k - x^*\| \\ &\leq \beta_k\|y^0 - x^*\| + (1 - \beta_k)\|y^k - x^*\| \\ &\leq \beta_k\|y^0 - x^*\| + (1 - \beta_k)\|y^0 - x^*\| = \|y^0 - x^*\|. \end{aligned}$$

The first inequality is just triangular inequality, the second inequality comes from the fact $x^* \in \text{Fix}\mathbf{T}$ and \mathbf{T} is nonexpansive, and the last inequality is from induction hypothesis. By induction, we have $\|y^k - x^*\| \leq \|y^0 - x^*\|$ for $k = 0, 1, \dots$

Define $\tilde{\mathbf{A}}x^k = y^{k-1} - x^k$, then $\tilde{\mathbf{A}}x^k \in \mathbf{A}x^k$ since $\mathbf{J}_{\mathbf{A}}y^{k-1} = x^k$. Observe for $k = 1, 2, \dots$, from monotone inequality we have

$$\|y^{k-1} - x^*\|^2 = \|\tilde{\mathbf{A}}x^k + (x^k - x^*)\|^2 = \|\tilde{\mathbf{A}}x^k\|^2 + 2\langle \tilde{\mathbf{A}}x^k, x^k - x^* \rangle + \|x^k - x^*\|^2 \geq \|x^k - x^*\|^2.$$

Therefore we conclude

$$\|x^k - x^*\| \leq \|y^{k-1} - x^*\| \leq \|y^0 - x^*\| = \|x^0 - x^*\|.$$

The latter statement holds directly from triangular inequality,

$$\|x^k - x^0\| \leq \|x^k - x^*\| + \|x^* - x^0\| \leq 2\|x^0 - x^*\|.$$

□

G.3 Proof of Theorem 5.1

The outline of the proofs originate from continuous proofs. To simplify calculations, instead of directly deriving discrete counterpart of Proposition 3.2, we consider rescaled conservation law for each cases. By considering dilated coordinate $W_1 = X - X_0$ for $p > 1$, $W_2 = t^p (X - X_0)$ for $0 < p < 1$, $W_3 = t(X - X_0)$ for $p = 1$, $\gamma \geq 1$ and $W_4 = t^\gamma (X - X_0)$ for $p = 1$, $0 < \gamma < 1$, we obtain below conservation laws.

$$\begin{aligned} E_1 &\equiv \|\tilde{\mathbf{A}}(X(t))\|^2 + \frac{2\gamma}{t^p} \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \left(\frac{\gamma^2}{t^{2p}} + \frac{\gamma p}{t^{p+1}} \right) \|X(t) - X_0\|^2 \\ &\quad + \int_0^t \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds + \int_0^t \frac{\gamma}{s^p} \left\| \dot{X}(s) + \frac{p}{s} (X(s) - X_0) \right\|^2 ds + \frac{1}{2} \int_0^t \frac{\gamma p(p-1)}{s^{p+2}} \|X(s) - X_0\|^2 ds \\ E_2 &\equiv t^{2p} \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\gamma t^p \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + (\gamma^2 - \gamma p t^{p-1}) \|X(t) - X_0\|^2 \\ &\quad + \int_0^t 2s^{2p} \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds + \int_0^t 2s^p (\gamma - p s^{p-1}) \left\| \dot{X}(s) \right\|^2 ds + \int_0^t \gamma p(p-1) s^{p-2} \|X(s) - X_0\|^2 ds \\ E_3 &\equiv t^2 \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\gamma t \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \gamma(\gamma-1) \|X(t) - X_0\|^2 \\ &\quad + \int_{t_0}^t 2s^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds + \int_{t_0}^t 2s(\gamma-1) \left\| \dot{X}(s) \right\|^2 ds \\ E_4 &\equiv t^{2\gamma} \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\gamma t^{2\gamma-1} \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \gamma(\gamma-1) t^{2\gamma-2} \|X(t) - X_0\|^2 \\ &\quad + \int_0^t 2s^{2\gamma} \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds + \int_0^t 2\gamma(\gamma-1) s^{2\gamma-3} \|X(s) - X_0\|^2 ds. \end{aligned}$$

Lyapunov style proof can be obtained by considering below functions

$$\begin{aligned} U_1(t) &= \|\tilde{\mathbf{A}}(X(t))\|^2 + \frac{2\gamma}{t^p} \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \left(\frac{\gamma^2}{t^{2p}} + \frac{\gamma p}{t^{p+1}} \right) \|X(t) - X_0\|^2 \\ U_2(t) &= t^{2p} \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\gamma t^p \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + (\gamma^2 - \gamma p t^{p-1}) \|X(t) - X_0\|^2 \\ U_3(t) &= t^2 \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\gamma t \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \gamma(\gamma-1) \|X(t) - X_0\|^2 \\ U_4(t) &= t^{2\gamma} \|\tilde{\mathbf{A}}(X(t))\|^2 + \gamma t^{2\gamma-1} \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \gamma(\gamma-1) t^{2\gamma-2} \|X(t) - X_0\|^2. \end{aligned}$$

The main blocks of the calculations corresponds to continuous cases, but there are some ‘errors’ in terms of $\tilde{\mathbf{A}}x^k$ and $x^k - x^0$ occur due to discretization. The proofs are done by showing these ‘errors’ don’t effect to the conclusions.

(0) Preparation.

For all cases, we will consider functions of the form

$$U^k = a_k \|\tilde{\mathbf{A}}x^k\|^2 + b_k \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle + c_k \|x^k - x^0\|^2, \quad (41)$$

and consider $U^{k+1} - U^k$. As similar calculations will be repeated, we first organize the repeating calculations. That is, we prove following equality is true.

$$\begin{aligned}
& U^{k+1} - U^k + \lambda_k \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle - \tau_k \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\
&= \left(a_{k+1} - \lambda_k \frac{k^p + \gamma}{k^p} \right) \|\tilde{\mathbf{A}}x^{k+1}\|^2 + \left(\lambda_k \frac{k^p}{k^p + \gamma} - a_k \right) \|\tilde{\mathbf{A}}x^k\|^2 \\
&+ \left(b_{k+1} - \lambda_k \frac{\gamma}{k^p} - \tau_k - c_{k+1} \frac{k^p + \gamma}{k^p} \right) \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle + (\tau_k - c_{k+1}) \langle \tilde{\mathbf{A}}x^{k+1}, x^k - x^0 \rangle \\
&+ (\tau_k - c_{k+1}) \langle \tilde{\mathbf{A}}x^k, x^{k+1} - x^0 \rangle + \left(\lambda_k \frac{\gamma}{k^p + \gamma} - b_k - \tau_k - c_{k+1} \frac{k^p}{k^p + \gamma} \right) \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\
&- c_{k+1} \frac{\gamma}{k^p} \|x^{k+1} - x^0\|^2 - c_{k+1} \frac{\gamma}{k^p + \gamma} \|x^k - x^0\|^2 + (c_{k+1} - c_k) \|x^k - x^0\|^2.
\end{aligned} \tag{42}$$

Observe, this can be shown by checking below equalities are true.

$$\begin{aligned}
& \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\
&= -\frac{k^p + \gamma}{k^p} \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \frac{\gamma}{k^p} \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle + \frac{k^p}{k^p + \gamma} \|\tilde{\mathbf{A}}x^k\|^2 + \frac{\gamma}{k^p + \gamma} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle
\end{aligned} \tag{43}$$

$$\begin{aligned}
& \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\
&= \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle + \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle - (\langle \tilde{\mathbf{A}}x^{k+1}, x^k - x^0 \rangle + \langle \tilde{\mathbf{A}}x^k, x^{k+1} - x^0 \rangle)
\end{aligned} \tag{44}$$

$$\begin{aligned}
& c_{k+1} \|x^{k+1} - x^0\|^2 - c_k \|x^k - x^0\|^2 \\
&= -c_{k+1} \frac{k^p + \gamma}{k^p} \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle - c_{k+1} \langle \tilde{\mathbf{A}}x^{k+1}, x^k - x^0 \rangle \\
&- c_{k+1} \langle \tilde{\mathbf{A}}x^k, x^{k+1} - x^0 \rangle - c_{k+1} \frac{k^p}{k^p + \gamma} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\
&- c_{k+1} \frac{\gamma}{k^p} \|x^{k+1} - x^0\|^2 - c_{k+1} \frac{\gamma}{k^p + \gamma} \|x^k - x^0\|^2 + (c_{k+1} - c_k) \|x^k - x^0\|^2.
\end{aligned} \tag{45}$$

- Proof for (43)

Recall, the method was defined as

$$\begin{aligned}
y^k &= \frac{k^p}{k^p + \gamma} (2x^k - y^{k-1}) + \frac{\gamma}{k^p + \gamma} x^0 \\
x^{k+1} &= \mathbf{J}_A y^k.
\end{aligned}$$

By considering $x^{k+1} + \tilde{\mathbf{A}}x^{k+1} = y^k$, substituting y^k and y^{k-1} we can rewrite the method as

$$x^{k+1} + \tilde{\mathbf{A}}x^{k+1} = \left(1 - \frac{\gamma}{k^p + \gamma} \right) (x^k - \tilde{\mathbf{A}}x^k) + \frac{\gamma}{k^p + \gamma} x^0. \tag{46}$$

By multiplying $k^p + \gamma$ to both sides and reorganizing we have,

$$(k^p + \gamma) \{ (x^{k+1} - x^0) + \tilde{\mathbf{A}}x^{k+1} \} = k^p \{ (x^k - x^0) - \tilde{\mathbf{A}}x^k \}.$$

By subtracting both sides by $k^p(x^{k+1} - x^0)$ and $(k^p + \gamma)(x^k - x^0)$ respectively, above equation can be rewritten as

$$\begin{aligned}
& -k^p(\tilde{\mathbf{A}}x^k + (x^{k+1} - x^k)) = (k^p + \gamma)\tilde{\mathbf{A}}x^{k+1} + \gamma(x^{k+1} - x^0) \\
& -(k^p + \gamma)(\tilde{\mathbf{A}}x^{k+1} + (x^{k+1} - x^k)) = k^p\tilde{\mathbf{A}}x^k + \gamma(x^k - x^0).
\end{aligned}$$

From above, we get the following

$$\begin{aligned}
& (k^p + \gamma)k^p \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\
&= (k^p + \gamma)k^p \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^k \rangle + (k^p + \gamma)k^p \langle \tilde{\mathbf{A}}x^{k+1}, \tilde{\mathbf{A}}x^k \rangle \\
&- (k^p + \gamma)k^p \langle \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle - (k^p + \gamma)k^p \langle \tilde{\mathbf{A}}x^{k+1}, \tilde{\mathbf{A}}x^k \rangle \\
&= -(k^p + \gamma)(\tilde{\mathbf{A}}x^{k+1}, -k^p(\tilde{\mathbf{A}}x^k + (x^{k+1} - x^k))) + k^p(\tilde{\mathbf{A}}x^k, -(k^p + \gamma)(\tilde{\mathbf{A}}x^{k+1} + (x^{k+1} - x^k))) \\
&= -(k^p + \gamma)\langle \tilde{\mathbf{A}}x^{k+1}, (k^p + \gamma)\tilde{\mathbf{A}}x^{k+1} + \gamma(x^{k+1} - x^0) \rangle + k^p\langle \tilde{\mathbf{A}}x^k, k^p\tilde{\mathbf{A}}x^k + \gamma(x^k - x^0) \rangle \\
&= -(k^p + \gamma) \left((k^p + \gamma) \|\tilde{\mathbf{A}}x^{k+1}\|^2 + \gamma \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle \right) + k^p \left(\|\tilde{\mathbf{A}}x^k\|^2 + \gamma \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \right).
\end{aligned}$$

Now dividing both sides by $k^p(k^p + \gamma)$, we get the desired result.

– Proof for (44)

This can be checked by just expanding the inner product of left hand side.

– Proof for (45)

First, observe

$$\begin{aligned}
& c_{k+1} \|x^{k+1} - x^0\|^2 - c_k \|x^k - x^0\|^2 \\
&= c_{k+1} \left(\|x^{k+1} - x^0\|^2 - \|x^k - x^0\|^2 \right) + (c_{k+1} - c_k) \|x^k - x^0\|^2 \\
&= c_{k+1} \langle x^{k+1} - x^k, (x^{k+1} - x^0) + (x^k - x^0) \rangle + (c_{k+1} - c_k) \|x^k - x^0\|^2 \\
&= c_{k+1} \left(\langle x^{k+1} - x^k, x^{k+1} - x^0 \rangle + \langle x^{k+1} - x^k, x^k - x^0 \rangle \right) + (c_{k+1} - c_k) \|x^k - x^0\|^2.
\end{aligned}$$

Reorganizing (40), we get two different expressions for $x^{k+1} - x^k$.

$$\begin{aligned}
x^{k+1} - x^k &= - \left(\tilde{\mathbf{A}}x^{k+1} + \frac{k^p}{k^p + \gamma} \tilde{\mathbf{A}}x^k \right) - \frac{\gamma}{k^p + \gamma} (x^k - x^0) \\
x^{k+1} - x^k &= - \left(\frac{k^p + \gamma}{k^p} \tilde{\mathbf{A}}x^{k+1} + \tilde{\mathbf{A}}x^k \right) - \frac{\gamma}{k^p} (x^{k+1} - x^0).
\end{aligned}$$

Now plugging these to previous equality, we get the desired result.

$$\begin{aligned}
& c_{k+1} \|x^{k+1} - x^0\|^2 - c_k \|x^k - x^0\|^2 \\
&= -c_{k+1} \left\langle \left(\frac{k^p + \gamma}{k^p} \tilde{\mathbf{A}}x^{k+1} + \tilde{\mathbf{A}}x^k + \frac{\gamma}{k^p} (x^{k+1} - x^0) \right), x^{k+1} - x^0 \right\rangle \\
&\quad + \left\langle \tilde{\mathbf{A}}x^{k+1} + \frac{k^p}{k^p + \gamma} \tilde{\mathbf{A}}x^k + \frac{\gamma}{k^p + \gamma} (x^k - x^0), x^k - x^0 \right\rangle + (c_{k+1} - c_k) \|x^k - x^0\|^2 \\
&= -c_{k+1} \frac{k^p + \gamma}{k^p} \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle - c_{k+1} \langle \tilde{\mathbf{A}}x^{k+1}, x^k - x^0 \rangle \\
&\quad - c_{k+1} \langle \tilde{\mathbf{A}}x^k, x^{k+1} - x^0 \rangle - c_{k+1} \frac{k^p}{k^p + \gamma} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\
&\quad - c_{k+1} \frac{\gamma}{k^p} \|x^{k+1} - x^0\|^2 - c_{k+1} \frac{\gamma}{k^p + \gamma} \|x^k - x^0\|^2 + (c_{k+1} - c_k) \|x^k - x^0\|^2
\end{aligned}$$

(i) $\|\tilde{\mathbf{A}}(x^k)\|^2 = \mathcal{O}(1)$ for $p > 0, \gamma > 0$.

Plugging

$$\begin{aligned}
a_k &= 1 + \frac{\gamma}{2k^p}, & b_k &= \frac{\gamma}{k^p}, & c_{k+1} &= \frac{\gamma k^p}{4(k^p + \frac{\gamma}{2})} \left(\frac{\gamma}{k^{2p}} - \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) \\
\lambda_k &= 1 + \frac{\gamma}{2k^p}, & \tau_k &= c_{k+1}
\end{aligned}$$

to (41) and (42), we obtain

$$\begin{aligned}
& U^{k+1} - U^k + \left(1 + \frac{\gamma}{2k^p} - \frac{\gamma k^p}{4(k^p + \frac{\gamma}{2})} \left(\frac{\gamma}{k^{2p}} - \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) \right) \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\
&= - \left(\frac{\gamma(2k^p + \gamma)}{2k^{2p}} + \frac{\gamma}{2} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \frac{\gamma(2k^p + \gamma)}{2k^p(k^p + \gamma)} \|\tilde{\mathbf{A}}x^k\|^2 \\
&\quad - \left(\frac{\gamma^2}{k^{2p}} + \frac{\gamma}{2} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle \\
&\quad - \left(\frac{\gamma^2}{k^p(k^p + \gamma)} + \frac{\gamma k^p}{2(k^p + \gamma)} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\
&\quad - \frac{\gamma^2}{2(2k^p + \gamma)} \left(\frac{\gamma}{k^{2p}} - \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) \|x^{k+1} - x^0\|^2
\end{aligned}$$

$$\begin{aligned}
& - \frac{\gamma^2 k^p}{2(k^p + \gamma)(2k^p + \gamma)} \left(\frac{\gamma}{k^{2p}} - \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) \|x^k - x^0\|^2 + (c_{k+1} - c_k) \|x^k - x^0\|^2 \\
= & - \frac{\gamma(2k^p + \gamma)}{2k^{2p}} \left\| \tilde{\mathbf{A}}x^{k+1} + \frac{1}{2k^p + \gamma} \left(\gamma + \frac{k^{2p}}{2} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) (x^{k+1} - x^0) \right\|^2 \\
& - \frac{\gamma(2k^p + \gamma)}{2k^p(k^p + \gamma)} \left\| \tilde{\mathbf{A}}x^k + \frac{1}{2k^p + \gamma} \left(\gamma + \frac{k^{2p}}{2} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right) (x^k - x^0) \right\|^2 \\
& - \frac{1}{2} \left(\frac{\gamma}{k^p} - \frac{\gamma}{(k+1)^p} \right) \|\tilde{\mathbf{A}}x^{k+1}\|^2 \\
& - \underbrace{\frac{\gamma}{2(2k^p + \gamma)} \left(2\gamma \left(\frac{1}{(k+1)^p} - \frac{1}{k^p} \right) + \frac{k^{2p}}{4} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right)^2 \right)}_{s_{k,1}} \|x^{k+1} - x^0\|^2 \\
& - \underbrace{\frac{\gamma k^p}{2(k^p + \gamma)(2k^p + \gamma)} \left(2\gamma \left(\frac{1}{(k+1)^p} - \frac{1}{k^p} \right) + \frac{k^{2p}}{4} \left(\frac{1}{k^p} \right)^2 - \frac{1}{(k+1)^p} \right)}_{s_{k,0}} \|x^k - x^0\|^2 \\
& + (c_{k+1} - c_k) \|x^k - x^0\|^2.
\end{aligned}$$

The continuous counterpart of above equality is

$$\begin{aligned}
& \dot{U}_1(t) + \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle \\
& = -\frac{\gamma}{t^p} \left\| \tilde{\mathbf{A}}(X(t)) + \left(\frac{\gamma}{t^p} + \frac{p}{t} \right) (X(t) - X_0) \right\|^2 - \frac{1}{2} \frac{\gamma p(p-1)}{t^{p+2}} \|X(t) - X_0\|^2,
\end{aligned}$$

which can be obtained by differentiating and reorganizing the conservation law for E_1 . Note, as $\frac{k^{2p}}{2} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) = \mathcal{O}(k^{2p}) \mathcal{O}\left(\frac{1}{k^{p+1}}\right) = \mathcal{O}\left(\frac{1}{k^{1-p}}\right)$, the order of the term $\frac{1}{2k^p + \gamma} \left(\gamma + \frac{k^{2p}}{2} \left(\frac{1}{k^p} - \frac{1}{(k+1)^p} \right) \right)$ corresponds to $\left(\frac{\gamma}{t^p} + \frac{p}{t} \right)$. Thus we can see the sum of first two terms on the right hand side of the equality for discrete setting corresponds to the first term of the right hand side of the equality for continuous setting.

We can observe that $s_{k,1}, s_{k,0} = \mathcal{O}\left(\frac{1}{k^{2p+1}}\right) + \mathcal{O}\left(\frac{1}{k^{p+2}}\right)$. And since $c_k = \mathcal{O}\left(\frac{1}{k^{2p}}\right) + \mathcal{O}\left(\frac{1}{k^{p+1}}\right)$, we have $c_{k+1} - c_k = \mathcal{O}\left(\frac{1}{k^{2p+1}}\right) + \mathcal{O}\left(\frac{1}{k^{p+2}}\right)$ as well. Therefore the coefficients of $\|x^{k+1} - x^0\|^2$ and $\|x^k - x^0\|^2$ are summable for $p > 0$. From Lemma G.1, we know $\|x^{k+1} - x^0\|^2$ and $\|x^k - x^0\|^2$ are bounded by $4\|x^0 - x^*\|^2$.

The term $\langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle$ on the left hand side is nonnegative from monotonicity of $\tilde{\mathbf{A}}$, and the coefficient is nonnegative as well. The coefficient of $\|\tilde{\mathbf{A}}x^{k+1}\|^2$ in the right hand side, $-\frac{1}{2} \left(\frac{\gamma}{k^p} - \frac{\gamma}{(k+1)^p} \right)$, is nonpositive. As a result, we get below inequality.

$$\begin{aligned}
U^{k+1} & \leq U^k + \frac{1}{2} \left(\frac{\gamma}{k^p} - \frac{\gamma}{(k+1)^p} \right) \|x^{k+1} - x^0\|^2 + (c_{k+1} - c_k) \|x^k - x^0\|^2 \\
& \leq U^1 + 2\|x^0 - x^*\|^2 \sum_{m=1}^{\infty} \left(\left(\frac{\gamma}{k^p} - \frac{\gamma}{(k+1)^p} \right) + 2(c_{k+1} - c_k) \right) = M.
\end{aligned}$$

As done in (8), by monotonicity of \mathbf{A} and Young's inequality

$$\begin{aligned}
M & \geq U^k = a_k \|\tilde{\mathbf{A}}x^k\|^2 + b_k \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle + c_k \|x^k - x^0\|^2 \\
& \geq a_k \|\tilde{\mathbf{A}}x^k\|^2 + b_k \langle \tilde{\mathbf{A}}x^k, x^* - x^0 \rangle \\
& \geq a_k \|\tilde{\mathbf{A}}x^k\|^2 - \frac{1}{2} \left(\|\tilde{\mathbf{A}}x^k\|^2 + b_k^2 \|x^* - x^0\|^2 \right) = \frac{1}{2} \left(1 + \frac{\gamma}{k^p} \right) \|\tilde{\mathbf{A}}x^k\|^2 + \frac{\gamma^2}{2k^{2p}} \|x^* - x^0\|^2.
\end{aligned}$$

Reorganizing, we get the desired result

$$\begin{aligned}\|\tilde{\mathbf{A}}x^k\|^2 &\leq 2\left(1 - \frac{\gamma}{k^p}\right)^{-1} \left(M + \frac{\gamma^2}{2k^{2p}} \|x^* - x^0\|^2\right) \\ &= 2\left(1 + \frac{\gamma}{k^p - \gamma}\right) \left(M + \frac{\gamma^2}{2k^{2p}} \|x^* - x^0\|^2\right) = \mathcal{O}(1).\end{aligned}$$

(ii) $\|\tilde{\mathbf{A}}(x^k)\|^2 = \mathcal{O}\left(\frac{1}{k^{2p}}\right)$ for $0 < p < 1, \gamma > 0$.
Plugging

$$\begin{aligned}a_{k+1} &= k^p \left(k^p + \frac{\gamma}{2}\right), & b_{k+1} &= \gamma k^p, & c_{k+1} &= \frac{k^p \gamma^2}{4\left(k^p + \frac{\gamma}{2}\right)} \\ \lambda_k &= a_{k+1}, & \tau_k &= c_{k+1}\end{aligned}$$

to (41) and (42) with $p = 1$, we obtain

$$\begin{aligned}U^{k+1} - U^k &+ \left(k^p \left(k^p + \frac{\gamma}{2}\right) - \frac{k^p \gamma^2}{4\left(k^p + \frac{\gamma}{2}\right)}\right) \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\ &= -\gamma \left(k^p + \frac{\gamma}{2}\right) \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \gamma^2 \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle - \frac{\gamma^3}{4\left(k^p + \frac{\gamma}{2}\right)} \|x^{k+1} - x^0\|^2 \\ &+ \underbrace{\left(\frac{k^{2p}\left(k^p + \frac{\gamma}{2}\right)}{k^p + \gamma} - (k-1)^p \left((k-1)^p + \frac{\gamma}{2}\right)\right)}_{=q_k} \|\tilde{\mathbf{A}}x^k\|^2 + \underbrace{\frac{\gamma \left(- (k-1)^p k^p + k^{2p} - \gamma(k-1)^p\right)}{k^p + \gamma}}_{=r_k} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\ &- \frac{\gamma^3 k^p}{4\left(k^p + \gamma\right)\left(k^p + \frac{\gamma}{2}\right)} \|x^k - x^0\|^2 + \frac{\gamma^3 \left(k^p - (k-1)^p\right)}{8\left((k-1)^p + \frac{\gamma}{2}\right)\left(k^p + \frac{\gamma}{2}\right)} \|x^k - x^0\|^2 \\ &= -\gamma \left(k^p + \frac{\gamma}{2}\right) \left\| \tilde{\mathbf{A}}x^{k+1} + \frac{\gamma}{2k^p + \gamma} (x^{k+1} - x^0) \right\|^2 + q_k \left\| \tilde{\mathbf{A}}x^k + \frac{r_k}{2q_k} (x^k - x^0) \right\|^2 \\ &- \underbrace{\left(\frac{r_k^2}{4q_k} + \frac{\gamma^3 k^p}{4\left(k^p + \gamma\right)\left(k^p + \frac{\gamma}{2}\right)} + \frac{\gamma^3 \left(k^p - (k-1)^p\right)}{8\left((k-1)^p + \frac{\gamma}{2}\right)\left(k^p + \frac{\gamma}{2}\right)}\right)}_{s_k} \|x^k - x^0\|^2.\end{aligned}$$

The continuous counterpart of this equality is

$$\begin{aligned}\dot{U}_2(t) + 2t^{2p} \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle \\ = -2t^p \left(\gamma - pt^{p-1}\right) \left\| \tilde{\mathbf{A}}(X(t)) + \frac{\gamma}{t^p} (X(t) - X_0) \right\|^2 - \gamma p(p-1)t^{p-2} \|X(t) - X_0\|^2\end{aligned}$$

which can be obtained by differentiating the conservation law for E_2 . The first term on the right hand side correspond to the sum of first two terms in the right hand side of discrete equality. Thus we may expect the order of the coefficients for the terms would match, and we will check the expectation is indeed true.

With some calculation, we can observe

$$s_k = \frac{2\gamma^2(k-1)^p k^{2p} \left((k-1)^p - k^p\right)^2}{4\left((k-1)^p + \frac{\gamma}{2}\right)\left(k^p + \frac{\gamma}{2}\right) d_k},$$

where

$$d_k = 2k^p(k-1)^p \left((k-1)^p + \frac{\gamma}{2}\right) - 2k^{3p} - \gamma k^{2p} + 2\gamma(k-1)^p \left((k-1)^p + \frac{\gamma}{2}\right).$$

By considering Newton expansion and from $0 < p < 1$, we see

$$\begin{aligned}d_k &= 2k^{3p} + \gamma k^{2p} + \mathcal{O}\left(k^{3p-1}\right) - 2k^{3p} - \gamma k^{2p} + 2\gamma k^{2p} + \mathcal{O}\left(k^p\right) \\ &= 2\gamma k^{2p} + \mathcal{O}\left(k^{3p-1}\right) + \mathcal{O}\left(k^p\right) = \mathcal{O}\left(k^{2p}\right).\end{aligned}$$

Thus we can check the leading order of numerator is $p + 2p + (2p - 2) = 5p - 2$, and leading order of the denominator is $p + p + 2p = 4p$. As $5p - 2 - 4p = p - 2$, we have $s_k \in \mathcal{O}(k^{p-2})$, which matches with the continuous counterpart. Therefore $\sum_{k=1}^{\infty} s_k < \infty$.

On the other hand, we see

$$\begin{aligned} q_k &= \frac{k^{2p} \left(k^p + \frac{\gamma}{2}\right)}{k^p + \gamma} - (k-1)^p \left((k-1)^p + \frac{\gamma}{2}\right) \\ &\leq k^{2p} - (k^p - pk^{p-1} + \mathcal{O}(k^{p-2})) \left((k^p - pk^{p-1} + \mathcal{O}(k^{p-2})) + \frac{\gamma}{2}\right) \\ &= -\frac{\gamma}{2}k^p + 2pk^{2p-1} + \mathcal{O}(k^{2p-2}). \end{aligned}$$

Since $p > 2p - 1$, we have $\lim_{k \rightarrow \infty} q_k = -\infty$. Note this matches with the continuous counterpart as well.

Therefore there is $N > 0$ such that for $k > N$, $q_k < 0$. Now for $k > N$ we have

$$U_{k+1} \leq U_k + s_k \|x^k - x^0\|^2 \leq U_N + 4 \left(\sum_{m=N}^{\infty} s_m \right) \|x^0 - x^*\|^2 = M.$$

Thus for $k > N$, by monotonicity of \mathbb{A} and Young's inequality

$$\begin{aligned} M &\geq U^{k+1} = k^p \left(k^p + \frac{\gamma}{2}\right) \|\tilde{\mathbb{A}}x^{k+1}\|^2 + \gamma k^p \langle \tilde{\mathbb{A}}x^{k+1}, x^{k+1} - x^0 \rangle + \frac{k^p \gamma^2}{4 \left(k^p + \frac{\gamma}{2}\right)} \|x^{k+1} - x^0\|^2 \\ &\geq k^p \left(k^p + \frac{\gamma}{2}\right) \|\tilde{\mathbb{A}}x^{k+1}\|^2 + \gamma k^p \langle \tilde{\mathbb{A}}x^{k+1}, x^* - x^0 \rangle \\ &\geq k^p \left(k^p + \frac{\gamma}{2}\right) \|\tilde{\mathbb{A}}x^{k+1}\|^2 - \frac{k^p}{2} \left(\left(k^p + \frac{\gamma}{2}\right) \|\tilde{\mathbb{A}}x^{k+1}\|^2 + \frac{\gamma^2}{k^p + \frac{\gamma}{2}} \|x^* - x^0\|^2 \right) \\ &= \frac{k^p \left(k^p + \frac{\gamma}{2}\right)}{2} \|\tilde{\mathbb{A}}x^{k+1}\|^2 - \frac{k^p \gamma^2}{2k^p + \gamma} \|x^* - x^0\|^2. \end{aligned}$$

Reorganizing, we get the desired result.

$$\|\tilde{\mathbb{A}}x^{k+1}\|^2 \leq \frac{2}{k^p \left(k^p + \frac{\gamma}{2}\right)} \left(M + \frac{\gamma^2}{2} \|x^0 - x^*\|^2 \right) = \mathcal{O}\left(\frac{1}{k^{2p}}\right).$$

(iii) $\|\tilde{\mathbb{A}}(x^k)\|^2 = \mathcal{O}\left(\frac{1}{k^2}\right)$ for $p = 1, \gamma \geq 1$.
Plugging

$$\begin{aligned} a_{k+1} &= k^2, & b_{k+1} &= \gamma \left(k - \frac{1}{2}(\gamma - 1) \right), & c_{k+1} &= \frac{1}{4}\gamma(\gamma - 1) \\ \lambda_k &= k(k+1), & \tau_k &= c_{k+1} \end{aligned}$$

to (41) and (42), we obtain

$$\begin{aligned}
& U^{k+1} - U^k + \left(k(k+1) - \frac{1}{4}\gamma(\gamma-1) \right) \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\
&= -(\gamma-1)(k+1) \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \frac{k^2(\gamma-1)}{k+\gamma} \|\tilde{\mathbf{A}}x^k\|^2 \\
&\quad - \frac{\gamma(\gamma-1)(k+\frac{\gamma}{4})}{k} \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle - \frac{\gamma(\gamma-1)(k-\frac{\gamma}{4})}{k+\gamma} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\
&\quad - \frac{\gamma^2(\gamma-1)}{4k} \|x^{k+1} - x^0\|^2 - \frac{\gamma^2(\gamma-1)}{4(k+\gamma)} \|x^k - x^0\|^2 \\
&= -(\gamma-1)(k+1) \left\| \tilde{\mathbf{A}}x^{k+1} + \frac{\gamma}{2(k+1)} \left(1 + \frac{\gamma}{4k}\right) (x^{k+1} - x^0) \right\|^2 \\
&\quad - \frac{k^2(\gamma-1)}{k+\gamma} \left\| \tilde{\mathbf{A}}x^k + \frac{\gamma}{2k} \left(1 - \frac{\gamma}{4k}\right) (x^k - x^0) \right\|^2 \\
&\quad - \underbrace{\frac{\gamma^2(\gamma-1)}{64k(k+1)} \left(8(\gamma-2) - \frac{\gamma^2}{k}\right) \|x^{k+1} - x^0\|^2}_{=s_{k,1}=\mathcal{O}(\frac{1}{k^2})} - \underbrace{\frac{\gamma^3(\gamma-1)}{64k(k+\gamma)} \left(8 - \frac{\gamma}{k}\right) \|x^k - x^0\|^2}_{=s_{k,0}=\mathcal{O}(\frac{1}{k^2})}.
\end{aligned}$$

The continuous counterpart of above equality is

$$\dot{U}_3(t) + 2t^2 \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle = -2t(\gamma-1) \left\| \tilde{\mathbf{A}}(X(t)) + \frac{\gamma}{t} (X(t) - X_0) \right\|^2$$

which can be obtained by differentiating the conservation law for E_3 . Note the terms match with same order of coefficients, except two $\|x^{k+1} - x^0\|^2$ and $\|x^k - x^0\|^2$, while these terms are summable as $s_{k,1}, s_{k,0} = \mathcal{O}(\frac{1}{k^2})$. Therefore we have

$$\begin{aligned}
U^{k+1} &\leq U^k - s_{k,1} \|x^{k+1} - x^0\|^2 - s_{k,0} \|x^k - x^0\|^2 \\
&\leq U^1 + 4 \|x^0 - x^*\|^2 \sum_{m=1}^{\infty} (s_{m,1} + s_{m,0}) = M.
\end{aligned}$$

Thus by monotonicity of \mathbf{A} and Young's inequality

$$\begin{aligned}
M &\geq U^k = k^2 \|\tilde{\mathbf{A}}x^{k+1}\|^2 + \gamma \left(k - \frac{1}{2}(\gamma-1) \right) \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle + \frac{1}{4}\gamma(\gamma-1) \|x^k - x^0\|^2 \\
&= k^2 \|\tilde{\mathbf{A}}x^{k+1}\|^2 + \gamma \left(k - \frac{1}{2}(\gamma-1) \right) \langle \tilde{\mathbf{A}}x^k, x^* - x^0 \rangle \\
&\geq k^2 \|\tilde{\mathbf{A}}x^k\|^2 - \frac{1}{2} \left(\left(k - \frac{1}{2}(\gamma-1) \right)^2 \|\tilde{\mathbf{A}}x^k\|^2 + \gamma^2 \|x^* - x^0\|^2 \right) \\
&\geq \frac{k^2}{2} \|\tilde{\mathbf{A}}x^k\|^2 - \frac{\gamma^2}{2} \|x^* - x^0\|^2.
\end{aligned}$$

Reorganizing, we get the desired result.

$$\|\tilde{\mathbf{A}}x^k\|^2 \leq \frac{2}{k^2} \left(M + \frac{\gamma^2}{2} \|x^0 - x^*\|^2 \right) = \mathcal{O}\left(\frac{1}{k^2}\right).$$

(iv) $\|\tilde{\mathbf{A}}(x^k)\|^2 = \mathcal{O}(\frac{1}{k^{2\gamma}})$ for $p = 1, 0 < \gamma < 1$.
Plugging

$$\begin{aligned}
a_k &= k^{2\gamma}, & b_k &= \gamma k \left(k - \frac{1}{4} \right)^{2\gamma-2}, & c_{k+1} &= \frac{1}{4}\gamma(\gamma-1)k^{2\gamma-2} \\
\lambda_k &= k^{2\gamma-1}(k+\gamma), & \tau_k &= c_{k+1}
\end{aligned}$$

to (41) and (42) with $p = 1$, we obtain

$$\begin{aligned}
& U^{k+1} - U^k + \left(k^{2\gamma-1}(k+\gamma) + \frac{1}{4}\gamma(1-\gamma)k^{2\gamma-2} \right) \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\
&= \underbrace{\left((k+1)^{2\gamma} - k^{2\gamma-2}(k+\gamma)^2 \right)}_{=q_k} \|\tilde{\mathbf{A}}x^{k+1}\|^2 \\
&+ \underbrace{\left(\gamma(k+1) \left(k + \frac{3}{4} \right)^{2\gamma-2} - \gamma k^{2\gamma-2}(k+\gamma) - \left(1 + \frac{k+\gamma}{k} \right) \frac{1}{4}\gamma(\gamma-1)k^{2\gamma-2} \right)}_{=s_{k,1}} \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle \\
&+ \underbrace{\left(\gamma k^{2\gamma-1} - \gamma k \left(k - \frac{1}{4} \right)^{2\gamma-2} - \left(1 + \frac{k}{k+\gamma} \right) \frac{1}{4}\gamma(\gamma-1)k^{2\gamma-2} \right)}_{=s_{k,0}} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\
&- \frac{1}{4}\gamma^2(\gamma-1)k^{2\gamma-3} \|x^{k+1} - x^0\|^2 - \frac{1}{4}\gamma^2(\gamma-1)k^{2\gamma-3} \frac{1}{1+\frac{\gamma}{k}} \|x^k - x^0\|^2 \\
&+ \frac{1}{4}\gamma(\gamma-1) \left(k^{2\gamma-2} - (k-1)^{2\gamma-2} \right) \|x^k - x^0\|^2
\end{aligned}$$

The continuous counterpart of this equality is

$$\dot{U}_4(t) + 2t^{2\gamma} \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle ds = -2\gamma(\gamma-1)t^{2\gamma-3} \|X(t) - X_0\|^2$$

which can be obtained by differentiating the conservation law for E_4 . Thus we may expect the order of the matching terms are equal, and the terms do not occur in the continuous version do not bother our desired conclusion. We check our expectation is true.

Terms $\|x^{k+1} - x^0\|^2$ and $\|x^k - x^0\|^2$ correspond to $\|X - X_0\|^2$. The coefficients for $\|x^{k+1} - x^0\|^2$ and $\|x^k - x^0\|^2$ are clearly $\mathcal{O}(k^{2\gamma-3})$, which equals the order of continuous counterpart. Since $\gamma < 1$, we know these terms are summable.

Next we observe $q_k \leq 0$. Observe

$$\begin{aligned}
q_k \leq 0 &\iff \frac{(k+1)^{2\gamma}}{k^{2\gamma}} \leq \frac{(k+\gamma)^2}{k^2} \\
&\iff \left(1 + \frac{1}{k} \right)^{2\gamma} \leq \left(1 + \frac{\gamma}{k} \right)^2 \iff \left(1 + \frac{1}{k} \right)^\gamma \leq 1 + \frac{\gamma}{k}.
\end{aligned}$$

To check the last inequality is true, consider $f(x) = x^\gamma$. Since this function is concave for $0 < \gamma < 1$, we see

$$\left(1 + \frac{1}{k} \right)^\gamma = f\left(1 + \frac{1}{k} \right) \leq f(1) + \frac{1}{k}f'(1) = 1 + \frac{\gamma}{k}.$$

Finally we focus on $s_{k,0}$ and $s_{k,1}$. As cross terms don't appear in continuous version, we may expect these terms are 'small', or in mathematical words, summable. From Cauchy-Schwarz inequality we know

$$\begin{aligned}
\langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle &\leq \|\tilde{\mathbf{A}}x^{k+1}\| \|x^{k+1} - x^0\| \\
\langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle &\leq \|\tilde{\mathbf{A}}x^k\| \|x^k - x^0\|.
\end{aligned}$$

Since $\|\tilde{\mathbf{A}}x^{k+1}\|$ and $\|\tilde{\mathbf{A}}x^k\|$ are bounded from the proof for case (i), we know two innerproduct terms are bounded. Thus if we show $\sum_{k=1}^{\infty} |s_{k,0}|, \sum_{k=1}^{\infty} |s_{k,1}| < \infty$, we can conclude U^k is bounded.

Considering Newton expansion, we see

$$\begin{aligned}
& \gamma(k+1) \left(k + \frac{3}{4} \right)^{2\gamma-2} - \gamma k^{2\gamma-2}(k+\gamma) \\
&= \gamma k^{2\gamma-1} + \gamma \left(1 + \frac{3}{2}(\gamma-1) \right) k^{2\gamma-2} + \mathcal{O}(k^{2\gamma-3}) - \gamma k^{2\gamma-1} - \gamma^2 k^{2\gamma-2} \\
&= \frac{1}{2}\gamma(\gamma-1)k^{2\gamma-2} + \mathcal{O}(k^{2\gamma-3})
\end{aligned}$$

Therefore

$$s_{k,1} = \frac{1}{2}\gamma(\gamma-1)k^{2\gamma-2} + \mathcal{O}(k^{2\gamma-3}) - \left(2 + \frac{\gamma}{k}\right) \frac{1}{4}\gamma(\gamma-1)k^{2\gamma-2} = \mathcal{O}(k^{2\gamma-3}).$$

With similar argument

$$\begin{aligned} s_{k,0} &= \gamma k^{2\gamma-1} - \gamma k \left(k - \frac{1}{4}\right)^{2\gamma-2} - \left(1 + \frac{k}{k+\gamma}\right) \frac{1}{4}\gamma(\gamma-1)k^{2\gamma-2} \\ &= \gamma k^{2\gamma-1} - \gamma k \left(k^{2\gamma-2} - (2\gamma-2) \frac{1}{4}k^{2\gamma-3} + \mathcal{O}(k^{2\gamma-4})\right) - \left(2 - \frac{\gamma}{k+\gamma}\right) \frac{1}{4}\gamma(\gamma-1)k^{2\gamma-2} \\ &= \frac{1}{2}\gamma(\gamma-1)k^{2\gamma-2} + \mathcal{O}(k^{2\gamma-3}) - \left(\frac{1}{2}\gamma(\gamma-1)k^{2\gamma-2} + \mathcal{O}(k^{2\gamma-3})\right) = \mathcal{O}(k^{2\gamma-3}) \end{aligned}$$

Therefore, we have

$$\begin{aligned} U^{k+1} &\leq U^k + s_{k,1} \|\tilde{\mathbf{A}}x^{k+1}\| \|x^{k+1} - x^0\| + s_{k,0} \|\tilde{\mathbf{A}}x^k\| \|x^k - x^0\| \\ &\quad - c_{k+1} \frac{\gamma}{k} \|x^{k+1} - x^0\|^2 - c_{k+1} \frac{\gamma}{k+\gamma} \|x^k - x^0\|^2 + (c_{k+1} - c_k) \|x^k - x^0\|^2 \\ &\leq \sum_{m=1}^{\infty} (|s_{m,1}| \|\tilde{\mathbf{A}}x^{k+1}\| \|x^{k+1} - x^0\| + |s_{m,0}| \|\tilde{\mathbf{A}}x^{k+1}\| \|x^{k+1} - x^0\|) \\ &\quad + \sum_{m=1}^{\infty} \left(\left| c_{m+1} \frac{\gamma}{m} \right| \|x^{k+1} - x^0\|^2 + \left(\left| c_{m+1} \frac{\gamma}{m+\gamma} \right| + |c_{k+1} - c_k| \right) \|x^k - x^0\|^2 \right) = M_1. \end{aligned}$$

By Young's inequality

$$\begin{aligned} M_1 + \frac{1}{4}\gamma(1-\gamma)k^{2\gamma-2} \|x^k - x^0\|^2 &\geq k^{2\gamma} \|\tilde{\mathbf{A}}x^k\|^2 + \gamma k \left(k - \frac{1}{4}\right)^{2\gamma-2} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle \\ &\geq k^{2\gamma} \|\tilde{\mathbf{A}}x^k\|^2 + \gamma k \left(k - \frac{1}{4}\right)^{2\gamma-2} \langle \tilde{\mathbf{A}}x^k, x^* - x^0 \rangle \\ &= k^{2\gamma} \|\tilde{\mathbf{A}}x^k\|^2 + \left\langle k^\gamma \tilde{\mathbf{A}}x^k, \gamma k^{1-\gamma} \left(k - \frac{1}{4}\right)^{2\gamma-2} (x^* - x^0) \right\rangle \\ &\geq k^{2\gamma} \|\tilde{\mathbf{A}}x^k\|^2 - \frac{1}{2}k^{2\gamma} \|\tilde{\mathbf{A}}x^k\|^2 - \frac{\gamma^2}{2} \left(k^{1-\gamma} \left(k - \frac{1}{4}\right)^{2\gamma-2}\right)^2 \|x^* - x^0\|^2. \end{aligned}$$

Since $k^{1-\gamma} \left(k - \frac{1}{4}\right)^{2\gamma-2} = \mathcal{O}(k^{\gamma-1})$ and $\gamma < 1$, this terms goes to zero as $k \rightarrow \infty$ thus there is some $M_2 > 0$ such that $\frac{1}{2} \left(\gamma k^{1-\gamma} \left(k - \frac{1}{4}\right)^{2\gamma-2}\right)^2 \|x^k - x^*\|^2 \leq M_2$ for all $k \geq 1$. Reorganizing terms, we obtain the desired result

$$\|\tilde{\mathbf{A}}x^k\|^2 \leq \frac{2}{k^{2\gamma}} \left(M_1 + M_2 + \gamma(1-\gamma)k^{2\gamma-2} \|x^* - x^0\|^2\right) = \mathcal{O}\left(\frac{1}{k^{2\gamma}}\right).$$

H Proof of convergence analysis for strongly monotone \mathbb{A}

H.1 Proof of Proposition 6.1

We take dilated coordinate $W(t) = C(t)(X(t) - X_0)$ as did in the proof of Proposition 3.2. Recall from (30), the second order version of the ODE was written as $0 = \ddot{W} - \beta(t)\dot{W} + C(t)\frac{d}{dt}\tilde{\mathbf{A}}(X(t))$. Now for we multiply $R(t)^2$ in the ODE and obtain

$$0 = R(t)^2 \left(\ddot{W} - \beta(t)\dot{W} + C(t)\frac{d}{dt}\tilde{\mathbf{A}}(X) \right).$$

Now taking inner product with \dot{W} and integrating we have

$$E_1 \equiv \frac{R(t)^2}{2} \|\dot{W}(t)\|^2 - \underbrace{\int_{t_0}^t R(s)^2 \left(\frac{\dot{R}(s)}{R(s)} \|\dot{W}(s)\|^2 \right) ds}_{*} - \int_{t_0}^t \beta(s) R(s)^2 \|\dot{W}(s)\|^2 ds$$

$$+ \int_{t_0}^t C(s) R(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{W}(s) \right\rangle ds.$$

Note the second term which is obtained from integration by parts, would not appear if $R(t) = 1$ as in Proposition 3.2. This is key term to exploit the condition \mathbf{A} is strongly monotone.

Now again with $\dot{W}(t) = C(t) (\dot{X}(t) + \beta(s)(X(t) - X_0))$, we rewrite the last term as

$$\int_{t_0}^t C(s) R(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{W}(s) \right\rangle ds$$

$$= \int_{t_0}^t C(s)^2 R(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds + \int_{t_0}^t \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), C(s) R(s)^2 \beta(s) W(s) \right\rangle ds.$$

Taking integration by parts to the second term we have

$$\int_{t_0}^t \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), C(s) R(s)^2 \beta(s) W(s) \right\rangle ds - [\langle \tilde{\mathbf{A}}(X(W(s), s)), C(s) R(s)^2 \beta(s) W(s) \rangle]_{t_0}^t$$

$$= - \int_{t_0}^t \left\langle \tilde{\mathbf{A}}(X(W, t)), \left(\beta(s)^2 + 2 \frac{\dot{R}(s)}{R(s)} \beta(s) + \dot{\beta}(s) \right) C(s) R(s)^2 W(s) + C(s) R(s)^2 \beta(s) \dot{W}(s) \right\rangle ds$$

$$= \int_{t_0}^t \beta(s) R(s)^2 \|\dot{W}(s)\|^2 ds + \int_{t_0}^t \left(\beta(s)^2 + \underbrace{2 \frac{\dot{R}(s)}{R(s)} \beta(s) + \dot{\beta}(s)}_{*} \right) R(s)^2 \langle \dot{W}(s), W(s) \rangle ds.$$

The fact $C(t) \tilde{\mathbf{A}}(X(W, t)) = -\dot{W}(t)$ is applied to the second equality. Note the fundamental theorem of calculus is valid since $C(s) R(s)^2 \beta(s) W(s)$ is differentiable, $\tilde{\mathbf{A}}(X(W(s), t))$ is Lipschitz continuous in $[t_0, t]$ by Lemma B.4, and so their inner product is absolutely continuous in $[t_0, t]$.

Now consider the second integrand except the term marked with *. From integration by parts we have

$$\int_{t_0}^t \left(\beta(s)^2 + \dot{\beta}(s) \right) R(s)^2 \langle \dot{W}(s), W(s) \rangle ds - \left[\left(\beta(s)^2 + \dot{\beta}(s) \right) R(s)^2 \frac{1}{2} \|W(s)\|^2 \right]_{t_0}^t$$

$$= -\frac{1}{2} \int_{t_0}^t \left(\left(2\beta(s)\dot{\beta}(s) + \ddot{\beta}(s) \right) R(s)^2 + \underbrace{\left(\beta(s)^2 + \dot{\beta}(s) \right) 2R(s)\dot{R}(s)}_{*} \right) \|W(s)\|^2 ds.$$

The integrand except * marked term can be rewritten as

$$\frac{1}{2} \int_{t_0}^t \left(2\beta(s)\dot{\beta}(s)R(s) + \ddot{\beta}(s)R(s) + 2\dot{\beta}(s)\dot{R}(s) \right) R(s) \|W(s)\|^2 ds$$

$$= \frac{1}{2} \int_{t_0}^t C(s)^2 \left(2\beta(s)\dot{\beta}(s)R(s) + \ddot{\beta}(s)R(s) + 2\dot{\beta}(s)\dot{R}(s) \right) R(s) \|X(s) - X_0\|^2 ds$$

$$= \frac{1}{2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 R(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds.$$

Now collecting the terms marked with *, we have

$$- \int_{t_0}^t R(s)^2 \left(\frac{\dot{R}(s)}{R(s)} \|\dot{W}(s)\|^2 \right) ds + \int_{t_0}^t 2 \frac{\dot{R}(s)}{R(s)} \beta(s) R(s)^2 \langle \dot{W}(s), W(s) \rangle ds - \int_{t_0}^t \beta(s)^2 R(s) \dot{R}(s) \|W(s)\|^2 ds$$

$$= - \int_{t_0}^t R(s)^2 \frac{\dot{R}(s)}{R(s)} \|\dot{W}(s) - \beta(s)W(s)\|^2 ds = - \int_{t_0}^t R(s)^2 C(s)^2 \frac{\dot{R}(s)}{R(s)} \|\dot{X}(s)\|^2 ds.$$

Collecting all results we have

$$\begin{aligned}
E_1 &\equiv \frac{R(t)^2}{2} \left\| \dot{W}(t) \right\|^2 + \left[\langle \tilde{\mathbf{A}}(X(W, s)), C(s)R(s)^2\beta(s)W(s) \rangle \right]_{t_0}^t + \left[\left(\beta(s)^2 + \dot{\beta}(s) \right) R(s)^2 \frac{1}{2} \|W(s)\|^2 \right]_{t_0}^t \\
&\quad + \int_{t_0}^t C(s)^2 R(s)^2 \left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle ds - \int_{t_0}^t R(s)^2 C(s)^2 \frac{\dot{R}(s)}{R(s)} \left\| \dot{X}(s) \right\|^2 ds \\
&\quad + \frac{1}{2} \int_{t_0}^t \frac{d}{ds} \left(C(s)^2 R(s)^2 \dot{\beta}(s) \right) \|X(s) - X_0\|^2 ds \\
&= \frac{C(t)^2 R(t)^2}{2} \left(\|\mathbf{A}(X(t))\|^2 + 2\beta(t) \langle \mathbf{A}(X(t)), X(t) - X_0 \rangle + \left(\beta(t)^2 + \dot{\beta}(t) \right) \|X(t) - X_0\|^2 \right) \\
&\quad + \int_{t_0}^t C(s)^2 R(s)^2 \left(\left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle - \frac{\dot{R}(s)}{R(s)} \left\| \dot{X}(s) \right\|^2 \right) ds - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 R(s)^2 \dot{\beta}(s)}{2} \right) \|X(s) - X_0\|^2 ds \\
&\quad - \underbrace{\frac{C(t_0)^2 R(t_0)^2}{2} \left(2\beta(t_0) \langle \mathbf{A}(X(t_0)), X(t_0) - X_0 \rangle + \left(\beta(t_0)^2 + \dot{\beta}(t_0) \right) \|X(t_0) - X_0\|^2 \right)}_{\text{constant}}.
\end{aligned}$$

Renaming $E = E_1 - \text{constant}$, we get the desired result.

H.2 Proof of Theorem 6.2

H.2.1 Proof of the inequality $V(t) \leq V(0)$

The basic structure of the proof is same as Appendix E.3. We do not repeat the whole proof here, instead we check the steps done in Appendix E.3 are also valid for the setup in Theorem 6.2.

- (i) V is nonincreasing for Lipschitz continuous μ -strongly monotone $\tilde{\mathbf{A}}$.

Recall V in Theorem 6.2 was defined as

$$V(t) = \frac{(e^{\mu t} - e^{-\mu t})^2}{2} \left\| \tilde{\mathbf{A}}(X(t)) \right\|^2 + 2\mu(1 - e^{-2\mu t}) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle - 2\mu^2 (1 - e^{-2\mu t}) \|X(t) - X_0\|^2. \quad (47)$$

We first check following equality is true.

$$V(t) = \frac{C(t)^2 R(t)^2}{2} \left(\left\| \tilde{\mathbf{A}}(X(t)) \right\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \left(\beta(t)^2 + \dot{\beta}(t) \right) \|X(t) - X_0\|^2 \right). \quad (48)$$

Recall we're considering (10), $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1}$. As $\frac{d}{dt} \log(1 - e^{-2\mu t}) = \frac{2\mu e^{-2\mu t}}{1 - e^{-2\mu t}} = \frac{2\mu}{e^{2\mu t} - 1} = \beta(t)$, we have

$$C(t) = e^{\int_{-\infty}^t \frac{2\mu}{e^{2\mu s} - 1} ds} = e^{\log(1 - e^{-2\mu t})} = 1 - e^{-2\mu t}.$$

As $\dot{\beta}(t) = -\frac{4\mu^2 e^{2\mu t}}{(e^{2\mu t} - 1)^2}$ and $R(t) = e^{\mu t}$,

$$\begin{aligned}
\frac{C(t)^2 R(t)^2}{2} &= \frac{1}{2} (1 - e^{-2\mu t})^2 e^{2\mu t} = \frac{(e^{\mu t} - e^{-\mu t})^2}{2} \\
C(t)^2 R(t)^2 \beta(t) &= e^{-2\mu t} (e^{2\mu t} - 1)^2 \frac{2\mu}{e^{2\mu t} - 1} = 2\mu (1 - e^{-2\mu t}) \\
\frac{C(t)^2 R(t)^2}{2} \left(\beta(t)^2 + \dot{\beta}(t) \right) &= \frac{(e^{\mu t} - e^{-\mu t})^2}{2} \left(\left(\frac{2\mu}{e^{2\mu t} - 1} \right)^2 - \frac{4\mu^2 e^{2\mu t}}{(e^{2\mu t} - 1)^2} \right) = 2\mu^2 (e^{-2\mu t} - 1).
\end{aligned}$$

This proves the desired equality. Now we show

$$V(t) = E - \int_{t_0}^t C(s)^2 R(s)^2 \left(\left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle - \frac{\dot{R}(s)}{R(s)} \left\| \dot{X}(s) \right\|^2 \right) ds, \quad (49)$$

where E is from Proposition 6.1 which was defined as

$$E = \frac{C(t)^2 R(t)^2}{2} \left(\|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + (\beta(t)^2 + \dot{\beta}(t)) \|X(t) - X_0\|^2 \right) \\ + \int_{t_0}^t C(s)^2 R(s)^2 \left(\left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle - \frac{\dot{R}(s)}{R(s)} \|\dot{X}(s)\|^2 \right) ds - \int_{t_0}^t \frac{d}{ds} \left(\frac{C(s)^2 R(s)^2 \dot{\beta}(s)}{2} \right) \|X(t) - X_0\|^2 ds.$$

From (48) and the definition of E , it is enough to show $\frac{d}{ds} \left(\frac{C(s)^2 R(s)^2 \dot{\beta}(s)}{2} \right) = 0$. Since

$$\frac{C^2(t) R^2(t) \dot{\beta}(t)}{2} = (1 - e^{-2\mu t})^2 e^{2\mu t} \left(-\frac{4\mu^2 e^{2\mu t}}{(e^{2\mu t} - 1)^2} \right) = -4\mu^2,$$

we see $\frac{d}{ds} \left(\frac{C(s)^2 R(s)^2 \dot{\beta}(s)}{2} \right) = 0$.

Now since $\tilde{\mathbf{A}}$ is Lipschitz continuous, we know E is constant from Proposition 6.1. Therefore from (49), for $t > 0$, $|h| < t$ we have

$$V(t+h) - V(t) = \int_t^{t+h} C(s)^2 R(s)^2 \left(\left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle - \frac{\dot{R}(s)}{R(s)} \|\dot{X}(s)\|^2 \right) ds.$$

As $\frac{\dot{R}(s)}{R(s)} = \mu$ and $\tilde{\mathbf{A}}$ is μ -strongly monotone, from (2) we see

$$\left\langle \frac{d}{ds} \tilde{\mathbf{A}}(X(s)), \dot{X}(s) \right\rangle - \frac{\dot{R}(s)}{R(s)} \|\dot{X}(s)\|^2 \geq 0.$$

Therefore $V(t+h) - V(t) \geq 0$ for $h > 0$, we get the desired result.

(ii) Calculation of $V(0)$ for Lipschitz continuous monotone $\tilde{\mathbf{A}}$
Plugging $t = 0$ to (47) we immediately obtain $V(0) = 0$.

(iii) $V(t) \leq 0$ holds for all $t \in [0, \infty)$ and general maximal μ -strongly monotone \mathbf{A}

We check the arguments in Appendix E.3.3 and Appendix E.3.4 are also valid here.

Define S as defined in Lemma E.1. Take $t \in S$, let $T > t$. As checked in Appendix C, the arguments used in the proof Proposition B.11 is also valid for the case $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1}$. This fact provides the required sequence $\{X_{\lambda_n}\}_{n \in \mathbb{N}}$, where X_{λ_n} converges to X uniformly on $[0, T]$, and \dot{X}_{λ_n} converges weakly to \dot{X} in $L^2([0, T], \mathbb{R}^n)$. As in Appendix E.3.3, denote V_λ as V for the solution with \mathbf{A}_λ . Then from (i) we know V_{λ_n} is nonincreasing for all $n \in \mathbb{N}$, we have $\limsup_{n \rightarrow \infty} V_{\lambda_n}(t) \leq \limsup_{n \rightarrow \infty} V_{\lambda_n}(0) = 0$.

Moreover, we can check the extension of $\tilde{\mathbf{A}}$ defined in Lemma E.1 is also valid. From (21) and (22), we know $\|\dot{X}_\lambda(t)\| \leq \frac{e^{\mu T}}{\sqrt{2}} \|m(\mathbf{A}(X_0))\|$ and $\|\mathbf{A}_\lambda(X_\lambda(t))\| \leq \sqrt{e^{2\mu T} + 1} \|m(\mathbf{A}(X_0))\|$ for all $\lambda > 0, t \in [0, T]$.

Therefore we can prove Corollary E.2 for the case $\beta(t) = \frac{2\mu}{e^{2\mu t} - 1}$ with the same proof, replacing $M_{\mathbf{A}}(T)$ by $\sqrt{e^{2\mu T} + 1} \|m(\mathbf{A}(X_0))\|$ and $M_{dot}(T)$ by $\frac{e^{\mu T}}{\sqrt{2}} \|m(\mathbf{A}(X_0))\|$. Thus $\tilde{\mathbf{A}}(X(t))$ is well-defined for $t = 0$, plugging $t = 0$ to (47) we obtain $V(0) = 0$.

Therefore we have

$$\limsup_{n \rightarrow \infty} V_{\lambda_n}(t) \leq \limsup_{n \rightarrow \infty} V_{\lambda_n}(0) = 0 = V(0),$$

it remains to show $V(t) \leq \limsup_{n \rightarrow \infty} V_{\lambda_n}(t)$. Observe, as equality $\dot{X}(t) = -\tilde{\mathbf{A}}(X(t)) - \beta(t)(X(t) - X_0)$ holds since $t \in S$, from (48) we have

$$V(t) = \frac{C(t)^2 R(t)^2}{2} \left(\|\dot{X}(t)\|^2 + \dot{\beta}(t) \|X(t) - X_0\|^2 \right).$$

Therefore it suffices to check Lemma E.4 is valid here with some U_{λ_n} . For some $a > 0$, Define $U_{\lambda_n} : [0, \infty) \rightarrow \mathbb{R}$ as

$$U_{\lambda_n}(t) = \underbrace{\left\| \dot{X}_{\lambda_n}(t) \right\|^2 - \dot{\beta}(t) \|X_{\lambda_n}(t) - X_0\|^2 + \int_a^t \left(\ddot{\beta}(s) - \frac{2\dot{\beta}(s)^2}{\beta(s)} \right) \|X_{\lambda_n}(s) - X_0\|^2 ds}_{f_n(t)}.$$

We proceed similar argument with Lemma B.5. Differentiating $\dot{X}_{\lambda_n}(t) = -\mathbf{A}_{\lambda_n}(X(t)) - \beta(t)(X_{\lambda_n}(t) - X_0)$, we have for almost all $t > 0$

$$\ddot{X}_{\lambda_n}(t) = -\frac{d}{dt}\mathbf{A}_{\lambda_n}(X(t)) - \dot{\beta}(t)(X_{\lambda_n}(t) - X_0) - \beta(t)\dot{X}_{\lambda_n}(t).$$

Therefore for almost all $t > 0$,

$$\begin{aligned} & \dot{U}_{\lambda_n}(t) \\ &= 2 \left\langle \dot{X}_{\lambda_n}(t), \ddot{X}_{\lambda_n}(t) \right\rangle - \dot{\beta}(t) \|X_{\lambda_n}(t) - X_0\|^2 - 2\dot{\beta}(t) \left\langle \dot{X}_{\lambda_n}(t), X_{\lambda_n}(t) - X_0 \right\rangle + \left(\dot{\beta}(t) - \frac{2\dot{\beta}(t)^2}{\beta(t)} \right) \|X_{\lambda_n}(t) - X_0\|^2 \\ &= 2 \left\langle \dot{X}_{\lambda_n}(t), -\frac{d}{dt}\mathbf{A}_{\lambda_n}(X(t)) - \dot{\beta}(t)(X_{\lambda_n}(t) - X_0) - \beta(t)\dot{X}_{\lambda_n}(t) \right\rangle \\ &\quad - 2\dot{\beta}(t) \left\langle \dot{X}_{\lambda_n}(t), X_{\lambda_n}(t) - X_0 \right\rangle - \frac{2\dot{\beta}(t)^2}{\beta(t)} \|X_{\lambda_n}(t) - X_0\|^2 \\ &= -2 \left\langle \dot{X}_{\lambda_n}(t), \frac{d}{dt}\mathbf{A}_{\lambda_n}(X(t)) \right\rangle - 2\beta(t) \left\| \dot{X}_{\lambda_n}(t) + \frac{\dot{\beta}(t)}{\beta(t)}(X_{\lambda_n}(t) - X_0) \right\|^2 \leq 0. \end{aligned}$$

Therefore U_{λ_n} is nonincreasing, we can prove $\|\dot{X}(t)\| \leq \limsup_{n \rightarrow \infty} \|\dot{X}_{\lambda_n}(t)\|$ with same argument in Lemma E.4. Therefore we have $V(t) \leq V(0) = 0$ for $t \in S$. Extending the result to $t \in [0, \infty)$ can be done with the same argument done in Appendix E.3.4.

H.2.2 Proof for convergence rate

Recall

$$V(t) = \frac{(e^{\mu t} - e^{-\mu t})^2}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\mu(1 - e^{-2\mu t}) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle - 2\mu^2(1 - e^{-2\mu t}) \|X(t) - X_0\|^2.$$

Observe

$$\begin{aligned} \frac{2V(t)}{1 - e^{-2\mu t}} &= (e^{2\mu t} - 1) \|\tilde{\mathbf{A}}(X(t))\|^2 + 4\mu \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle - 4\mu^2 \|X(t) - X_0\|^2 \\ &= (e^{\mu t} - 1) \underbrace{\left(\|\tilde{\mathbf{A}}(X(t))\|^2 - 4\mu \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + 4\mu^2 \|X(t) - X_0\|^2 \right)}_{=\|\tilde{\mathbf{A}}(X(t)) - 2\mu(X(t) - X_0)\|^2} \\ &\quad + e^{\mu t} \underbrace{(e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t))\|^2 + e^{\mu t} \left(4\mu \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle - 4\mu^2 \|X(t) - X_0\|^2 \right)}_{=p(t)}. \end{aligned}$$

From the law of cosines, we have $\|X(t) - X_0\|^2 = \|X(t) - X_\star\|^2 - 2 \langle X(t) - X_0, X_0 - X_\star \rangle - \|X_\star - X_0\|^2$. Applying this to $p(t)$ we have

$$\begin{aligned} p(t) &= 4\mu e^{\mu t} (\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_\star \rangle - \langle \tilde{\mathbf{A}}(X(t)), X_0 - X_\star \rangle - \mu \|X(t) - X_0\|^2) \\ &= 4\mu e^{\mu t} (\langle \tilde{\mathbf{A}}(X(t)) + \mu(X(t) - X_\star), X(t) - X_\star \rangle - \langle \tilde{\mathbf{A}}(X(t)) - 2\mu(X(t) - X_0), X_0 - X_\star \rangle - \mu \|X_0 - X_\star\|^2). \end{aligned}$$

Thus

$$\begin{aligned} \frac{2V(t)}{1 - e^{-2\mu t}} &= (e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t)) - 2\mu(X(t) - X_0)\|^2 + e^{\mu t} (e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t))\|^2 + p(t) \\ &= (e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t)) - 2\mu(X(t) - X_0)\|^2 - 4\mu e^{\mu t} \langle \tilde{\mathbf{A}}(X(t)) - 2\mu(X(t) - X_0), X_0 - X_\star \rangle \\ &\quad + e^{\mu t} (e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t))\|^2 + 4\mu e^{\mu t} \langle \tilde{\mathbf{A}}(X(t)) - \mu(X(t) - X_\star), X(t) - X_\star \rangle + 4\mu^2 e^{\mu t} \|X_0 - X_\star\|^2 \\ &= (e^{\mu t} - 1) \left\| \tilde{\mathbf{A}}(X(t)) - 2\mu(X(t) - X_0) - \frac{2\mu e^{\mu t}}{e^{\mu t} - 1} (X_0 - X_\star) \right\|^2 - \frac{4\mu^2 e^{2\mu t}}{e^{\mu t} - 1} \|X_0 - X_\star\|^2 \\ &\quad + e^{\mu t} (e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t))\|^2 + 4\mu e^{\mu t} (\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_\star \rangle - \mu \|X(t) - X_\star\|^2) + 4\mu^2 e^{\mu t} \|X_0 - X_\star\|^2. \end{aligned}$$

Now since $\tilde{\mathbf{A}}$ is μ -strongly monotone, $\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_\star \rangle - \mu \|X(t) - X_\star\|^2 \geq 0$. From previous section we know $0 = V(0) \geq V(t)$ for all $t > 0$. Therefore for all $t > 0$

$$\begin{aligned} 0 &\geq \frac{2V(t)}{1 - e^{-2\mu t}} \geq e^{\mu t}(e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t))\|^2 + \left(4\mu^2 e^{\mu t} - \frac{4\mu^2 e^{2\mu t}}{e^{\mu t} - 1}\right) \|X_0 - X_\star\|^2 \\ &= e^{\mu t}(e^{\mu t} - 1) \|\tilde{\mathbf{A}}(X(t))\|^2 - \frac{4\mu^2 e^{\mu t}}{e^{\mu t} - 1} \|X_0 - X_\star\|^2. \end{aligned}$$

Organizing, we conclude

$$\|\tilde{\mathbf{A}}(X(t))\|^2 \leq 4 \left(\frac{\mu}{e^{\mu t} - 1} \right)^2 \|X_0 - X_\star\|^2.$$

H.2.3 Informal derivation of ODE from the method

Assume $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a continuous monotone operator. In [54], the method OS-PPM is presented as

$$\begin{aligned} x^k &= \mathbf{J}_{h\mathbf{A}} y^{k-1} \\ y^k &= \left(1 - \frac{1}{s_k}\right) \left\{ x^k - \frac{1}{\nu} (y^{k-1} - x^k) \right\} + \frac{1}{s_k} y^0 \end{aligned}$$

where $y^0 = x^0$, $\nu = 1 + 2h\mu$ and $s_k = 1 + \nu^2 + \dots + \nu^{2k} = \frac{\nu^{2k+2} - 1}{\nu^2 - 1}$. Using $y^{k-1} = x^k + h\mathbf{A}x^k$, substituting y^k and y^{k-1} this method can be expressed in a single line,

$$x^{k+1} + h\mathbf{A}x^{k+1} = \left(1 - \frac{1}{s_k}\right) \left(x^k - \frac{h}{\nu} \mathbf{A}x^k \right) + \frac{1}{s_k} x^0.$$

Reorganizing and dividing both sides by h , we have

$$\frac{x^{k+1} - x^k}{h} = -\mathbf{A}x^{k+1} - \left(\frac{1}{\nu} - \frac{1}{\nu s_k} \right) \mathbf{A}x^k - \frac{1}{hs_k} (x^k - x^0).$$

Identifying $x^0 = X_0$, $2hk = t$, $x^k = X(t)$,

$$\frac{X(t+2h) - X(t)}{h} = -\mathbf{A}(X(t+2h)) - \left(\frac{1}{\nu} - \frac{1}{\nu s_k} \right) \mathbf{A}(X(t)) - \frac{1}{hs_k} (X(t) - X_0).$$

Now observe

$$\begin{aligned} \lim_{h \rightarrow 0^+} hs_k &= \lim_{h \rightarrow 0^+} h \frac{\nu^{2k+2} - 1}{\nu^2 - 1} = \lim_{h \rightarrow 0^+} h \frac{(1 + 2h\mu)^{2k+2} - 1}{(1 + 2h\mu)^2 - 1} \\ &= \lim_{h \rightarrow 0^+} \frac{(1 + 2h\mu)^2 (1 + 2h\mu)^{t/h} - 1}{4\mu(1 + h\mu)} = \frac{e^{2\mu t} - 1}{4\mu} \\ \lim_{h \rightarrow 0^+} \frac{1}{\nu} &= \lim_{h \rightarrow 0^+} \frac{1}{1 + 2h\mu} = 1 \\ \lim_{h \rightarrow 0^+} \frac{1}{\nu s_k} &= \lim_{h \rightarrow 0^+} \frac{1}{\nu} \frac{1}{hs_k} h = 1 \times \frac{4\mu}{e^{2\mu t} - 1} \times 0 = 0. \end{aligned}$$

Taking limit $h \rightarrow 0^+$ and organizing,

$$2\dot{X}(t) = -\mathbf{A}(X(t)) - (1 + 0)\mathbf{A}(X(t)) - \frac{4\mu}{e^{2\mu t} - 1} (X(t) - X_0).$$

By diving both sides by 2, we get the desired anchor ODE

$$\dot{X}(t) = -\mathbf{A}(X(t)) - \frac{2\mu}{e^{2\mu t} - 1} (X(t) - X_0).$$

I Proof omitted in Section 7

I.1 Proof for Theorem 7.1

We first check

$$\|\tilde{\mathbf{A}}(X(t))\|^2 \leq 4\beta(t)^2 \|X_0 - X_\star\|^2 = \mathcal{O}(\beta(t)^2),$$

where

$$\begin{aligned} \dot{X}(t) &= -\tilde{\mathbf{A}}(X(t)) - \beta(t)(X(t) - X_0) \\ \beta(t) &= -\frac{\|\tilde{\mathbf{A}}(X(t))\|^2}{2\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle}. \end{aligned}$$

By definition of $\beta(t)$, $\Phi(t)$ defined in (8) becomes zero, i.e.

$$0 \equiv \Phi(t) = \|\tilde{\mathbf{A}}(X(t))\|^2 + 2\beta(t)\langle \tilde{\mathbf{A}}(X(t)), X - X_0 \rangle.$$

Plugging $\Phi(t) = 0$ to inequality (8) we have

$$0 \geq \frac{1}{2} \|\tilde{\mathbf{A}}(X(t))\|^2 - 2\beta(t)^2 \|X_0 - X_\star\|^2.$$

Reorganizing, we get the desired result. Other results need some works, we provide the proof with steps into subsections.

I.1.1 $\beta(t) > 0$ for $t > 0$

Taking inner product with $\tilde{\mathbf{A}}(X)$ to the ODE and applying $\frac{1}{2}\Phi(t) = 0$ we have

$$\langle \dot{X}(t), \tilde{\mathbf{A}}(X(t)) \rangle = -\|\tilde{\mathbf{A}}(X(t))\|^2 - \beta(t)\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle = -\frac{1}{2}\|\tilde{\mathbf{A}}(X(t))\|^2.$$

As \mathbf{A} is assumed to be continuous, taking limit $t \rightarrow 0+$ we have

$$\lim_{t \rightarrow 0+} \langle \dot{X}(t), \tilde{\mathbf{A}}(X(t)) \rangle = \left\langle \lim_{t \rightarrow 0+} \dot{X}(t), \tilde{\mathbf{A}}(X_0) \right\rangle = -\frac{1}{2} \lim_{t \rightarrow 0+} \|\tilde{\mathbf{A}}(X(t))\|^2 = -\frac{1}{2} \|\tilde{\mathbf{A}}(X_0)\|^2.$$

On the other hand, by assumption we have $\lim_{t \rightarrow 0+} \dot{X}(t) = \lim_{t \rightarrow 0+} \frac{X(t) - X_0}{t}$ and $\|\tilde{\mathbf{A}}(X_0)\| \neq 0$, thus

$$\lim_{t \rightarrow 0+} \frac{1}{t} \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle = \left\langle \tilde{\mathbf{A}}(X_0), \lim_{t \rightarrow 0+} \dot{X}(t) \right\rangle = -\frac{1}{2} \|\tilde{\mathbf{A}}(X_0)\|^2 < 0. \quad (50)$$

Therefore there is $\epsilon > 0$ such that $\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle < 0$ for $t \in (0, \epsilon)$, thus for $t \in (0, \epsilon)$ we have $\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle \neq 0$ so $\beta(t)$ is well-defined and satisfies

$$\beta(t) = -\frac{\|\tilde{\mathbf{A}}(X(t))\|^2}{2\langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle} > 0.$$

Observe the denominator of $\beta(t)$ is zero when $\|\tilde{\mathbf{A}}(X(t))\| = 0$. As β is assumed to be well-defined, we have $\|\tilde{\mathbf{A}}(X(t))\| \neq 0$ for all $t > 0$. Since β is continuous as \mathbf{A} and X are continuous, by intermediate value theorem we have $\|\tilde{\mathbf{A}}(X(t))\| > 0$ for all $t > 0$.

I.1.2 Proof for the main statements

We first show following lemma.

Lemma I.1. *Following equality holds for almost every $t > 0$.*

$$\left(\dot{\beta}(t) + \beta(t)^2 \right) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle = \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle. \quad (51)$$

Proof. Since $\tilde{\mathbf{A}}$ is Lipschitz continuous by assumption, by Lemma B.4 we know $\tilde{\mathbf{A}}(X(t))$ is differentiable almost everywhere. Differentiating $\Phi(t)$ we have

$$\begin{aligned}
0 &= \dot{\Phi}(t) \\
&= 2 \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \tilde{\mathbf{A}}(X(t)) \right\rangle + 2\dot{\beta}(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + 2\beta(t) \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \right\rangle + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \rangle \\
&= 2 \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \tilde{\mathbf{A}}(X(t)) + \beta(t)(X(t) - X_0) \right\rangle + 2\dot{\beta}(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle \\
&\quad + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), -\tilde{\mathbf{A}}(X(t)) - \beta(t)(X(t) - X_0) \rangle \\
&= -2 \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle - 2\beta(t) \|\tilde{\mathbf{A}}(X(t))\|^2 + 2(\dot{\beta}(t) - \beta(t)^2) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle \\
&= -2 \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle + 4\beta(t)^2 \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + 2(\dot{\beta}(t) - \beta(t)^2) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle \\
&= -2 \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X(t)), \dot{X}(t) \right\rangle + 2(\dot{\beta}(t) + \beta(t)^2) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle.
\end{aligned}$$

for $t \in (0, \infty)$ almost everywhere. The equalities come from the ODE $\dot{X}(t) = -\tilde{\mathbf{A}}(X(t)) - \beta(t)(X(t) - X_0)$ and the fact $\Phi(t) = 0$. Reorganizing, we have the desired equation (51). \square

Now we show the upper bounds of $\beta(t)$ for each monotone and strongly monotone case. Observe from (51) and the definition of $\beta(t)$, we have for almost all $t \in (0, \infty)$

$$-\frac{\dot{\beta}(t) + \beta(t)^2}{2\beta(t)} \|\tilde{\mathbf{A}}(X(t))\|^2 = \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X), \dot{X} \right\rangle. \quad (52)$$

(i) When $\tilde{\mathbf{A}}$ is monotone.

From (52) and (1) we have for almost all $t \in (0, \infty)$

$$-\frac{\dot{\beta}(t) + \beta(t)^2}{2\beta(t)} \|\tilde{\mathbf{A}}(X(t))\|^2 = \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X), \dot{X} \right\rangle \geq 0.$$

Since $\beta(t) > 0$ we have

$$\dot{\beta}(t) + \beta(t)^2 \leq 0.$$

almost everywhere. Now dividing both sides by $\beta(t)^2$ we have

$$1 \leq -\frac{\dot{\beta}(t)}{\beta(t)^2}$$

holds almost everywhere. Since $-\frac{\dot{\beta}(t)}{\beta(t)^2} = \frac{d}{dt} \left(\frac{1}{\beta(t)} \right)$, integrating above inequality both side from δ to t we have

$$t - \delta \leq \frac{1}{\beta(t)} - \frac{1}{\beta(\delta)} \implies \beta(t) \leq \frac{1}{t - \delta + \frac{1}{\beta(\delta)}}.$$

By the way, from (50) we have

$$\lim_{t \rightarrow 0^+} t\beta(t) = - \lim_{t \rightarrow 0^+} \frac{\|\tilde{\mathbf{A}}(X(t))\|^2}{2 \left\langle \tilde{\mathbf{A}}(X(t)), \frac{X(t) - X_0}{t} \right\rangle} = 1,$$

thus $\lim_{\delta \rightarrow 0^+} \beta(\delta) = \infty$, so $\lim_{\delta \rightarrow 0^+} \frac{1}{\beta(\delta)} = 0$. Therefore taking limit $\delta \rightarrow 0^+$ we have

$$\beta(t) \leq \lim_{\delta \rightarrow 0^+} \frac{1}{t - \delta + \frac{1}{\beta(\delta)}} = \frac{1}{t}.$$

(ii) When $\tilde{\mathbf{A}}$ is μ -strongly monotone.

From (52) and (2) for almost all $t \in (0, \infty)$ we have

$$-\frac{\dot{\beta}(t) + \beta(t)^2}{2\beta(t)} \|\tilde{\mathbf{A}}(X(t))\|^2 = \left\langle \frac{d}{dt} \tilde{\mathbf{A}}(X), \dot{X} \right\rangle \geq \mu \|\dot{X}(t)\|^2. \quad (53)$$

On the other hand, observe

$$\begin{aligned} \|\dot{X}(t)\|^2 &= \|\tilde{\mathbf{A}}(X(t)) + \beta(t)(X(t) - X_0)\|^2 \\ &= \underbrace{\|\tilde{\mathbf{A}}(X(t))\|^2}_{=\Phi(t)=0} + 2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle + \beta(t)^2 \|X(t) - X_0\|^2 = \beta(t)^2 \|X(t) - X_0\|^2. \end{aligned}$$

From Cauchy-Schwarz inequality we see

$$\|\tilde{\mathbf{A}}(X(t))\|^2 = -2\beta(t) \langle \tilde{\mathbf{A}}(X(t)), X(t) - X_0 \rangle \leq 2\beta(t) \|\tilde{\mathbf{A}}(X(t))\| \|X(t) - X_0\|,$$

therefore $\|\tilde{\mathbf{A}}(X(t))\| \leq 2\beta(t) \|X(t) - X_0\|$. Combining above observations, we have for almost all $t > 0$

$$\frac{4 \|\dot{X}(t)\|^2}{\|\tilde{\mathbf{A}}(X(t))\|^2} = \frac{4\beta(t)^2 \|X(t) - X_0\|^2}{\|\tilde{\mathbf{A}}(X(t))\|^2} \geq 1. \quad (54)$$

From (53) and (54), we get an inequality for $\beta(t)$ and $\dot{\beta}(t)$

$$-\frac{\dot{\beta}(t) + \beta(t)^2}{2\beta(t)} = \frac{\dot{\beta}(t)}{2\beta(t)} + \frac{\beta(t)}{2} \geq \frac{\mu}{4}.$$

Moving $\frac{\beta(t)}{2}$ to the right-hand side and reorganizing, we have for almost all $t > 0$

$$1 \leq -\frac{\dot{\beta}(t)}{\beta(t)(\frac{\mu}{2} + \beta(t))} = -\frac{\dot{\beta}(t)}{\frac{\mu}{2}} \left(\frac{1}{\beta(t)} - \frac{1}{\frac{\mu}{2} + \beta(t)} \right) \implies \frac{\mu}{2} \leq -\frac{\dot{\beta}(t)}{\beta(t)} + \frac{\dot{\beta}(t)}{\frac{\mu}{2} + \beta(t)}.$$

Integrating above inequality both sides from δ to t we have

$$\frac{\mu}{2}(t - \delta) \leq \left[-\log \beta(t) + \log \left(\frac{\mu}{2} + \beta(t) \right) \right]_{\delta}^t = \log \frac{\frac{\mu}{2} + \beta(t)}{\beta(t)} - \log \frac{\frac{\mu}{2} + \beta(\delta)}{\beta(\delta)}.$$

As observed in case (i) we know $\lim_{\delta \rightarrow 0+} \beta(\delta) = \infty$, taking limit $\delta \rightarrow 0+$ both sides we have

$$\frac{\mu t}{2} \leq \log \frac{\frac{\mu}{2} + \beta(t)}{\beta(t)} = \log \left(1 + \frac{\mu/2}{\beta(t)} \right) \implies e^{\mu t/2} \leq 1 + \frac{\mu/2}{\beta(t)}.$$

Organizing, we get the desired result.

$$\beta(t) \leq \frac{\mu/2}{e^{\mu t/2} - 1}.$$

I.2 Correspondence between the ODE (11) and the discrete method in Theorem 7.2

Observe, the case $h = 1$ for below method corresponds to the method provided in Theorem 7.2.

$$\begin{aligned} x^k &= \mathbb{J}_{h\mathbf{A}} y^{k-1} \\ \beta_k &= \begin{cases} \frac{\|h\tilde{\mathbf{A}}x^k\|^2}{-\langle h\tilde{\mathbf{A}}x^k, x^k - x^0 \rangle + \|h\tilde{\mathbf{A}}x^k\|^2} & \text{if } \|\tilde{\mathbf{A}}x^k\|^2 \neq 0 \\ 0 & \text{if } \|\tilde{\mathbf{A}}x^k\|^2 = 0 \end{cases} \\ y^k &= (1 - \beta_k)(2x^k - y^{k-1}) + \beta_k x^0. \end{aligned}$$

We now show when \mathbb{A} is continuous and $\|\tilde{\mathbb{A}}x^k\|^2 \neq 0$ for all $k \geq 0$, we obtain the ODE (11) when we take limit $h \rightarrow 0+$. Note when \mathbb{A} is continuous \mathbb{A} equals to $\tilde{\mathbb{A}}$.

From $h\tilde{\mathbb{A}}x^{k+1} + x^{k+1} = y^k$, substituting y^k and y^{k-1} we get a single line expression

$$h\tilde{\mathbb{A}}x^{k+1} + x^{k+1} = (1 - \beta_k)(x^k - h\tilde{\mathbb{A}}x^k) + \beta_k x^0.$$

Reorganizing and dividing both sides by h , we have

$$\frac{x^{k+1} - x^k}{h} = -\tilde{\mathbb{A}}x^{k+1} - (1 - \beta_k)\tilde{\mathbb{A}}x^k - \frac{\beta_k}{h}(x^k - x^0).$$

Identify $x^0 = X_0$, $2hk = t$, and $x^k = X(t)$. Then we see

$$\frac{\beta_k}{h} = \frac{h^2 \|\tilde{\mathbb{A}}x^k\|^2}{-h^2 \langle \tilde{\mathbb{A}}x^k, x^k - x^0 \rangle + h^3 \|\tilde{\mathbb{A}}x^k\|^2} = \frac{\|\tilde{\mathbb{A}}(X(t))\|^2}{-\langle \tilde{\mathbb{A}}(X(t)), X(t) - X_0 \rangle + h \|\tilde{\mathbb{A}}(X(t))\|^2}.$$

Thus $\beta_k = \mathcal{O}(h)$, we have $\lim_{h \rightarrow 0+} \beta_k = 0$. Now taking limit $h \rightarrow 0+$ we have

$$2\dot{X}(t) = -2\tilde{\mathbb{A}}(X(t)) - \frac{\|\tilde{\mathbb{A}}(X(t))\|^2}{\langle \tilde{\mathbb{A}}(X(t)), X(t) - X_0 \rangle} (X(t) - X_0).$$

Dividing both sides by 2, we obtain (11).

I.2.1 Correspondence between convergence rates in Theorem 7.1 and Theorem 7.2

From identification above, we see $\beta(t)$ corresponds to $\frac{\beta_k}{2h}$. Therefore we see with identification $x^0 = X_0$, $x^* = X_*$, $2hk = t$, and $x^k = X(t)$ we have

$$\|h\tilde{\mathbb{A}}(x^{k+1})\|^2 \leq \beta_k^2 \|x^0 - x^*\|^2 = 4h^2 \frac{\beta_k^2}{(2h)^2} \|x^0 - x^*\|^2 \xrightarrow{\text{divide by } h^2, h \rightarrow 0+} \|\tilde{\mathbb{A}}(X(t))\|^2 \leq 4\beta(t)^2 \|X_0 - X_*\|^2.$$

And we can also check that the bound $\beta_k \leq \frac{1}{k+1}$ for the monotone case, is equivalent to $\frac{\beta_k}{2h} \leq \frac{1}{2hk + \mathcal{O}(h)}$ and corresponds to $\beta(t) \leq \frac{1}{t}$ as well.

Now suppose \mathbb{A} is μ -strongly monotone and L -Lipschitz continuous. Then $h\mathbb{A}$ is $h\mu$ -strongly monotone and hL -Lipschitz continuous, we have the following inequality from Theorem 7.2

$$\frac{\beta_k}{2h} \leq \frac{\frac{\mu}{2(1+(hL)^2)}}{\left(1 + \frac{h\mu}{1+(hL)^2}\right)^k - 1 + \frac{h\mu}{1+(hL)^2}}. \quad (55)$$

Recalling the identification $2hk = t$, we see

$$\left(1 + \frac{h\mu}{1+(hL)^2}\right)^k = \left(1 + \frac{h\mu}{1+(hL)^2}\right)^{\frac{t}{2h}} = \left(\left(1 + \frac{h\mu}{1+(hL)^2}\right)^{\frac{1+(hL)^2}{h\mu}}\right)^{\frac{\mu t}{2(1+(hL)^2)}} \xrightarrow{h \rightarrow 0+} e^{\mu t/2}.$$

Therefore identifying $\beta(t) = \frac{\beta_k}{2h}$ and taking limit $h \rightarrow 0+$ to the inequality (55), we have

$$\beta(t) \leq \frac{\mu/2}{e^{\mu t/2} - 1}.$$

I.3 Proof of Theorem 7.2

Recall, the method was defined as

$$\begin{aligned} x^k &= \mathbb{J}_{\mathbb{A}} y^{k-1} \\ \beta_k &= \begin{cases} \frac{\|\tilde{\mathbb{A}}x^k\|^2}{-\langle \tilde{\mathbb{A}}x^k, x^k - x^0 \rangle + \|\tilde{\mathbb{A}}x^k\|^2} & \text{if } \|\tilde{\mathbb{A}}x^k\|^2 \neq 0 \\ 0 & \text{if } \|\tilde{\mathbb{A}}x^k\|^2 = 0 \end{cases} \\ y^k &= (1 - \beta_k)(2x^k - y^{k-1}) + \beta_k x^0. \end{aligned}$$

First, we assume $\tilde{\mathbf{A}}x^k \neq 0$ for all $k \geq 0$. Define

$$\tilde{\Phi}^k = (1 - \beta_k) \|\tilde{\mathbf{A}}x^k\|^2 + \beta_k \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle$$

for $k = 1, 2, \dots$, and

$$\Phi^k = \|\tilde{\mathbf{A}}x^k\|^2 + \beta_{k-1} \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle$$

for $k = 2, 3, \dots$. Note by definition of β_k , we have $\tilde{\Phi}^k = 0$. Our goal is to prove

$$\Phi^{k+1} \leq 0.$$

Then with the same argument with (8) we can conclude $\|\tilde{\mathbf{A}}x^{k+1}\|^2 \leq \beta_k^2 \|x^0 - x^*\|^2$. To do so, we first show following lemma.

Lemma I.2. *For $k \geq 1$, following is true.*

$$(1 - \beta_k) \tilde{\Phi}^k - \Phi^{k+1} = (1 - \beta_k) \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle. \quad (56)$$

Proof. Recall $y^k = x^{k+1} + \tilde{\mathbf{A}}x^{k+1}$. Substituting y^k and y^{k-1} , the method is equivalent to

$$x^{k+1} + \tilde{\mathbf{A}}x^{k+1} = (1 - \beta_k)(x^k - \tilde{\mathbf{A}}x^k) + \beta_k x^0.$$

From above we can get two different expression of $(1 - \beta_k)(x^{k+1} - x^k)$.

$$\begin{aligned} (1 - \beta_k)(x^{k+1} - x^k) &= -(1 - \beta_k)(\tilde{\mathbf{A}}x^{k+1} + \tilde{\mathbf{A}}x^k) - \beta_k (\tilde{\mathbf{A}}x^{k+1} + (x^{k+1} - x^0)) \\ &= -(1 - \beta_k) [(\tilde{\mathbf{A}}x^{k+1} + \tilde{\mathbf{A}}x^k) - \beta_k (\tilde{\mathbf{A}}x^k - (x^k - x^0))]. \end{aligned}$$

With reorganizing, first equality can be obtained by subtracting both sides by $\beta_k(x^{k+1} - x^0)$ and the second equality can be obtained by multiplying both sides by $(1 - \beta_k)$.

From this we have

$$\begin{aligned} &(1 - \beta_k) \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \\ &= \langle \tilde{\mathbf{A}}x^{k+1}, (1 - \beta_k)(x^{k+1} - x^k) \rangle - \langle \tilde{\mathbf{A}}x^k, (1 - \beta_k)(x^{k+1} - x^k) \rangle \\ &= \langle \tilde{\mathbf{A}}x^{k+1}, -(1 - \beta_k)(\tilde{\mathbf{A}}x^{k+1} + \tilde{\mathbf{A}}x^k) - \beta_k \{\tilde{\mathbf{A}}x^{k+1} + (x^{k+1} - x^0)\} \rangle \\ &\quad - (1 - \beta_k) \langle \tilde{\mathbf{A}}x^k, -(\tilde{\mathbf{A}}x^{k+1} + \tilde{\mathbf{A}}x^k) + \beta_k \{\tilde{\mathbf{A}}x^k - (x^k - x^0)\} \rangle \\ &= -(1 - \beta_k) \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \beta_k \langle \tilde{\mathbf{A}}x^{k+1}, \tilde{\mathbf{A}}x^{k+1} + (x^{k+1} - x^0) \rangle \\ &\quad + (1 - \beta_k) (\|\tilde{\mathbf{A}}x^k\|^2 - \beta_k \langle \tilde{\mathbf{A}}x^k, \tilde{\mathbf{A}}x^k - (x^k - x^0) \rangle) \\ &= (1 - \beta_k) ((1 - \beta_k) \|\tilde{\mathbf{A}}x^k\|^2 + \beta_k \langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle) - \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \beta_k \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle \\ &= (1 - \beta_k) \tilde{\Phi}^k - \Phi^{k+1}. \end{aligned}$$

□

Since \mathbf{A} is monotone we have $\langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle \geq 0$, we see that the right-hand side of (56) is greater or equal to 0 if $1 - \beta_k \geq 0$. Thus it remains to show $1 - \beta_k \geq 0$. As the index is quite confusing, we provide it as a lemma to avoid confusion while proceeding the proof.

Lemma I.3. *If $\tilde{\mathbf{A}}x^k \neq 0$ for all $k \geq 1$, then*

$$\langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle < 0, \quad \beta_k \in (0, 1), \quad \Phi^{k+1} \leq 0$$

for all $k \geq 1$.

Proof. Proof by induction. From $x^1 = \mathbf{J}_A y^0 = \mathbf{J}_A x^0$, we have $x^1 + \tilde{\mathbf{A}}x^1 = x^0$ and so $x^1 - x^0 = -\tilde{\mathbf{A}}x^1$. Applying these facts we have

$$\begin{aligned} \langle \tilde{\mathbf{A}}x^1, x^1 - x^0 \rangle &= -\|\tilde{\mathbf{A}}x^1\|^2 < 0, \\ \beta_1 &= \frac{\|\tilde{\mathbf{A}}x^1\|^2}{-\langle \tilde{\mathbf{A}}x^1, x^1 - x^0 \rangle + \|\tilde{\mathbf{A}}x^1\|^2} = \frac{1}{2} \in (0, 1). \end{aligned}$$

As $1 - \beta_1 > 0$, from (56) we have

$$(1 - \beta_1)\tilde{\Phi}^1 - \Phi^2 = (1 - \beta_1)\langle \tilde{\mathbf{A}}x^2 - \tilde{\mathbf{A}}x^1, x^2 - x^1 \rangle \geq 0.$$

As $\tilde{\Phi}^1 = 0$ by definition, we have $\Phi^2 \leq 0$. Therefore the statement is true for $k = 1$.

Now suppose the statements are true for k . By induction hypothesis, we know

$$0 \geq \Phi^{k+1} = \|\tilde{\mathbf{A}}x^{k+1}\|^2 + \beta_k \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle.$$

As $\beta_k > 0$ from induction hypothesis and $\tilde{\mathbf{A}}x^{k+1} \neq 0$ by assumption, reorganizing $\Phi^k \leq 0$ we have

$$\langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle \leq -\frac{\|\tilde{\mathbf{A}}x^{k+1}\|^2}{\beta_k} < 0.$$

And therefore

$$\beta_{k+1} = \frac{\|\tilde{\mathbf{A}}x^{k+1}\|^2}{\underbrace{-\langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^0 \rangle}_{>0} + \|\tilde{\mathbf{A}}x^{k+1}\|^2} \in (0, 1).$$

Since $1 - \beta_{k+1} > 0$, by (56) we have

$$(1 - \beta_{k+1})\tilde{\Phi}^{k+1} - \Phi^{k+2} = (1 - \beta_{k+1})\langle \tilde{\mathbf{A}}x^{k+2} - \tilde{\mathbf{A}}x^{k+1}, x^{k+2} - x^{k+1} \rangle \geq 0.$$

Since $\tilde{\Phi}^{k+1} = 0$ by definition, we have $\Phi^{k+2} \leq 0$. Therefore the statements are true for $k + 1$, by induction, we get the desired result. \square

Suppose $\|\tilde{\mathbf{A}}x^k\|^2 \neq 0$ for all $k \geq 1$. From the lemma we know $\Phi^{k+1} \leq 0$ for $k \geq 1$, so with the same argument of (8) we have for $x^* \in \text{Zer}\mathbf{A}$

$$\begin{aligned} 0 \geq \Phi^{k+1} &= \|\tilde{\mathbf{A}}x^{k+1}\|^2 + \beta_k \langle \tilde{\mathbf{A}}x^{k+1}, x^{k+1} - x^* \rangle - \beta_k \langle \tilde{\mathbf{A}}x^{k+1}, x^0 - x^* \rangle \\ &\geq \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \beta_k \langle \tilde{\mathbf{A}}x^{k+1}, x^0 - x^* \rangle \\ &\geq \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \left(\frac{1}{2} \|\tilde{\mathbf{A}}x^{k+1}\|^2 + \frac{\beta_k^2}{2} \|x^0 - x^*\|^2 \right) = \frac{1}{2} \|\tilde{\mathbf{A}}x^{k+1}\|^2 - \frac{\beta_k^2}{2} \|x^0 - x^*\|^2. \end{aligned}$$

Organizing, we get

$$\|\tilde{\mathbf{A}}x^{k+1}\|^2 \leq \beta_k^2 \|x^0 - x^*\|^2.$$

Now we show the upper bound of β_k . Observe from (56) and the fact $\tilde{\Phi} = 0$, we have

$$\begin{aligned} (1 - \beta_k) \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle &= (1 - \beta_k)\tilde{\Phi}^k - \Phi^{k+1} \\ &= \frac{\beta_k}{\beta_{k+1}} \tilde{\Phi}^{k+1} - \Phi^{k+1} = \left(\frac{\beta_k}{\beta_{k+1}} - \beta_k - 1 \right) \|\tilde{\mathbf{A}}x^{k+1}\|^2. \end{aligned}$$

As $1 - \beta_k \neq 0$ for $k \geq 1$ by Lemma I.3, dividing both sides by $1 - \beta_k$, we get the discrete counterpart of (52),

$$\langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle = \frac{\frac{\beta_k}{\beta_{k+1}} - \beta_k - 1}{1 - \beta_k} \|\tilde{\mathbf{A}}x^{k+1}\|^2. \quad (57)$$

(i) When \mathbf{A} is monotone.

From (57) and monotonicity of \mathbf{A} we have

$$0 \leq \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle = \frac{\frac{\beta_k}{\beta_{k+1}} - \beta_k - 1}{1 - \beta_k} \|\tilde{\mathbf{A}}x^{k+1}\|^2.$$

From Lemma I.3 we have $1 - \beta_k > 0$, therefore

$$\frac{\beta_k}{\beta_{k+1}} - \beta_k - 1 \geq 0 \implies \frac{1}{\beta_k} + 1 \leq \frac{1}{\beta_{k+1}}.$$

Summing up, as $\beta_1 = \frac{1}{2}$ and $\frac{1}{\beta_k} > 0$ we have

$$\frac{1}{\beta_1} + (k - 1) = k + 1 \leq \frac{1}{\beta_k} \implies \beta_k \leq \frac{1}{k + 1}.$$

(ii) When \mathbf{A} is μ -strongly monotone.

Since \mathbf{A} is μ -strongly monotone, from (57) we have

$$\mu \|x^{k+1} - x^k\|^2 \leq \langle \tilde{\mathbf{A}}x^{k+1} - \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle = \frac{\beta_k - \beta_{k+1}}{1 - \beta_k} \|\tilde{\mathbf{A}}x^{k+1}\|^2.$$

Define $r_k = \frac{\|x^{k+1} - x^k\|^2}{\|\tilde{\mathbf{A}}x^{k+1}\|^2}$ for $k = 0, 1, \dots$. Dividing both sides by $\|\tilde{\mathbf{A}}x^{k+1}\|^2$ and organizing, we have

$$r_k \mu \leq \frac{\beta_k - \beta_{k+1}}{1 - \beta_k} \implies r_k \mu (1 - \beta_k) \leq \beta_k - \beta_{k+1} = \beta_k \left(\frac{1}{\beta_{k+1}} - 1 \right) - \beta_k - (1 - \beta_k).$$

Dividing both sides by β_k and reorganizing, we get a recursive inequality for $\frac{1}{\beta_k} - 1$

$$(1 + r_k \mu) \left(\frac{1}{\beta_k} - 1 \right) + 1 \leq \frac{1}{\beta_{k+1}} - 1. \quad (58)$$

We now prove an upper bound of β_k from above inequality.

Lemma I.4. *Suppose \mathbf{A} be a μ -strongly monotone operator. Let β_k be a sequence defined as Theorem 7.2 and let*

$r_k = \frac{\|x^{k+1} - x^k\|^2}{\|\tilde{\mathbf{A}}x^{k+1}\|^2}$ for $k = 0, 1, \dots$. Then following holds for $k = 1, 2, \dots$.

$$\beta_k \leq \frac{1}{\sum_{j=1}^{k-1} \prod_{i=j}^{k-1} (1 + r_i \mu) + 2}.$$

Proof. First observe, the statement is equivalent to

$$\frac{1}{\beta_k} - 1 \geq \sum_{j=1}^{k-1} \prod_{i=j}^{k-1} (1 + r_i \mu) + 1.$$

The proof can be done by induction with (58).

When $k = 1$, recalling $\beta_1 = \frac{1}{2}$ from the proof of Lemma I.3, we can check the inequality is true.

Now suppose the inequality is true for $k = m$. Then from (58) we have

$$\begin{aligned} \frac{1}{\beta_{m+1}} - 1 &\geq (1 + r_m \mu) \left(\frac{1}{\beta_m} - 1 \right) + 1 \\ &\geq (1 + r_m \mu) \left(\sum_{j=1}^{m-1} \prod_{i=j}^{m-1} (1 + r_i \mu) + 1 \right) + 1 \\ &= \left(\sum_{j=1}^{m-1} \prod_{i=j}^m (1 + r_i \mu) + (1 + r_m \mu) \right) + 1 = \sum_{j=1}^m \prod_{i=j}^m (1 + r_i \mu) + 1. \end{aligned}$$

Therefore, we get the desired result. \square

Under the identification considered in Appendix I.2, the continuous counterpart of r_k is

$$r_k = \frac{\|x^{k+1} - x^k\|^2}{\|h\tilde{\mathbf{A}}x^{k+1}\|^2} = 4 \frac{\|x^{k+1} - x^k\|^2}{(2h)^2} \frac{1}{\|\tilde{\mathbf{A}}x^{k+1}\|^2} \xrightarrow{h \rightarrow 0+} \frac{4 \|\dot{X}(t)\|^2}{\|\tilde{\mathbf{A}}(X(t))\|^2},$$

and is greater or equal to 1 by (54). We obtained an exponential convergence rate in continuous setup from this fact. In the same spirit, we can get an exponential convergence rate for discrete setup if there is a positive lower bound for r_k , we provide it as a corollary of Lemma I.4.

Corollary I.5. Consider the setup of Lemma I.4. Suppose there is $r \geq 0$ such that $r_k \geq r$ for $k = 0, 1, \dots$. Then following is true for $k = 1, 2, \dots$

$$\beta_k \leq \frac{r\mu}{(1+r\mu)^k - 1 + r\mu}.$$

Proof. From $r_k \geq r$ we have

$$\begin{aligned} \sum_{j=1}^{k-1} \prod_{i=j}^{k-1} (1+r_i\mu) + 2 &\geq \sum_{j=1}^{k-1} \prod_{i=j}^{k-1} (1+r\mu) + 2 \\ &= \sum_{j=1}^{k-1} (1+r\mu)^{k-j} + 2 = \sum_{l=0}^{k-1} (1+r\mu)^l + 1 = \frac{(1+r\mu)^k - 1 + r\mu}{r\mu}. \end{aligned}$$

Applying Lemma I.4, we get the desired result. \square

We now show $r_k \geq \frac{1}{1+L^2}$ holds when \mathbf{A} is furthermore L -Lipschitz continuous.

(iii) When \mathbf{A} is μ -strongly monotone and L -Lipschitz continuous.

Recall from the proof of Lemma I.2, we know

$$x^{k+1} - x^k = -\mathbf{A}x^{k+1} - (1 - \beta_k)\mathbf{A}x^k - \beta_k(x^k - x^0).$$

Taking inner product with $\tilde{\mathbf{A}}x^k$ both sides we have

$$\langle \tilde{\mathbf{A}}x^k, x^{k+1} - x^k \rangle = -\langle \mathbf{A}x^{k+1}, \tilde{\mathbf{A}}x^k \rangle - \tilde{\Phi}^k = -\langle \mathbf{A}x^{k+1}, \tilde{\mathbf{A}}x^k \rangle.$$

From above equality we can check

$$\|x^{k+1} - x^k\|^2 + \|\mathbf{A}x^{k+1} - \tilde{\mathbf{A}}x^k\|^2 = \|\mathbf{A}x^{k+1}\|^2 + \|x^{k+1} - x^k + \tilde{\mathbf{A}}x^k\|^2 \geq \|\mathbf{A}x^{k+1}\|^2.$$

As \mathbf{A} is L -Lipschitz continuous, we have

$$(1+L^2)\|x^{k+1} - x^k\|^2 \geq \|x^{k+1} - x^k\|^2 + \|\mathbf{A}x^{k+1} - \tilde{\mathbf{A}}x^k\|^2 \geq \|\mathbf{A}x^{k+1}\|^2.$$

Dividing both sides by $(1+L^2)\|\mathbf{A}x^{k+1}\|^2$ we get a lowerbound for r_k

$$r_k = \frac{\|x^{k+1} - x^k\|^2}{\|\mathbf{A}x^{k+1}\|^2} \geq \frac{1}{1+L^2}.$$

Applying Corollary I.5 we get the desired result

$$\beta_k \leq \frac{\mu/(1+L^2)}{(1+\mu/(1+L^2))^k - 1 + \mu/(1+L^2)}.$$

Note if we take limit $\mu \rightarrow 0^+$, with substitution $\alpha = \frac{\mu}{1+L^2}$ we have

$$\lim_{\mu \rightarrow 0^+} \frac{\mu/(1+L^2)}{(1+\mu/(1+L^2))^k - 1 + \mu/(1+L^2)} = \frac{1}{\lim_{\alpha \rightarrow 0^+} \frac{1}{\alpha} \left((1+\alpha)^k - 1 \right) + 1} = \frac{1}{k+1},$$

which is the bound for the monotone case.

I.3.1 If there is $k \geq 0$ such that $\tilde{\mathbf{A}}x^k = 0$

Suppose $\tilde{\mathbf{A}}x^k = 0$ for some k . Let x^N be the very first iterate such that $\tilde{\mathbf{A}}x^N = 0$. Then from previous argument we know Theorem 7.2 is true for $k < N$. Thus it remains to show the statements are true for $k \geq N$.

From $\tilde{\mathbf{A}}x^N = 0$ we know $y^{N-1} = x^N + \tilde{\mathbf{A}}x^N = x^N$. And since $\tilde{\mathbf{A}}x^N = 0$ implies $\beta_N = 0$, from the definition of the method we have $y^N = 2x^N - y^{N-1} = x^N$. Therefore

$$x^{N+1} = \mathbf{J}_{\mathbf{A}}y^N = \mathbf{J}_{\mathbf{A}}x^N = x^N,$$

we conclude $x^k = x^N \in \text{Zer}\mathbf{A}$ for all $k \geq N$. Thus $\beta_k = 0$, $\|\tilde{\mathbf{A}}x^k\| = 0$ for all $k \geq N$, the Theorem 7.2 is trivially true for $k \geq N$.

J Details of experiment in Section 7.1

We solve a compressed sensing problem of Shi et al. [63] which is formulated as an ℓ_1 -regularized least-squared problem

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{2} \|A_{(i)}x - b_i\|^2 + \rho \|x\|_1 \right\}.$$

We solve this problem in decentralized manner due to the problem setup where the network of local agents are as Figure 2 and each agents communicate only with their neighbors, the nodes connected to each agents by edge. We use Metropolis-Hastings matrix as our mixing matrix $W \in \mathbb{R}^{n \times n}$ and apply PG-EXTRA. Let $W_{i,j}$ denote (i, j) -th entry of W and $N_i \subseteq \{1, 2, \dots, n\}$ denote the index of the agents in the neighborhood of agent i . Consider

$$\begin{aligned} \mathbf{x}_i^{k+1} &= \text{Prox}_{\alpha\rho\|\cdot\|_1} \left(\sum_{j \in N_i} W_{i,j} \mathbf{x}_j^k - \alpha A_{(i)}^\top (A_{(i)} \mathbf{x}_i^k - b_{(i)}) - \mathbf{w}_i^k \right) \\ \mathbf{w}_i^{k+1} &= \mathbf{w}_i^k + \frac{1}{2} \left(x_i^k - \sum_{j \in N_i} W_{i,j} x_j^k \right), \quad k = 0, 1, \dots \end{aligned} \quad (\text{PG-EXTRA})$$

to the problem above. Under suitable choice of parameters, PG-EXTRA can be seen as a fixed-point iteration of an averaged operator with respect to $\|\cdot\|_M$ [74, Theorem 2], where the metric matrix M is defined as

$$M = \begin{bmatrix} (1/\alpha)I & U^\top \\ U & \alpha I \end{bmatrix},$$

and U is a symmetric definite matrix with $U^2 = \frac{1}{2}(I - W)$. That is, denoting $\mathbf{x}^k, \mathbf{w}^k \in \mathbb{R}^{d \times n}$ as the vertical stack of \mathbf{x}_i^k 's and \mathbf{w}_i^k 's respectively [58, Chapter 11.3], PG-EXTRA can be rewritten as $(\mathbf{x}^{k+1}, \mathbf{w}^{k+1}) = \mathbf{T}(\mathbf{x}^k, \mathbf{w}^k)$. Using this \mathbf{T} , we proceed the experiment with the Halpern method

$$(\mathbf{x}^{k+1}, \mathbf{w}^{k+1}) = \beta_k (\mathbf{x}^0, \mathbf{w}^0) + (1 - \beta_k) \mathbf{T}(\mathbf{x}^k, \mathbf{w}^k).$$

When \mathbf{T} is an averaged operator, it is nonexpansive, we know $\mathbf{T} = \mathbf{R}_\mathbf{A} = 2\mathbf{J}_\mathbf{A} - \mathbf{I}$ for some maximal monotone operator \mathbf{A} [54, Lemma 2.1]. Considering the equivalence discussed in Lemma D.2, we see above Halpern method is equivalent to our presented algorithms of the form

$$\begin{aligned} x^{k+1} &= \mathbf{J}_\mathbf{A} y^k \\ y^{k+1} &= (1 - \beta_k) (2x^{k+1} - y^k) + \beta_k x^0, \end{aligned}$$

by corresponding $y^k = (\mathbf{x}^{k+1}, \mathbf{w}^{k+1})$. Note the operator norm $\|\tilde{\mathbf{A}}x^k\|_M^2$ can be calculated by considering below equation

$$\frac{1}{2} (\mathbf{T}y^{k-1} - y^{k-1}) = \mathbf{J}_\mathbf{A} y^{k-1} - y^{k-1} = x^k - (\tilde{\mathbf{A}}x^k - x^k) = -\tilde{\mathbf{A}}x^k.$$

We use the anchor coefficients $\beta_k = \frac{1}{k+1}$, $\beta_k = \frac{\gamma}{k^p + \gamma}$ with $p = 1.5$, $\gamma = 2.0$ and the adaptive choice of β_k in Theorem 7.2 with M -norm. Note, in the experiment the adaptive coefficient is calculated by considering below equation

$$\begin{aligned} \frac{1}{2} \frac{\|\mathbf{T}y^{k-1} - y^{k-1}\|_M^2}{\|\mathbf{T}y^{k-1} - y^{k-1}\|_M^2 + \langle \mathbf{T}y^{k-1} - y^{k-1}, y^{k-1} - x^0 \rangle_M} &= \frac{1}{2} \frac{4 \|\tilde{\mathbf{A}}x^k\|_M^2}{4 \|\tilde{\mathbf{A}}x^k\|_M^2 + \langle -2\tilde{\mathbf{A}}x^k, \tilde{\mathbf{A}}x^k + x^k - x^0 \rangle_M} \\ &= \frac{\|\tilde{\mathbf{A}}x^k\|_M^2}{-\langle \tilde{\mathbf{A}}x^k, x^k - x^0 \rangle_M + \|\tilde{\mathbf{A}}x^k\|_M^2}. \end{aligned}$$

We choose the dimension of signal $d = 100$, the number of agents $n = 20$, the number of measurement for each agent $m_i = 4$, ℓ_1 -regularization parameter $\rho = 0.01$, and algorithm parameter $\alpha = 0.01$.

K Broader Impacts

Our work focuses on the theoretical aspects of convex optimization algorithms. There are no negative social impacts that we anticipate from our theoretical results.

L Limitations

Our analysis concerns convex optimization. Although this assumption is standard in optimization theory, many functions that arise in machine learning practice are not convex.