J. Korean Math. Soc. **61** (2024), No. 2, pp. 395–408
https://doi.org/10.4134/JKMS.j230284
pISSN: 0304-9914 / eISSN: 2234-3008

SHADOWING PROPERTY FOR ADMM FLOWS

Yoon Mo Jung, Bomi Shin, and Sangwoon Yun

Abstract. There have been numerous studies on the characteristics of the solutions of ordinary differential equations for optimization methods, including gradient descent methods and alternating direction methods of multipliers. To investigate computer simulation of ODE solutions, we need to trace pseudo-orbits by real orbits and it is called shadowing property in dynamics. In this paper, we demonstrate that the flow induced by the alternating direction methods of multipliers (ADMM) for a C^2 strongly convex objective function has the eventual shadowing property. For the converse, we partially answer that convexity with the eventual shadowing property guarantees a unique minimizer. In contrast, we show that the flow generated by a second-order ODE, which is related to the accelerated version of ADMM, does not have the eventual shadowing property.

1. Introduction

Gradient-based optimization methods are often interpreted by ordinary differential equations (ODEs), since ODEs can provide insights into the dynamics of the method. However, computer simulations used to track the ODE's solutions do not yield an exact trajectory but a pseudo-trajectory. If the pseudo-trajectory is close to the real trajectory, then we can confidently say that the method is well represented by the ODE. In the literature of dynamical systems, this phenomenon is called the shadowing property, which plays an important role in the study of stability. Roughly speaking, it allows us to trace a set of points which stay near a true orbit, even when perturbed by error or noise. This phenomenon occurs in various applications, such as computer simulations and physics experiments.

Received May 31, 2023; Revised September 6, 2023; Accepted November 11, 2023.

2020 *Mathematics Subject Classification.* Primary 37N40; Secondary 90C31, 90C25.

Key words and phrases. Shadowing property, sensitivity, asymptotic behavior, long time behavior, ADMM, optimization flow, optimization methods, convex programming, proximal method.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2016R1A5A1008055). The work of Y. M. Jung was supported by NRF of Korea (No. 2022R1A2C1010537). The work of B. Shin was supported by NRF of Korea (No. 2021R1C1C2005241). The work of S. Yun was supported by NRF of Korea (No. 2022R1A2C1011503).

We investigate an ODE in $\mathbb{R}^n$ that appears as a continuous limit of the alternating direction methods of multipliers (ADMM) [6]. It is given as follows:

$$\begin{cases} \dot{X} = -\lambda \nabla V(X), \\ X(0) = x_0, \end{cases} \tag{1.1}$$

where λ is an invertible $n \times n$ matrix and $V : \mathbb{R}^n \to \mathbb{R}$ represents a differentiable objective function of the corresponding optimization problem. To specify the induced flow, we call *the ADMM flow for the objective function* V *and the matrix* λ. Owing to its significant applications in machine learning and related topics, the dynamics of the flow induced by ADMM have been recently studied. For instance, the direct method of Lyapunov was employed to study the stability properties of the flow [6]. Moreover, the nonsmooth ADMM cases have also been analyzed, where the objective function is not differentiable [7, 13].

In what follows, we briefly describe how to obtain (1.1). First, we consider the optimization problem

$$\min_{x \in \mathbb{R}^n} V(x) := g_1(x) + g_2(Ax) \tag{1.2}$$

under the following assumptions:

(H1) $g_1 : \mathbb{R}^n \to \mathbb{R}$ and $g_2 : \mathbb{R}^m \to \mathbb{R}$ are convex and smooth.
(H2) $A \in M_{m \times n}(\mathbb{R})$ has full column rank.

The alternating direction method of multipliers (ADMM) to solve (1.2) is given as follows [4]:

$$x^{k+1} = \operatorname*{arg\,min}_{x \in \mathbb{R}^n} g_1(x) + \frac{\rho}{2}\|Ax - z^k + u^k\|^2, \tag{1.3a}$$

$$z^{k+1} = \operatorname*{arg\,min}_{z \in \mathbb{R}^m} g_2(z) + \frac{\rho}{2}\|Ax^{k+1} - z + u^k\|^2, \tag{1.3b}$$

$$u^{k+1} = u^k + Ax^{k+1} - z^{k+1}, \tag{1.3c}$$

where $\rho > 0$ is a penalty parameter. The detailed derivation of the ADMM flow (1.1) is given in Appendix.

In this paper, we investigate the eventual shadowing property of the ADMM flow generated by (1.1). In Section 2, we present our results on the ADMM flow. In Section 3, we show that the ADMM flow for a C^2 strongly convex objective function V and a positive definite matrix λ has the eventual shadowing property. Furthermore, if the ADMM flow for a convex objective function V and a positive definite matrix λ has the eventual shadowing property, then the objective function has a unique minimizer. We note that related results can be found in [3], [11] or [17], under C^2 assumption on the objective functions. In Section 4, we examine the dynamical system generated by a second order ODE which can be used to model the accelerated version of ADMM. We prove that a flow associated to this accelerated model does not have the eventual shadowing property.

2. Preliminary definitions and statements

In a metric space M, a *flow* of M is a continuous map $\varphi : \mathbb{R} \times M \to M$ such that $\varphi(0, x) = x$ and $\varphi(s+t, x) = \varphi(s, \varphi(t, x))$ for every $x \in M$ and $s, t \in \mathbb{R}$. As usual, we denote the time t-map as $\varphi_t : M \to M$, defined by $\varphi_t(x) = \varphi(t, x)$. We say that $\sigma \in X$ is a *fixed point* of φ if $\varphi_t(\sigma) = \sigma$ for every $t \in \mathbb{R}$. A typical example of flow is the solution $\varphi_t(x) = X(t, x)$ of the ODE $\dot{X} = F(X)$, where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a C^1 map. Thus, the equation (1.1) associated with ADMM generates a flow. Note that there is a equivalence relation between the zeroes of F and the fixed points of the flow, specifically,

$$F(\sigma) = 0 \iff \sigma \text{ is a fixed point of } \varphi. \tag{2.1}$$

Now, we introduce the notions of shadowing we will study. Consider a continuous map $f : M \to M$, and let $\Delta > 0$, $\epsilon > 0$. We consider the iteration f^n for $n \in \mathbb{N} \cup \{0\}$ defined by $f^0 = id_M$ (the identity of M) and $f^n = f \circ \overset{n}{\cdots} \circ f$ for $n \in \mathbb{N}$. A sequence $(x_n)_{n \geq 0}$ is called a Δ*-pseudo-orbit* if $d(f(x_n), x_{n+1}) \leq \Delta$ for every $n \geq 0$. The sequence is said to be ϵ*-shadowed* if there is an $x \in M$ such that $d(f^n(x), x_n) \leq \epsilon$ for every $n \geq 0$. The sequence is said to be *eventually* ϵ*-shadowed* if there are x and $N \in \mathbb{N}$ such that $d(f^n(x), x_n) \leq \Delta$, $\forall n \geq N$.

Definition 2.1. We say that a continuous map $f : M \to M$ of a metric space M has the *shadowing* (resp. *eventual shadowing*) *property* [9] if for every $\epsilon > 0$ there is $\Delta > 0$ such that every Δ-pseudo orbit can be ϵ-shadowed (resp. eventually ϵ-shadowed).

It is well known that the analysis of shadowing on flows becomes more complicated than that for maps or homeomorphisms due to the presence of the reparametrization of system (see [15] for details). The following definition extends Definition 2.1 to flow. Note that the definition of shadowing for flows is equivalent to the usual one if the flow has no fixed points (see Lemma 3.2 in [15]).

Definition 2.2. A flow $\varphi : \mathbb{R} \times M \to M$ has the *shadowing* (resp. *eventual shadowing*) *property* if its time t-map $\varphi_t(x) = \varphi(t, x)$ has the shadowing (resp. eventual shadowing) property for some $t > 0$.

For optimization problems, we remind a few definitions. We say that V is *convex* if

$$V(sx + (1-s)y) \leq sV(x) + (1-s)V(y), \quad \forall x, y \in \mathbb{R}^n, \ \ \forall 0 \leq s \leq 1$$

and *strongly convex* if $V - \sigma\| \cdot \|^2$ is convex for some $\sigma > 0$. If V is C^2, then V is strongly convex if and only if there is $\mu > 0$ such that

$$\nabla^2 V(x) \succeq \mu I.$$

Here, $\nabla^2 V(x) \succeq \mu I$ means $\nabla^2 V(x) - \mu I$ is a positive semidefinite matrix.

Our first result is stated as follows.

Theorem 2.3. *The ADMM flow for a C^2 strongly convex map $V : \mathbb{R}^n \to \mathbb{R}$ and a positive definite $n \times n$ matrix λ has the eventual shadowing property.*

We may question whether the converse is also true; specifically, the eventual shadowing property of the ADMM flow implies strong convexity of the corresponding objective function. While we do not have a definitive answer to this question, we have a partial result that shows the eventual shadowing property guarantees a unique minimizer. This characteristic distinguishes strong convexity from convexity.

Theorem 2.4. *Let $V : \mathbb{R}^n \to \mathbb{R}$ be a convex function. If the ADMM flow for V and a positive definite $n \times n$ matrix λ has the eventual shadowing property, then V has a unique minimizer.*

3. Proof of the theorems

Theorem 2.3 is a consequence of the following two lemmas. To delve into them, we adopt a few notations and definitions. For a symmetric positive semidefinite matrix Q, we denote its minimal eigenvalue by $\gamma(Q)$. The Hessian of a C^2 map $F : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ is denoted by $\nabla^2 F(x)$. The *mininorm* $m(L)$ of a linear operator L is defined by

$$m(L) = \inf_{\|x\|=1} \|L(x)\|.$$

Definition 3.1. A flow φ of a metric space M is *point contracting* if there are $x_* \in M$ and $K, \rho > 0$ such that

$$d(\varphi_t(x), x_*) \leq K e^{-\rho t} d(x, x_*), \quad \forall x \in M, t \geq 0.$$

In the literature, a flow φ of a metric space X is said to be *contracting* when there is $\rho > 0$ such that $d(\varphi_t(x), \varphi_t(y)) \leq e^{-\rho t} d(x, y)$ for every $t \geq 0$, $x, y \in X$ (see [8, 10]). It can be proved that all such flows have a singular point $x_* \in X$ (i.e., $\varphi_t(x_*) = x_*$ for all $t \in \mathbb{R}$) if X is complete. Hence, every contracting flow of a complete metric space is point contracting.

We also utilize the *Gronwall inequality* and state it below for the reader's convenience (see [18]).

Gronwall inequality: Let I denote an interval of the form $[a, \infty)$ or $[a, b]$, with $a < b$. Let β and u be real-valued continuous functions defined on I. If u is differentiable in the interior I^o of I and satisfies the differential inequality

$$\frac{du}{dt}(t) \leq \beta(t) u(t), \quad \forall t \in I^o,$$

then

$$u(t) \leq e^{\int_a^t \beta(s) ds} u(a), \quad \forall t \in I.$$

With these notations and facts, we have our first lemma.

Lemma 3.2. *Let $V : \mathbb{R}^n \to \mathbb{R}$ be C^2 strongly convex, and let λ be a positive definite matrix. Then, the ADMM flow for V and λ is contracting.*

Proof. Since λ is positive definite and thus invertible, $\lambda = A^\top A$ for some invertible matrix A (cf. p. 2 in [2]). Since V is strongly convex, V has a unique minimizer X_*.

Now, let X be a solution of (1.1). Then,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|(A^\top)^{-1}(X-X_*)\|^2 &= \langle (A^\top)^{-1}\dot{X}, (A^\top)^{-1}(X-X_*)\rangle \\ &= \langle A^{-1}(A^\top)^{-1}\dot{X}, X-X_*\rangle \\ &= \langle (A^\top A)^{-1}\dot{X}, X-X_*\rangle \\ &= \langle \lambda^{-1}\dot{X}, X-X_*\rangle \\ &= -\langle \nabla V(X), X-X_*\rangle.\end{aligned}$$

In summary,

$$\frac{1}{2}\frac{d}{dt}\|(A^\top)^{-1}(X_*-X)\|^2 = \langle \nabla V(X), X_*-X\rangle. \tag{3.1}$$

On the other hand, by Taylor expansion, one has

$$V(X_*)-V(X) = \langle \nabla V(X), X_*-X\rangle + \frac{1}{2}(X_*-X)^\top \nabla^2 V(U)(X_*-X),$$

where $U = sX_* + (1-s)X$ for some $s \in [0,1]$. Then,

$$\begin{aligned}-\langle \nabla V(X), X_*-X\rangle &= V(X)-V(X_*) + \frac{1}{2}(X_*-X)^\top \nabla^2 V(U)(X_*-X) \\ &\geq \frac{1}{2}(X_*-X)^\top \nabla^2 V(U)(X_*-X) \\ &\geq \frac{\mu}{2}\|X_*-X\|^2 \\ &\geq \frac{\mu}{2}(m(A^\top))^2\|(A^\top)^{-1}(X_*-X)\|^2 \\ &= \rho\|(A^\top)^{-1}(X_*-X)\|^2.\end{aligned}$$

The second inequality comes from the strong convexity with the positive scalar $\mu \leq \inf_{U\in\mathbb{R}^n}\gamma(\nabla V(U))$ and $\rho := \frac{\mu}{2}(m(A^\top))^2$. Hence,

$$\langle \nabla V(X), X_*-X\rangle \leq -\rho\|(A^\top)^{-1}(X_*-X)\|^2. \tag{3.2}$$

By combining (3.1) with (3.2), we obtain

$$\frac{d}{dt}\|(A^\top)^{-1}(X_*-X)\|^2 \leq -2\rho\|(A^\top)^{-1}(X_*-X)\|^2.$$

By letting $I = [0,\infty)$, $u(t) = \|(A^\top)^{-1}(X_*-X)\|^2$, $\beta(t) = -2\rho$ for all t, we rewrite

$$u'(t) \leq \beta(t)u(t), \quad \forall t \in I^o.$$

By applying Gronwall inequality,

$$\|(A^\top)^{-1}(X_*-X)\| \leq e^{-2\rho t}\|(A^\top)^{-1}(X(0)-X_*)\|, \quad \forall t \geq 0.$$

Therefore,

$$\|X - X_*\| \le Ke^{-2\rho t}\|X(0) - X_*\|, \quad \forall t \ge 0,$$

where

$$K = (m((A^\top)^{-1}))^{-1}\|(A^\top)^{-1}\|.$$

From the definition of the ADMM flow, we conclude the lemma. □

A shadowing lemma for contracting flows was obtained in [8, 10]. Since the contracting flow is a special case of the point contracting flow, the following lemma is an extended result.

Lemma 3.3. *Every point contracting flow of a metric space has the eventual shadowing property.*

Proof. Let φ be a point contracting flow of a metric space M, and $x_* \in M$, $K, \rho > 0$ are from Definition 3.1. Fix $T > 0$ such that $0 < a < 1$ where $a = Ke^{-2\rho T}$ and denote $f = \varphi_T$. Then,

$$d(f(x), x_*) \le ad(x, x_*), \quad \forall x \in M. \tag{3.3}$$

Replacing $x = x_*$ we get $f(x_*) = x_*$.

Fix $\epsilon > 0$ and $\delta > 0$ such that

$$\delta \sum_{i=0}^{\infty} a^i < \frac{\epsilon}{2}. \tag{3.4}$$

Let $(x_n)_{n\ge 0}$ be a δ-pseudo-orbit of f. Fix $N \in \mathbb{N}$ such that

$$a^n d(x_0, x_*) < \frac{\epsilon}{2}, \quad \forall n \ge N. \tag{3.5}$$

By using (3.3), we have that

$$d(x_n, x_*) \le d(f(x_{n-1}), x_*) + d(f(x_{n-1}), x_n) \le ad(x_{n-1}, x_*) + \delta$$

and hence

$$d(x_n, x_*) \le a^n d(x_0, x_*) + \delta \sum_{i=0}^{n-1} a^i, \quad \forall n \ge 1.$$

Since $f(x_*) = x_*$, $f^n(x_*) = x_*$ for $n \in \mathbb{N}$, (3.4) and (3.5) imply

$$d(f^n(x_*), x_n) < \frac{\epsilon}{2} + \delta \sum_{i=0}^{\infty} a^i < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \quad \forall n \ge N.$$

This completes the proof. □

Proof of Theorem 2.3. Let $V : \mathbb{R}^n \to \mathbb{R}$ be a C^2 strictly convex map and λ be a positive definite $n \times n$ matrix. Then, the ADMM flow for V and λ is point contracting by Lemma 3.2 and so has the eventual shadowing property by Lemma 3.3. □

Next, we prove Theorem 2.4. In order to do it, we provide some technical lemmas. First, let us define a technical set. Given a map of a metric space $f : M \to M$, we define the *nonwandering set* $\Omega(f)$ of f by

$$\{x \in M : x = \lim_{n\to\infty} f^{k_n}(x_n) \text{ for a sequence } x_n \in M \text{ with } x_n \to x \text{ and } k_n \in \mathbb{N}\}.$$

Now, we have the following lemma for the nonwandering set for a flow.

Lemma 3.4. *Let $V : \mathbb{R}^n \to \mathbb{R}$ be C^2 convex and λ be a positive definite $n \times n$ matrix. Then, $\Omega(\varphi_T) = \arg\min_{X\in\mathbb{R}^n} V(X)$ for every $T > 0$, where φ is the ADMM flow generated by V and λ.*

Proof. Since V is C^2, the equation (1.1) generates a flow φ and we consider the time-T map φ_T, the nonwandering set $\Omega(\varphi_T)$ of φ_T. On the other hand, the convexity of V guarantees

$$\arg\min_{X\in\mathbb{R}^n} V(X) = \{X \in \mathbb{R}^n : \nabla V(X) = 0\} \overset{(2.1)}{\subset} \Omega(\varphi_T).$$

It remains to prove that $\nabla V(x_0) = 0$ for every $x_0 \in \Omega(\varphi_T)$.

Fix $x_0 \in \Omega(\varphi_T)$. Then, there are sequences $x_i \in \mathbb{R}^n$ and $k_i \in \mathbb{N}$ such that

$$x_i \to x_0 \quad \text{and} \quad X^i(k_iT) \to x_0 \quad \text{ as } n \to \infty,$$

where X^i is the solution of (1.1) with the initial value x_i. It follows that

$$V(X^i(0)) \to V(x_0) \quad \text{and} \quad V(X^i(k_iT)) \to V(x_0)$$

and also

$$|V(X^i(0)) - V(X^i(k_iT))| \to 0 \quad \text{as} \quad i \to \infty.$$

Then, by the Rolle's theorem, there is a sequence $t_i \in [0, k_iT]$ such that

$$\frac{d(V\circ X^i)}{dt}(t_i) = 0 \quad \text{for all sufficiently large } i.$$

Since

$$\frac{d(V\circ X^i)}{dt}(t_i) = \left\langle \nabla V(X^i(t_i)), \frac{dX^i}{dt}(t_i)\right\rangle = \langle \nabla V(X^i(t_i)), \lambda\nabla V(X^i(t_i))\rangle,$$

we have

$$\langle \nabla V(X^i(t_i)), \lambda\nabla V(X^i(t_i))\rangle = 0.$$

In addition, from the positive definiteness of λ, we conclude that

$$\nabla V(X^i(t_i)) = 0.$$

Then, $X^i(t)$ is a fixed point for all $t \in \mathbb{R}$ and so $X^i(t) = x_i$ for all $t \in \mathbb{R}$. In particular, $x_i = X^i(t_i)$ is a zero of ∇V (i.e., $\nabla V(x_i) = 0$) for all sufficiently large i. Since $x_0 = \lim_{i\to\infty} x_i$, the continuity of ∇V implies $\nabla V(x_0) = \lim_{i\to\infty} \nabla V(x_i) = 0$. This completes the proof. □

We introduce the following auxiliary definition. Given $\delta > 0$ and $f : X \to X$, a *δ-chain* is a finite sequence $\{x_0, \dots, x_k\}$ such that $d(f(x_{i-1}), x_i) \le \delta$ for $1 \le i \le k$. We say that the chain is from x to y if $x_0 = x$ and $x_k = y$.

Definition 3.5. We say that a map $f : M \to M$ of a metric space M has the *finite shadowing property, if for every $\epsilon > 0$, there is $\delta > 0$ such that every δ-chain $\{x_0, \dots, x_k\}$ has $x \in M$ satisfying $d(f^i(x), x_i) \le \epsilon$ for $0 \le i \le k$.*

The following lemma is an extended version of Theorem 2.6 in [5] for a noncompact metric space.

Lemma 3.6. *If $f : M \to M$ is a homeomorphism with the eventual shadowing property of metric space, then $f|_{\Omega(f)} : \Omega(f) \to \Omega(f)$ has the (finite) shadowing property.*

Proof. Given $x, y \in M$ and $\delta > 0$, we write $x \sim_\delta y$ if there is a δ-chain from x to y. We write $x \sim y$ if $x \sim_\delta y$ for every $\delta > 0$. Define the *chain recurrent set* of f as

$$CR(f) = \{x \in M : x \sim x\}.$$

This set is closed and invariant, i.e., $f(CR(f)) = CR(f)$. We claim that $\Omega(f) = CR(f)$.

Note that $\Omega(f) \subseteq CR(f)$ is clear, and the equality $\Omega(f) = CR(f)$ does not hold in general [1]. So, it suffices to prove $CR(f) \subset \Omega(f)$. Take $x \in CR(f)$ and $\epsilon > 0$. Let $\delta > 0$ be given by the eventual shadowing property for this ϵ. Since $x \sim x$, there is a sequence $\{x_0 = x, \dots, x_k = x\}$ with $d(f(x_{i-1}), x_i) \le \delta$ for $1 \le i \le k$. Defining

$$x_{n(k+1)+i} = x_i \quad \text{for } 0 \le i \le k-1 \text{ and } n \in \mathbb{N},$$

we obtain a sequence $(x_r)_{r \ge 0}$ with $d(f(x_r), x_{r+1}) \le \delta$ for every $r \ge 0$. Then, by the eventual shadowing property, there are $y \in M$ and $N \in \mathbb{N}$ such that $d(f^l(y), x_l) \le \epsilon$ for every $l \ge N$. Taking $z = f^{Nk}(y)$ we get $d(z, x) = d(f^{Nk}(y), x) = d(f^{Nk}(y), x_{Nk}) \le \epsilon$ and $d(f^{Nk}(z), x) = d(f^{2Nk}(y), x_{2Nk}) \le \epsilon$. Since ϵ is arbitrary, $x \in \Omega(f)$. This proves the claim.

Now, we back to the proof of the lemma. Fix $\epsilon > 0$. Let $\delta > 0$ be given by the eventual shadowing property of f for this ϵ. Take a δ-chain $\{x_0, x_1, \dots, x_k\} \subset \Omega(f)$. We must find $x \in \Omega(f)$ such that $d(f^i(x), x_i) \le \epsilon$ for every $0 \le i \le k$. For this, we proceed as follows.

First note that $\sim_\delta$ is an equivalence relation in $CR(f)$ and then in $\Omega(f)$ by the claim. So, we can write

$$\Omega(f) = \bigcup_{\alpha \in \Lambda} B_\alpha,$$

where $\{B_\alpha : \alpha \in \Lambda\}$ (for some index set Λ) is the collection of equivalence classes of $\sim_\delta$ restricted to $\Omega(f)$. More precisely, $B_\alpha = \{y \in \Omega(f) : x \sim_\delta y\}$ for every $\alpha \in \Lambda$ and $x \in B_\alpha$. Choose $\xi \in \Lambda$ such that $x_0 \in B_\xi$. It follows from the definition of $\sim_\delta$ that $x_i \in B_\xi$ for every $0 \le i \le k$. In particular, $x_0 \sim_\delta x_k$, so we can complete the sequence $\{x_0, x_1, \dots, x_k\}$ with another δ-chain $\{x_k, x_{k+1}, \dots, x_{k+l} = x_0\}$. This results in a sequence $\{x_0, x_1, \dots, x_{k+l}\}$ and we can generate a δ-pseudo-orbit $(x_r)_{r \ge 0}$ by setting $x_{n(k+l)+s} = x_s$ for

$n \geq 0$, $0 \leq s \leq k+l-1$. Then, by the eventual shadowing property, there are $x \in M$ and $N \in \mathbb{N}$ such that

$$d(f^r(x), x_r) \leq \epsilon, \quad \forall r \geq N.$$

Fix $n \geq N$ and let $z = f^{n(k+l)}(y)$. Then, $n(k+l)+s \geq N$ for every $0 \leq s \leq k$, so

$$d(f^s(z), x_s) = d(f^{n(k+l)+s}(y), x_{n(k+l)+s}) \leq \epsilon, \quad \forall 0 \leq s \leq k.$$

Hence the initial finite sequence $\{x_1, \dots, x_k\}$ can be ϵ-shadowed. Therefore, $f|_{\Omega(f)}$ has the finite shadowing property. □

Proof of Theorem 2.4. Let $V : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Suppose that the ADMM flow φ for V and a positive definite $n \times n$ real matrix λ has the eventual shadowing property. By definition, there is $T > 0$ such that $f = \varphi_T$ has the eventual shadowing property. By Lemma 3.4, one has that $\Omega(f) = \operatorname{argmin}_{X \in \mathbb{R}^n} V(X)$. It follows that $\Omega(f)$ consists of the zeroes of ∇V and hence $f|_{\Omega(f)}$ is the identity of $\Omega(f)$. Since f has the eventual shadowing property, $f|_{\Omega(f)}$ has the finite shadowing property by Lemma 3.6. Since V is convex, $\operatorname{argmin}_{X \in \mathbb{R}^n} V(X) = \Omega(f)$ is convex (hence connected). Combining the fact that $f|_{\Omega(f)}$ is the identity of $\Omega(f)$, the finite shadowing property of $f|_{\Omega(f)}$ and the connectedness of $\Omega(f)$, we conclude that $\Omega(f)$ reduces to a single point by Theorem 2.3.2 in [1]. Therefore V has a unique minimizer. □

4. Nonshadowing for a flow by a second order ODE

In this section, we study the shadowing property for a flow of the dynamical system generated by a second order ODE. The second order ODE can be used to model the accelerated ADMM method [6]:

$$(A^\top A)\left(\ddot{X} + \frac{r}{t}\dot{X}\right) + \nabla V(X) = 0,$$

where A is an invertible $n \times n$ matrix, $V : \mathbb{R}^n \to \mathbb{R}$ is a differentiable objective function and $r \in \mathbb{R}$. If $A = I$, then the above ODE corresponds to Nesterov's accelerated gradient method [16].

By letting $Y = \dot{X}$, we reformulate it as the following system of first-order ODEs:

$$\begin{cases} \dot{X} = Y, \\ \dot{Y} = -(A^\top A)^{-1}\nabla V(X) - \frac{r}{t}Y. \end{cases}$$

However, this equation is nonautonomous if $r \neq 0$. In such a case, the equation generates the evolution process [14]. Since our primary interest is on flows rather than processes, we henceforth assume that $r = 0$. Under this assumption, we consider the autonomous ODE in $\mathbb{R}^{2n}$,

$$\begin{cases} \dot{X} = Y, \\ \dot{Y} = -\lambda \nabla V(X), \end{cases} \tag{4.1}$$

with a invertible $n \times n$ matrix λ. This equation generates a flow φ in $\mathbb{R}^{2n}$ that depends on both the objective function V and the matrix λ. On the contrary, we have the following opposite result for this flow.

Theorem 4.1. *The flow φ generated by (4.1) with a C^2 strongly convex function $V : \mathbb{R}^n \to \mathbb{R}$ and a positive definite $n \times n$ matrix λ does not have the eventual shadowing property.*

Proof. On the contrary, we assume that φ has the eventual shadowing property. As before, since λ is positive definite, we can write $\lambda = A^\top A$ for some invertible matrix A. We define $E : \mathbb{R}^{2n} \to \mathbb{R}$ by

$$E(X, Y) = V(X) + \frac{1}{2}\|(A^{-1})^\top(Y)\|^2 \quad \text{for } (X, Y) \in \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}.$$

Similar to [6], we take the time derivative of E and evaluate at a solution (X, Y) of (4.1):

$$\begin{aligned}
\frac{dE}{dt}(t) &= \langle \nabla_X E, \dot{X} \rangle + \langle \nabla_Y E, \dot{Y} \rangle \\
&= \langle \nabla V(X), Y \rangle - \langle A^{-1}(A^{-1})^\top Y, (A^\top A)\nabla V(X) \rangle \\
&= \langle \nabla V(X), Y \rangle - \langle Y, (A^{-1})(A^{-1})^\top (A^\top A)\nabla V(X) \rangle \\
&= \langle \nabla V(X), Y \rangle - \langle Y, (A^\top A)^{-1}(A^\top A)\nabla V(X) \rangle \\
&= \langle \nabla V(X), Y \rangle - \langle Y, \nabla V(X) \rangle \\
&= 0.
\end{aligned}$$

Thus, E is constant along the orbits of the flow φ. Moreover, since V is strongly convex, V has a unique minimizer X_*. It follows that E has a unique critical point at $Z^* = (X_*, 0)$ and it is a minimizer with the minimum value $E(Z^*) = V(X_*)$.

Since the level sets of E are concentric spheres around $Z_* = (X_*, 0)$, these sets are compact and do not contain Z_*. So, we can take a positive number

$$\Delta = \frac{1}{4}\mathrm{dist}(Z^*, E^{-1}(V(X_*) + 1)). \tag{4.2}$$

Then, by choosing $0 < \epsilon < \Delta$ such that

$$d(Z^*, Z) < \epsilon \quad \Longrightarrow \quad \mathrm{diam}(E^{-1}(E(Z))) < \Delta, \tag{4.3}$$

we can select δ for the eventual shadowing property of the time 1-map φ_1 with respect to this ϵ. By jumping along the level sets of E from Z_* to $E^{-1}(V(X_*)+1)$, we can construct a δ-pseudo-orbit $\{Z_k\}_{k\geq 0}$ such that $Z_i = Z_*$ for infinitely many i in $\mathbb{N}$, and $Z_j \in E^{-1}(V(X_*)+1)$ for infinitely many j in $\mathbb{N}$. This process is illustrated in Figure 1. The eventual shadowing property implies that there is $Z \in \mathbb{R}^{2n}$ such that

$$d(\varphi_i(Z), Z_i) < \epsilon \quad \text{for all sufficiently large } i.$$

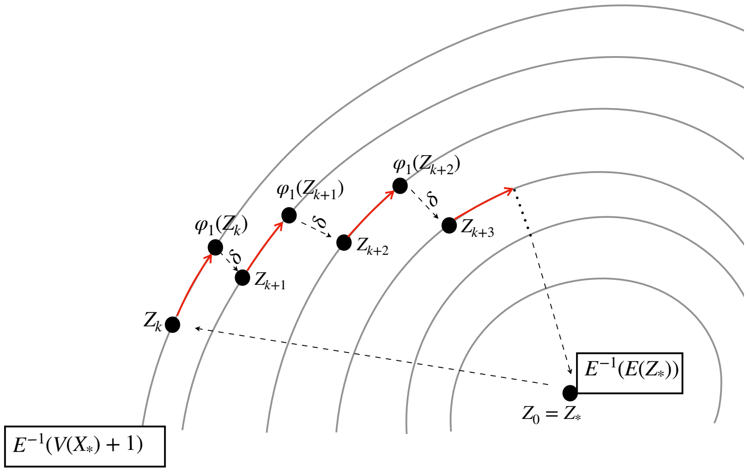

FIGURE 1. A δ-pseudo-orbit of φ_1 by jumping the levels sets

In particular, there are time i-map φ_i and time j-map φ_j with $i < j \in \mathbb{N}$ such that

$$d(\varphi_i(Z), Z_*) < \epsilon \quad \text{and} \quad d(\varphi_j(Z), Z_j) < \epsilon \quad \text{with} \quad Z_j \in E^{-1}(V(X_*) + 1).$$

Since E is constant along the orbits of φ, one has $\varphi_j(Z) \in E^{-1}(E(\varphi_i(Z)))$. Finally,

$$\begin{aligned} d(Z_*, Z_j) &\leq d(Z_*, \varphi_i(Z)) + d(\varphi_i(Z), \varphi_j(Z)) + d(\varphi_j(Z), Z_j) \\ &\leq 2\epsilon + \operatorname{diam}(E^{-1}(E(\varphi_i(Z)))) \\ &\overset{(4.3)}{<} 2\Delta + \Delta \\ &= 3\Delta \\ &\overset{(4.2)}{<} \frac{3}{4}\operatorname{dist}(Z_*, E^{-1}(V(X_*) + 1)) \\ &< \operatorname{dist}(Z_*, E^{-1}(V(X_*) + 1)). \end{aligned}$$

It contradicts to $Z_j \in E^{-1}(V(X_*) + 1)$ and we draw our conclusion. □

5. Conclusion

In this study, we have analyzed the characteristics of solutions for the ordinary differential equations for the alternating direction method of multipliers. Our analysis is based on the shadowing property, which is a tool used in dynamical systems to investigate computer simulations of ODE solutions and trace pseudo-orbits by real orbits. We have showed that the flow induced by the

ADMM for a C^2 strongly convex objective function has the eventual shadowing property. As for the converse, we provided a partial answer that the convexity with the eventual shadowing property implies a unique minimizer.

However, we have also found that the flow generated by a second-order ODE, which is related to the accelerated version of ADMM, does not possess the eventual shadowing property. Investigating the shadowing properties for the general accelerated optimization method will require a theory of evolution processes [12,19], as these properties have been little studied. This may be the focus of our future work.

6. Appendix

In this appendix, we derive the ADMM flow of the form (1.1) for the completion. From (1.3a) and (1.3b), we have

$$0 = \nabla g_1(x^{k+1}) + \rho A^\top (Ax^{k+1} - z^k + u^k), \tag{6.1a}$$

$$0 = \nabla g_2(z^{k+1}) - \rho(Ax^{k+1} - z^{k+1} + u^k). \tag{6.1b}$$

Multiplying (6.1b) by $A^\top$ and adding it to (6.1a), we get

$$0 = \nabla g_1(x^{k+1}) + A^\top \nabla g_2(z^{k+1}) + \rho A^\top (z^{k+1} - z^k). \tag{6.2}$$

Let $x^k = X(k\epsilon)$, $z^k = Z(k\epsilon)$ and $u^k = U(k\epsilon)$ be discretizations for the continuum trajectories $X(t)$, $Z(t)$ and $U(t)$ defined on $t \geq 0$. By taking $t = k\epsilon$, we get that

$$\begin{cases} x^k = X(t), \\ x^{k+1} = X(t+\epsilon), \end{cases} \quad \begin{cases} z^k = Z(t), \\ z^{k+1} = Z(t+\epsilon), \end{cases} \quad \begin{cases} u^k = U(t), \\ u^{k+1} = U(t+\epsilon). \end{cases} \tag{6.3}$$

Plugging (6.3) in (1.3c) yields that

$$U(t+\epsilon) - U(t) = AX(t+\epsilon) - Z(t+\epsilon).$$

By letting $\epsilon \to 0$ and the continuity of $X(\cdot)$, $Z(\cdot)$, and $U(\cdot)$, we get

$$Z(t) = AX(t). \tag{6.4}$$

Replacing $\rho = \epsilon^{-1}$ and putting (6.3) and (6.4) in (6.2) imply that

$$0 = \nabla g_1(X(t+\epsilon)) + A^\top \nabla g_2(AX(t+\epsilon)) + A^\top A \left(\frac{X(t+\epsilon) - X(t)}{\epsilon} \right). \tag{6.5}$$

Taking the limit as $\epsilon \to 0$ in the above equation yields that

$$0 = \nabla g_1(X(t)) + A^\top \nabla g_2(AX(t)) + A^\top A \dot{X}(t)$$

and so

$$\dot{X}(t) = -(A^\top A)^{-1} \left(\nabla g_1(X(t)) + A^\top \nabla g_2(AX(t)) \right).$$

Therefore, we obtain (1.1) by taking $V(x) = g_1(x) + g_2(Ax)$ and $\lambda = (A^\top A)^{-1}$.

References

[1] N. Aoki and K. Hiraide, *Topological Theory of Dynamical Systems*, North-Holland Mathematical Library, 52, North-Holland, Amsterdam, 1994.

[2] R. Bhatia, *Positive Definite Matrices*, Princeton Series in Applied Math., 2007.

[3] A. Bielecki and J. Ombach, *Shadowing property in analysis of neural networks dynamics*, J. Comput. Appl. Math. **164/165** (2004), 107–115. https://doi.org/10.1016/S0377-0427(03)00486-2

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning **3** (2011), 1–122.

[5] M. H. Dong, K. Lee, and N.-T. Nguyen, *Expanding measures for homeomorphisms with eventually shadowing property*, J. Korean Math. Soc. **57** (2020), no. 4, 935–955. https://doi.org/10.4134/JKMS.j190453

[6] G. França, D. P. Robinson, and R. E. Vidal, *ADMM and accelerated ADMM as continuous dynamical system*, Proceedings of the 35th International Conference on Machine Learning **80** (2018), 1559–1567.

[7] G. França, D. P. Robinson, and R. E. Vidal, *A nonsmooth dynamical systems perspective on accelerated extensions of ADMM*, IEEE Trans. Automat. Control **68** (2023), no. 5, 2966–2978. https://doi.org/10.1109/tac.2023.3238857

[8] M. Garg and R. Das, *Continuous semi-flows with the almost average shadowing property*, Chaos Solitons Fractals **105** (2017), 1–7. https://doi.org/10.1016/j.chaos.2017.10.005

[9] C. Good and J. Meddaugh, *Orbital shadowing, internal chain transitivity and ω-limit sets*, Ergodic Theory Dynam. Systems **38** (2018), no. 1, 143–154. https://doi.org/10.1017/etds.2016.30

[10] L. F. He, Z. H. Wang, and H. Li, *Continuous semiflows with the shadowing property*, Acta Math. Appl. Sinica **19** (1996), no. 2, 297–303.

[11] A. Jung, *A fixed-point of view on gradient methods for big data*, Frontiers in Applied Mathematics and Statistics **3** (2017), 18.

[12] W. Jung, K. Lee, and C. A. Morales, *Dynamics of G-processes*, Stoch. Dyn. **20** (2020), no. 1, 2050037, 30 pp. https://doi.org/10.1142/S0219493720500379

[13] Y. M. Jung, B. Shin, and S. Yun, *Global attractor and limit points for nonsmooth ADMM*, Appl. Math. Lett. **128** (2022), Paper No. 107890, 8 pp. https://doi.org/10.1016/j.aml.2021.107890

[14] P. E. Kloeden and M. Rasmussen, *Nonautonomous dynamical systems*, Mathematical Surveys and Monographs, 176, Amer. Math. Soc., Providence, RI, 2011. https://doi.org/10.1090/surv/176

[15] M. Komuro, *One-parameter flows with the pseudo-orbit tracing property*, Monatsh. Math. **98** (1984), no. 3, 219–253. https://doi.org/10.1007/BF01507750

[16] Y. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR **269** (1983), no. 3, 543–547.

[17] A. Orbieto and A. Lucchi, *Shadowing properties of optimization algorithms*, 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

[18] B. G. Pachpatte, *Inequalities for differential and integral equations*, Mathematics in Science and Engineering, 197, Academic Press, Inc., San Diego, CA, 1998.

[19] R. Vasisht and R. Das, *Specification and shadowing properties for non-autonomous systems*, J. Dyn. Control Syst. **28** (2022), no. 3, 481–492. https://doi.org/10.1007/s10883-021-09535-4

Yoon Mo Jung
Department of Mathematics
Sungkyunkwan University
Suwon-si 16419, Korea
Email address: yoonmojung@skku.edu

Bomi Shin
Institute of Basic Science
Sungkyunkwan University
Suwon-si 16419, Korea
Email address: bmshin@skku.edu

Sangwoon Yun
Department of Mathematics Education
Sungkyunkwan University
Seoul 03063, Korea
Email address: yswmathedu@skku.edu