

基于不定系数常微分方程的加速梯度算法研究

谢旭东

北京邮电大学理学院, 北京

收稿日期: 2024年3月18日; 录用日期: 2024年6月5日; 发布日期: 2024年6月13日

摘要

随着人工智能功能的日益强大, 对机器学习、深度学习的研究也日趋完善。深度学习的核心思想是求解优化问题, 而随着当前实际问题规模的日益扩大, 梯度类算法由于运算成本相对较低, 日益受到人们的关注和青睐, 设计新的加速梯度算法是加快问题求解的关键。本文基于常微分方程, 得到了一系列新的加速梯度算法。本文考虑了一类系数不定的二阶常微分方程, 证明了系数在满足什么条件下, 该微分方程能够快速收敛。对于该微分方程, 本文采用两种不同的离散格式: 显式欧拉格式、隐式欧拉格式进行离散, 分别得到了两类不同的优化算法, 分别证明了算法的收敛速率。最后, 本文通过数值实验证明了算法的有效性。

关键词

加速梯度算法, 常微分方程, Layapunov函数

Research on Accelerated Gradient Algorithm Based on Ordinary Differential Equations with Uncertain Coefficients

Xudong Xie

School of Science, Beijing University of Posts and Telecommunications, Beijing

Received: Mar. 18th, 2024; accepted: Jun. 5th, 2024; published: Jun. 13th, 2024

Abstract

With the increasingly powerful functions of artificial intelligence, research on machine learning and deep learning is also becoming increasingly sophisticated. The core idea of deep learning is to solve optimization problems, and with the increasing scale of practical problems, gradient algorithms are increasingly receiving attention and favor due to their relatively low computational

costs. Designing new accelerated gradient algorithms is the key to accelerating problem solving. This article presents a series of new accelerated gradient algorithms based on ordinary differential equations. This article considers a class of second-order ordinary differential equations with uncertain coefficients and proves under what conditions the coefficients satisfy that the differential equation can converge quickly. For this differential equation, this paper adopts two different discretization schemes: explicit Euler scheme and implicit Euler scheme for discretization, and obtains two different optimization algorithms, respectively proving the convergence rate of the algorithms. Finally, this article demonstrates the effectiveness of the algorithm through numerical experiments.

Keywords

Accelerated Gradient Algorithm, Ordinary Differential Equations, Layapunov Functions

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着计算机与网络的飞速发展，机器学习在各个领域中起着越来越大的作用，正在不断改变着我们的生活。不管是机器学习还是神经网络，它们往往都可以归结为求解最优化问题[1]。

机器学习中的优化问题根据目标函数的凸性，可分为凸优化问题和非凸优化问题；根据目标函数是否带有约束，可分为约束优化问题和无约束优化问题。对于非凸优化问题，在实际应用中，有时候可以通过如局部线性化、凸松弛等技巧将非凸优化问题转化为凸优化问题，从而利用凸优化的相关算法来求解。对于约束优化问题，可以通过如罚函数等方法将其转化为无约束优化问题。本文考虑的是无约束凸优化问题。

机器学习在现代社会扮演着越来越重要的角色，对医疗健康、金融领域、智能交通、自然语言处理等领域都产生了深远的影响。而作为机器学习的根本，对求解优化问题相关算法的研究在近年来日益受到人们的关注。

本文考虑的问题为无约束优化问题：

$$\min_{x \in R^n} f(x) \quad (1)$$

其中 $f(x)$ 是光滑的强凸函数。定义 R^n 上的一类光滑凸函数为 $\mathcal{F}_L^1$ ，若 $\forall x, y \in R^n$ ，函数梯度满足 Lipschitz 连续：

$$\|\nabla f(y) - \nabla f(x)\| \leq L \|y - x\| \quad (2)$$

其中 L 是 Lipschitz 系数。定义 R^n 上的强凸函数 $f \in \mathcal{S}_\mu^1$ ，若存在常数 μ ，使得函数 $g(x) = f(x) - \frac{\mu}{2} \|x\|^2$ 为凸函数，则称函数 $f(x)$ 为 μ -强凸函数。若函数 $f(x)$ 为 μ -强凸函数，则 $\forall x, y \in R^n$ ，满足：

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2 \quad (3)$$

最基础的梯度算法是梯度下降法：

$$x_{k+1} = x_k - s \nabla f(x_k) \quad (4)$$

其中 s 是步长。梯度下降法的基本思想是沿着函数梯度的反方向逐步迭代更新参数，以减小目标函数的值。然而，梯度下降法的算法性能在很大程度上受到目标函数的约束。为了解决这一问题，提高算法的收敛速率，近几十年来，学者们从不同的角度出发，设计了不同的加速梯度算法，其中较为典型的有重球方法[2]与 Nesterov 加速梯度算法(NAG 算法) [3]，其中 NAG 算法具有优秀的算法性能，在机器学习中被广泛运用。然而，算法的加速原理至今未能得到合理的解释。

常微分方程(Ordinary Differential Equations, ODE)与一阶优化算法之间存在密切的联系。在无约束优化问题中，求解优化问题的过程可以被视为一个微分方程的求解过程。因此，微分方程理论为我们提供了一种理论框架，帮助我们理解梯度类算法在优化问题中的工作原理，并为算法的收敛性和稳定性提供理论支持。基于这一思想，人们开始从 ODE 的角度来解释算法的加速原理，并设计新的加速梯度算法。Attouch [4]在重球方法对应 ODE 的基础上引入含有 Hessian 矩阵的项，得到惯性动力系统。Su 等人[5]将 NAG 算法的迭代点视为一个连续变量在某些特定时间点取值的近似，利用泰勒展开，得到了 NAG 算法对应连续时间下的 ODE。Shi B 等人[6]在此基础上，通过更高阶的泰勒展开，分别得到了 NAG 算法与重球方法对应的高精度 ODE，高精度 ODE 更能反应算法的新能。Chen 等人[7]以 ODE 为工具，给出了一阶凸优化算法的统一收敛性分析框架，并提出了一种新的动态惯性牛顿系统，以此得到了新的加速梯度算法[8]。

本文基于文献[6]中提出的 Layapunov 函数构建方法，考虑一类系数不定的 ODE：

$$\ddot{X} + \alpha \dot{X} + \beta \nabla^2 f(X) \dot{X} + \nabla f(X) = 0 \quad (5)$$

首先，通过构造合适的 Layapunov 函数，证明该 ODE 在系数 α, β 在满足什么条件下能够快速收敛。其次，本文使用两种不同的离散格式：显式欧拉格式和隐式欧拉格式对该 ODE 进行离散，得到两种不同的优化算法，并分别构造合适的 Layapunov 函数，证明这两类算法的收敛速率。最后，通过数值实验，验证所得算法的实用性。

2. 系数不定的二阶常微分方程与收敛速率

在本章中，我们考虑一个系数不定的二阶 ODE：

$$\ddot{X} + \alpha \dot{X} + \beta \nabla^2 f(X) \dot{X} + \nabla f(X) = 0 \quad (6)$$

定理 2.1. 证明了系数 α, β 在满足什么条件时，ODE 能够快速收敛。

定理 2.1. 假设目标函数 $f \in \mathcal{S}_{\mu, L}^1$ ， $X(t)$ 为 ODE: $\ddot{X} + \alpha \dot{X} + \beta \nabla^2 f(X) \dot{X} + \nabla f(X) = 0$ 的一个解，当系

$$\text{数 } \alpha, \beta \text{ 满足 } \begin{cases} \frac{3}{4}\alpha^2 \leq \frac{\mu}{2} \\ \beta \leq \frac{8}{3\alpha} \\ \alpha \geq 0 \end{cases} \text{ 时, } X(t) \text{ 满足 } f(X(t)) - f(x^*) \leq C \|x_0 - x^*\|^2 e^{-\frac{\alpha}{4}t}, \text{ 其中}$$

$$C = \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2 \right) L^2 + \frac{3}{4}\alpha^2 \right) \|X_0 - x^*\|^2.$$

证明：设 Layapunov 函数为：

$$\varepsilon(t) = f(X) - f(x^*) + \frac{1}{4} \|\dot{X}\|^2 + \frac{1}{4} \|\dot{X} + \alpha(X - x^*) + \beta \nabla f(X)\|^2.$$

$$\begin{aligned}\frac{d\varepsilon}{dt} &= \langle \nabla f(X), \dot{X} \rangle + \frac{1}{2} \langle \dot{X}, -\alpha \dot{X} - \beta \nabla^2 f(X) \dot{X} - \nabla f(X) \rangle + \frac{1}{2} \langle \dot{X} + \alpha(X - x^*) + \beta \nabla f(X), -\nabla f(X) \rangle \\ &= -\frac{1}{2} \alpha \|\dot{X}\|^2 - \frac{1}{2} \beta \|\nabla f(X)\|^2 - \frac{1}{2} \alpha \langle \nabla f(X), X - x^* \rangle - \frac{1}{2} \beta \dot{X}^T \nabla^2 f(X) \dot{X} \\ &\leq -\frac{1}{2} \alpha \|\dot{X}\|^2 - \frac{1}{2} \beta \|\nabla f(X)\|^2 - \frac{1}{2} \alpha \langle \nabla f(X), X - x^* \rangle.\end{aligned}$$

由 $f(x)$ 的强凸性可得

$$\begin{aligned}\frac{1}{2} \alpha \langle \nabla f(X), X - x^* \rangle &\geq \frac{1}{2} \alpha [f(X) - f(x^*)] + \frac{\mu \alpha}{4} \|X - x^*\|^2 \\ &\geq \frac{1}{4} \alpha [f(X) - f(x^*)] + \frac{\mu \alpha}{4} \|X - x^*\|^2,\end{aligned}$$

故

$$\begin{aligned}\frac{d\varepsilon}{dt} &\leq -\frac{1}{2} \alpha \|\dot{X}\|^2 - \frac{1}{2} \beta \|\nabla f(X)\|^2 - \frac{1}{4} \alpha [f(X) - f(x^*)] - \frac{\mu \alpha}{8} \|X - x^*\|^2 \\ &= -\alpha \left(\frac{1}{4} [f(X) - f(x^*)] + \frac{1}{2} \|\dot{X}\|^2 + \frac{1}{4} \mu \|X - x^*\|^2 + \frac{\beta}{2\alpha} \|\nabla f(X)\|^2 \right).\end{aligned}\quad (7)$$

由柯西 - 施瓦茨不等式,

$$\|\dot{X} + \alpha(X - x^*) + \beta \nabla f(X)\|^2 \leq 3(\|\dot{X}\|^2 + \alpha^2 \|X - x^*\|^2 + \beta^2 \|\nabla f(X)\|^2).$$

故

$$\varepsilon(t) \leq f(X) - f(x^*) + \|\dot{X}\|^2 + \frac{3}{4} \alpha^2 \|X - x^*\|^2 + \frac{3}{4} \beta^2 \|\nabla f(X)\|^2. \quad (8)$$

对比式(7)和式(8)中的各项, 当系数 α, β 满足 $\begin{cases} \frac{3}{4} \alpha^2 \leq \mu \\ \beta \leq \frac{8}{3\alpha} \end{cases}$ 时,

$$\frac{d\varepsilon}{dt} \leq -\frac{\alpha}{4} \varepsilon(t),$$

即

$$\varepsilon(t) \leq e^{-\frac{\alpha}{4}t} \varepsilon(0),$$

$$f(X(t)) - f(x^*) \leq \varepsilon(t) \leq e^{-\frac{\alpha}{4}t} \varepsilon(0).$$

设 $X(0) = X_0$, $\dot{X}(0) = \nabla f(X_0)$, 由于 $f \in \mathcal{S}_{\mu,L}^1$, 则有

$$f(X_0) - f(x^*) \leq \frac{L}{2} \|X_0 - x^*\|^2, \quad \|\nabla f(X_0)\| \leq L \|X_0 - x^*\|.$$

$$\begin{aligned}\varepsilon(0) &= f(X_0) - f(x^*) + \frac{1}{4} \|\nabla f(X_0)\|^2 + \frac{1}{4} \|\nabla f(X_0) + \alpha(X_0 - x^*) + \beta \nabla f(X_0)\|^2 \\ &\leq f(X_0) - f(x^*) + \left(1 + \frac{3}{4} \beta^2\right) \|\nabla f(X_0)\|^2 + \frac{3}{4} \alpha^2 \|X_0 - x^*\|^2 \\ &= \left(\frac{L}{2} + \left(1 + \frac{3}{4} \beta^2\right) L^2 + \frac{3}{4} \alpha^2\right) \|X_0 - x^*\|^2.\end{aligned}$$

综上,

$$f(X(t)) - f(x^*) \leq C \|x_0 - x^*\|^2 e^{-\frac{\alpha}{4}t},$$

$$\text{其中 } C = \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2 \right) L^2 + \frac{3}{4}\alpha^2 \right) \|X_0 - x^*\|^2.$$

3. 一类新的加速梯度算法

在第二章中, 我们证明了系数不定的 ODE:

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

在系数 α, β 在满足什么条件时, ODE 能够快速收敛。式(6)等价于:

$$\dot{X} = V, \dot{V} = -\alpha V - \beta \nabla^2 f(X) V - \nabla f(X) \quad (9)$$

本章将使用两种不同的离散格式: 显式欧拉格式和隐式欧拉格式对式(9)进行离散得到相应的优化算法, 并分别证明它们的收敛速率。

3.1. 显式欧拉格式

第一种离散格式是显式欧拉格式, 它的近似规则如下:

$$x_{k+1} - x_k = \sqrt{s} v_k, \sqrt{s} \nabla^2 f(x_k) v_k \approx \nabla f(x_{k+1}) - \nabla f(x_k).$$

由显式欧拉格式离散式(9)可得如下优化算法:

$$\begin{cases} x_{k+1} - x_k = \sqrt{s} v_k \\ v_{k+1} - v_k = -\alpha \sqrt{s} v_k - \beta (\nabla f(x_{k+1}) - \nabla f(x_k)) - \sqrt{s} \nabla f(x_k) \end{cases} \quad (10)$$

定理 3.1. 假设目标函数 $f \in \mathcal{S}_{\mu, L}^1$, 对于显式欧拉格式产生的迭代点序列 $\{x_k\}_{k \in \mathbb{N}}$, 当系数 α, β 满足

$$\begin{cases} \frac{3}{4}\alpha^2 \leq \frac{\mu}{2} \\ \beta \leq \frac{8}{3\alpha} \\ \alpha \geq 0 \end{cases}, \text{ 步长 } s \text{ 满足 } s \leq \min \left\{ \frac{4}{9L^2\beta^2}, \frac{\alpha^2}{(3\alpha^2 + 2L)^2}, \frac{(\alpha\gamma - \beta\gamma L)^2}{16L^2\gamma^4}, \frac{1}{L} \right\} \text{ 时, 收敛速率满足}$$

$$f(x_k) - f(x^*) \leq C \left(1 - \frac{\alpha\sqrt{s}}{8} \right)^k \|x_0 - x^*\|^2,$$

$$\text{其中 } C = \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2 \right) L^2 + \frac{3}{4}\alpha^2 \right) \|X_0 - x^*\|^2.$$

证明: 设 Layapunov 函数为:

$$\varepsilon(k) = f(x_k) - f(x^*) + \frac{1}{4} \|v_k\|^2 + \frac{1}{4} \left\| \alpha(x_k - x^*) + v_k + \beta \nabla f(x_k) \right\|^2.$$

$$\begin{aligned} & \varepsilon(k+1) - \varepsilon(k) \\ &= \frac{1}{4} \langle v_{k+1} - v_k, v_{k+1} + v_k \rangle + f(x_{k+1}) - f(x_k) \\ &+ \frac{1}{4} \left\langle \alpha(x_{k+1} - x_k) + v_{k+1} - v_k + \beta(\nabla f(x_{k+1}) - \nabla f(x_k)), \alpha(x_{k+1} + x_k - 2x^*) + v_{k+1} + v_k + \beta(\nabla f(x_{k+1}) + \nabla f(x_k)) \right\rangle \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{1}{2} \langle v_{k+1} - v_k, v_k \rangle + \frac{1}{4} \|v_{k+1} - v_k\|^2 + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2 \\
 &\quad - \frac{1}{2} \langle \sqrt{s} \nabla f(x_k), \alpha(x_k - x^*) + v_k + \beta \nabla f(x_k) \rangle + \frac{1}{4} \|\sqrt{s} \nabla f(x_k)\|^2 \\
 &= -\frac{\alpha \sqrt{s}}{2} \|v_k\|^2 - \frac{\beta}{2\sqrt{s}} \langle \nabla f(x_{k+1}) - \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{4} \|\alpha \sqrt{s} v_k + \beta(\nabla f(x_{k+1}) - \nabla f(x_k)) + \sqrt{s} \nabla f(x_k)\|^2 \\
 &\quad + \frac{Ls}{2} \|v_k\|^2 - \frac{\alpha \sqrt{s}}{2} \langle \nabla f(x_k), x_k - x^* \rangle - \frac{\beta \sqrt{s}}{2} \|\nabla f(x_k)\|^2 + \frac{s}{4} \|\nabla f(x_k)\|^2.
 \end{aligned}$$

由柯西 - 施瓦茨不等式,

$$\begin{aligned}
 &\|\alpha \sqrt{s} v_k + \beta(\nabla f(x_{k+1}) - \nabla f(x_k)) + \sqrt{s} \nabla f(x_k)\|^2 \\
 &\leq 3\alpha^2 s \|v_k\|^2 + 3\beta^2 \|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 + 3s \|\nabla f(x_k)\|^2.
 \end{aligned}$$

故

$$\begin{aligned}
 \varepsilon(k+1) - \varepsilon(k) &\leq -\frac{\alpha \sqrt{s}}{4} \left(\|v_k\|^2 + \langle \nabla f(x_k), x_k - x^* \rangle + \frac{\beta}{\alpha} \|\nabla f(x_k)\|^2 \right) \\
 &\quad - \frac{\beta}{2\sqrt{s}} \langle \nabla f(x_{k+1}) - \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{3}{4} \beta^2 \|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 \\
 &\quad + \left(\frac{3\alpha^2 s}{4} + \frac{Ls}{2} - \frac{\alpha \sqrt{s}}{4} \right) \|v_k\|^2 + s \|\nabla f(x_k)\|^2 \\
 &\quad - \frac{\alpha \sqrt{s}}{4} \langle \nabla f(x_k), x_k - x^* \rangle - \frac{\beta \sqrt{s}}{4} \|\nabla f(x_k)\|^2.
 \end{aligned}$$

由于 $f \in \mathcal{S}_{\mu, L}^1$, 则有

$$\begin{aligned}
 \|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 &\leq L \langle \nabla f(x_{k+1}) - \nabla f(x_k), x_{k+1} - x_k \rangle, \\
 \|\nabla f(x_k)\|^2 &\leq L \langle \nabla f(x_k), x_k - x^* \rangle.
 \end{aligned}$$

故

$$\begin{aligned}
 \varepsilon(k+1) - \varepsilon(k) &\leq -\frac{\alpha \sqrt{s}}{4} \left(\|v_k\|^2 + \langle \nabla f(x_k), x_k - x^* \rangle + \frac{\beta}{\alpha} \|\nabla f(x_k)\|^2 \right) \\
 &\quad - \left(\frac{\beta}{2L\sqrt{s}} - \frac{3}{4} \beta^2 \right) \|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 \\
 &\quad - \left(\frac{\alpha \sqrt{s}}{4} - \frac{3\alpha^2 s}{4} - \frac{Ls}{2} \right) \|v_k\|^2 \\
 &\quad - \left(\frac{\alpha \sqrt{s}}{4L} - s \right) \|\nabla f(x_k)\|^2.
 \end{aligned} \tag{11}$$

对于式(11), 若步长 s 满足条件

$$\begin{cases} \frac{\beta}{2L\sqrt{s}} - \frac{3}{4} \beta^2 \geq 0 \\ \frac{\alpha \sqrt{s}}{4} - \frac{3\alpha^2 s}{4} - \frac{Ls}{2} \geq 0, \\ \frac{\alpha \sqrt{s}}{4L} - s \geq 0 \end{cases}$$

可得

$$\varepsilon(k+1) - \varepsilon(k) \leq -\frac{\alpha\sqrt{s}}{4} \left(\|v_k\|^2 + \langle \nabla f(x_k), x_k - x^* \rangle + \frac{\beta}{\alpha} \|\nabla f(x_k)\|^2 \right), \quad (12)$$

由柯西 - 施瓦茨不等式,

$$\left\| \alpha(x_k - x^*) + v_k + \beta \nabla f(x_k) \right\|^2 \leq 3\alpha^2 \|x_k - x^*\|^2 + 3\|v_k\|^2 + 3\beta^2 \|\nabla f(x_k)\|^2,$$

故

$$\begin{aligned} \varepsilon(k) &\leq \|v_k\|^2 + f(x_k) - f(x^*) + \frac{3}{4}\alpha^2 \|x_k - x^*\|^2 + \frac{3}{4}\beta^2 \|\nabla f(x_k)\|^2 \\ &\leq \|v_k\|^2 + f(x_k) - f(x^*) + \frac{\mu}{2} \|x_k - x^*\|^2 + \frac{2\beta}{\alpha} \|\nabla f(x_k)\|^2 \\ &\leq \|v_k\|^2 + \langle \nabla f(x_k), x_k - x^* \rangle + \frac{2\beta}{\alpha} \|\nabla f(x_k)\|^2. \end{aligned} \quad (13)$$

对比式(12)和式(13)中的各项, 可得

$$\varepsilon(k+1) - \varepsilon(k) \leq -\frac{\alpha\sqrt{s}}{8} \varepsilon(k).$$

即

$$\varepsilon(k) \leq \left(1 - \frac{\alpha\sqrt{s}}{8} \right)^k \varepsilon(0),$$

$$f(x_k) - f(x^*) \leq \varepsilon(k) \leq \left(1 - \frac{\alpha\sqrt{s}}{8} \right)^k \varepsilon(0).$$

设 $X(0) = X_0$, $\dot{X}(0) = \nabla f(X_0)$, 由于 $f \in \mathcal{S}_{\mu,L}^1$, 则有

$$f(X_0) - f(x^*) \leq \frac{L}{2} \|X_0 - x^*\|^2, \quad \|\nabla f(X_0)\| \leq L \|X_0 - x^*\|.$$

$$\begin{aligned} \varepsilon(0) &= f(X_0) - f(x^*) + \frac{1}{4} \|\nabla f(X_0)\|^2 + \frac{1}{4} \|\nabla f(X_0) + \alpha(X_0 - x^*) + \beta \nabla f(X_0)\|^2 \\ &\leq f(X_0) - f(x^*) + \left(1 + \frac{3}{4}\beta^2 \right) \|\nabla f(X_0)\|^2 + \frac{3}{4}\alpha^2 \|X_0 - x^*\|^2 \\ &= \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2 \right) L^2 + \frac{3}{4}\alpha^2 \right) \|X_0 - x^*\|^2. \end{aligned}$$

综上,

$$f(x_k) - f(x^*) \leq C \left(1 - \frac{\alpha\sqrt{s}}{8} \right)^k \|x_0 - x^*\|^2,$$

$$\text{其中 } C = \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2 \right) L^2 + \frac{3}{4}\alpha^2 \right) \|X_0 - x^*\|^2.$$

3.2. 隐式欧拉格式

第二种离散格式是隐式欧拉格式, 它的近似规则如下:

$$x_{k+1} - x_k = \sqrt{s}v_{k+1}, \sqrt{s}\nabla^2 f(x_{k+1})v_{k+1} \approx \nabla f(x_{k+1}) - \nabla f(x_k).$$

由隐式欧拉格式离散式(9)可得如下优化算法:

$$\begin{cases} x_{k+1} - x_k = \sqrt{s}v_{k+1} \\ v_{k+1} - v_k = -\alpha\sqrt{s}v_{k+1} - \beta(\nabla f(x_{k+1}) - \nabla f(x_k)) - \sqrt{s}\nabla f(x_{k+1}) \end{cases} \quad (14)$$

定理 3.2. 假设目标函数 $f \in \mathcal{S}_{\mu,L}^1$, 对于隐式欧拉格式产生的迭代点序列 $\{x_k\}_{k \in N}$, 当系数 α, β 满足

$$\begin{cases} \frac{3}{4}\alpha^2 \leq \frac{\mu}{2} \\ \beta \leq \frac{8}{3\alpha} \\ \alpha \geq 0 \end{cases}, \text{ 步长 } s \text{ 满足 } 0 \leq s \leq \frac{1}{L} \text{ 时, 收敛速率满足}$$

$$f(x_k) - f(x^*) \leq \left(\frac{1}{1 + \frac{\alpha\sqrt{s}}{4}} \right)^k C \|x_0 - x^*\|^2.$$

$$\text{其中 } C = \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2 \right) L^2 + \frac{3}{4}\alpha^2 \right) \|x_0 - x^*\|^2.$$

证明: 设 Layapunov 函数为:

$$\varepsilon(k) = f(x_k) - f(x^*) + \frac{1}{4}\|v_k\|^2 + \frac{1}{4}\|\alpha(x_k - x^*) + v_k + \beta\nabla f(x_k)\|^2.$$

$$\begin{aligned} & \varepsilon(k+1) - \varepsilon(k) \\ &= \frac{1}{4}\langle v_{k+1} - v_k, v_{k+1} + v_k \rangle + f(x_{k+1}) - f(x_k) \\ &+ \frac{1}{4}\langle \alpha(x_{k+1} - x_k) + v_{k+1} - v_k + \beta(\nabla f(x_{k+1}) - \nabla f(x_k)), \alpha(x_{k+1} + x_k - 2x^*) + v_{k+1} + v_k + \beta(\nabla f(x_{k+1}) + \nabla f(x_k)) \rangle \\ &\leq \frac{1}{2}\langle v_{k+1} - v_k, v_{k+1} \rangle - \frac{1}{4}\|v_{k+1} - v_k\|^2 + \langle \nabla f(x_{k+1}), x_{k+1} - x_k \rangle \\ &- \frac{1}{2}\langle \sqrt{s}\nabla f(x_{k+1}), \alpha(x_{k+1} - x^*) + v_{k+1} + \beta\nabla f(x_{k+1}) \rangle - \frac{1}{4}\|\sqrt{s}\nabla f(x_{k+1})\|^2 \\ &= -\frac{\alpha\sqrt{s}}{2}\|v_{k+1}\|^2 - \frac{\beta}{2\sqrt{s}}\langle \nabla f(x_{k+1}) - \nabla f(x_k), x_{k+1} - x_k \rangle \\ &- \frac{1}{4}\|\alpha\sqrt{s}v_{k+1} + \beta(\nabla f(x_{k+1}) - \nabla f(x_k)) + \sqrt{s}\nabla f(x_{k+1})\|^2 - \frac{\alpha\sqrt{s}}{2}\langle \nabla f(x_{k+1}), x_{k+1} - x^* \rangle \\ &- \frac{\beta\sqrt{s}}{2}\|\nabla f(x_{k+1})\|^2 - \frac{s}{4}\|\nabla f(x_{k+1})\|^2. \end{aligned}$$

由柯西 - 施瓦茨不等式,

$$\begin{aligned} & \|\alpha\sqrt{s}v_k + \beta(\nabla f(x_{k+1}) - \nabla f(x_k)) + \sqrt{s}\nabla f(x_k)\|^2 \\ & \leq 3\alpha^2s\|v_k\|^2 + 3\beta^2\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 + 3s\|\nabla f(x_k)\|^2. \end{aligned}$$

故

$$\begin{aligned}\varepsilon(k+1) - \varepsilon(k) &\leq -\frac{\alpha\sqrt{s}}{2} \left(\|v_{k+1}\|^2 + \langle \nabla f(x_{k+1}), x_{k+1} - x^* \rangle + \frac{\beta}{\alpha} \|\nabla f(x_{k+1})\|^2 \right) \\ &\quad - \frac{\beta}{2\sqrt{s}} \langle \nabla f(x_{k+1}) - \nabla f(x_k), x_{k+1} - x_k \rangle - \frac{3\alpha^2 s}{4} \|v_{k+1}\|^2 \\ &\quad - \frac{3\beta^2}{4} \|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 - s \|\nabla f(x_{k+1})\|^2.\end{aligned}$$

由于 $f \in \mathcal{S}_{\mu,L}^1$, 则有

$$\langle \nabla f(x_{k+1}) - \nabla f(x_k), x_{k+1} - x_k \rangle \geq \mu \|x_{k+1} - x_k\|^2.$$

故

$$\varepsilon(k+1) - \varepsilon(k) \leq -\frac{\alpha\sqrt{s}}{2} \left(\|v_{k+1}\|^2 + \langle \nabla f(x_{k+1}), x_{k+1} - x^* \rangle + \frac{\beta}{\alpha} \|\nabla f(x_{k+1})\|^2 \right). \quad (15)$$

由柯西 - 施瓦茨不等式,

$$\|\alpha(x_k - x^*) + v_k + \beta \nabla f(x_k)\|^2 \leq 3\alpha^2 \|x_k - x^*\|^2 + 3\|v_k\|^2 + 3\beta^2 \|\nabla f(x_k)\|^2.$$

故

$$\begin{aligned}\varepsilon(k+1) &\leq \|v_{k+1}\|^2 + f(x_{k+1}) - f(x^*) + \frac{3}{4}\alpha^2 \|x_{k+1} - x^*\|^2 + \frac{3}{4}\beta^2 \|\nabla f(x_{k+1})\|^2 \\ &\leq \|v_{k+1}\|^2 + f(x_{k+1}) - f(x^*) + \frac{\mu}{2} \|x_{k+1} - x^*\|^2 + \frac{2\beta}{\alpha} \|\nabla f(x_{k+1})\|^2 \\ &\leq \|v_{k+1}\|^2 + \langle \nabla f(x_{k+1}), x_{k+1} - x^* \rangle + \frac{2\beta}{\alpha} \|\nabla f(x_{k+1})\|^2.\end{aligned} \quad (16)$$

对比式(15)和式(16)中的各项, 可得

$$\varepsilon(k+1) - \varepsilon(k) \leq -\frac{\alpha\sqrt{s}}{4} \varepsilon(k+1).$$

即

$$\begin{aligned}\varepsilon(k+1) &\leq \frac{1}{1 + \frac{\alpha\sqrt{s}}{4}} \varepsilon(k), \\ f(x_k) - f(x^*) &\leq \varepsilon(k) \leq \left(\frac{1}{1 + \frac{\alpha\sqrt{s}}{4}} \right)^k \varepsilon(0).\end{aligned}$$

设 $X(0) = X_0$, $\dot{X}(0) = \nabla f(X_0)$, 由于 $f \in \mathcal{S}_{\mu,L}^1$, 则有

$$f(X_0) - f(x^*) \leq \frac{L}{2} \|X_0 - x^*\|^2, \quad \|\nabla f(X_0)\| \leq L \|X_0 - x^*\|.$$

$$\begin{aligned}\varepsilon(0) &= f(X_0) - f(x^*) + \frac{1}{4} \|\nabla f(X_0)\|^2 + \frac{1}{4} \|\nabla f(X_0) + \alpha(X_0 - x^*) + \beta \nabla f(X_0)\|^2 \\ &\leq f(X_0) - f(x^*) + \left(1 + \frac{3}{4}\beta^2\right) \|\nabla f(X_0)\|^2 + \frac{3}{4}\alpha^2 \|X_0 - x^*\|^2 \\ &= \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2\right)L^2 + \frac{3}{4}\alpha^2\right) \|X_0 - x^*\|^2.\end{aligned}$$

综上,

$$f(x_k) - f(x^*) \leq \left(\frac{1}{1 + \frac{\alpha\sqrt{s}}{4}} \right)^k C \|x_0 - x^*\|^2,$$

其中 $C = \left(\frac{L}{2} + \left(1 + \frac{3}{4}\beta^2 \right) L^2 + \frac{3}{4}\alpha^2 \right) \|x_0 - x^*\|^2$.

4. 数值实验

在本章中, 我们将对第三章中所得的算法进行数值实验, 验证算法的性能。同时使用 NAG 算法与梯度下降法作为对比, 证明算法的实用性。

由于本文考虑的目标函数 $f \in \mathcal{S}_{\mu,L}^1$, 为了方便获取函数的 μ, L 值, 我们选取了 CUTest 数据集中的无约束二次函数问题作为测试函数。由于我们提出的算法中含有不定系数 α, β , 且它们往往与目标函数的 μ, L 值相关, 对于每个测试问题, 事先将各个系数调整为合适的值。每种算法都使用相同的固定步长。

算法终止条件: $\|\nabla f(x_k)\| \leq 10^{-8}$ 或 $\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq 10^{-16}$ 。

Table 1. Comparison of average number of iterations and computation time of various algorithms when the objective function is a quadratic function

表 1. 目标函数为二次函数时, 各个算法的平均迭代次数与运算时间对比

	显式欧拉格式	隐式欧拉格式	NAG 算法	梯度下降法
平均迭代次数	8941	1507	1836	10,447
平均运行时间(s)	0.007131	0.001376	0.001427	0.007559

由表 1 可知, 隐式欧拉格式在平均迭代次数上表现明显优于其他算法, 在平均运行时间上略优于 NAG 算法, 明显优于梯度下降法, 这一结果表明隐式欧拉格式为加速梯度算法, 且数值表现优于 NAG 算法。显示欧拉格式由于受到步长的限制, 在数值实验中的表现与梯度下降法类似, 没有起到加速效果。

5. 结论

本文主要研究的是从 ODE 中得到新的加速梯度算法。本文考虑的不再是一个固定的 ODE, 而是一个系数不定的 ODE:

$$\ddot{X} + \alpha \dot{X} + \beta \nabla^2 f(X) \dot{X} + \nabla f(X) = 0,$$

通过构造合适的 Layapunov 函数, 证明了系数 α, β 在满足什么条件下, 该 ODE 具有较快的收敛速率。同时, 本文使用了两种不同的离散格式: 显式欧拉格式、隐式欧拉格式对该 ODE 进行离散, 得到了相应的优化算法, 并通过构造合适的 Layapunov 函数, 分别证明了这两种算法的收敛速率。最后, 本文对于提出的优化算法进行了初步的数值实验。分别使用 CUTest 数据集中的无约束二次函数问题作为测试集, 验证了算法的性能, 并以 NAG 算法和梯度下降法作为对比。结果表明, 当目标函数为二次函数时, 隐式欧拉格式算法的数值表现优于 NAG 算法。

参考文献

- [1] Boyd, S.P. and Vandenberghe, L. (2004) Convex Optimization. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511804441>

-
- [2] Polyak, B.T. (1964) Some Methods of Speeding Up the Convergence of Iteration Methods. *USSR Computational Mathematics and Mathematical Physics*, **4**, 1-17. [https://doi.org/10.1016/0041-5553\(64\)90137-5](https://doi.org/10.1016/0041-5553(64)90137-5)
 - [3] Nesterov, Y. (1983) A Method of Solving a Convex Programming Problem with Convergence Rate $O(1/k^2)$. *Doklady Akademii Nauk SSSR*, **269**, 543.
 - [4] Alvarez, F., Attouch, H., Bolte, J., *et al.* (2002) A Second-Order Gradient-Like Dissipative Dynamical System with Hessian-Driven Damping: Application to Optimization and Mechanics. *Journal de Mathématiques Pures et Appliquées*, **81**, 747-779. [https://doi.org/10.1016/S0021-7824\(01\)01253-3](https://doi.org/10.1016/S0021-7824(01)01253-3)
 - [5] Su, W., Boyd, S. and Candes, E.J. (2016) A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights. *Journal of Machine Learning Research*, **17**, 1-43.
 - [6] Shi, B., Du, S.S., Jordan, M.I., *et al.* (2022) Understanding the Acceleration Phenomenon via High-Resolution Differential Equations. *Mathematical Programming*, **195**, 79-148. <https://doi.org/10.1007/s10107-021-01681-8>
 - [7] Chen, L. and Luo, H. (2018) A Unified Convergence Analysis of First Order Convex Optimization Methods via Strong Lyapunov Functions. arXiv preprint arXiv:2108.00132
 - [8] Chen, L. and Luo, H. (2019) First Order Optimization Methods Based on Hessian-Driven Nesterov Accelerated Gradient Flow. arXiv preprint arXiv:1912.09276