

Exploiting Numerical Analytical Concepts for Continuous Optimization: Numerical Stability and Discrete Chain Rule

(連続最適化のための数値解析学諸概念の活用：数値的安定性と離散連鎖律)

数理情報学専攻 48216206 牛山 寛生

指導教員 松尾 宇泰 教授

1 はじめに

連続最適化手法はステップ幅 0 の極限を考えることで連続力学系と結びつけられる。この視点に着目する研究は古くから行われていたが、近年 Su–Boyd–Candès [3] が加速勾配法に対して連続極限 ODE を導出して以降、特に活発になっており、様々な新しい ODE が考案され、その離散化として最適化手法が導出されている。本研究では最適化に対する力学系アプローチに数値解析学を取り込むことを目指す。本研究は 3 つの内容から構成されているが、本要旨ではそのうち 1 つを重点的に説明し、残りの 2 つは発表で詳細に扱う。

2 安定な数値解法の最適化への応用

2.1 概要

最適化手法の最急降下法は勾配流に対する陽的 Euler 法と解釈できる。本研究では陽的 Euler 法よりも安定性の高い数値解法である Runge–Kutta–Chebyshev 法 (RKC) を勾配流に適応することで、刻み幅の制限を緩和し、より速く収束する最適化手法の構成を試みる。

本研究の貢献は、RKC を最適化手法として使うことを提案したこと、RKC と最適化手法の Chebyshev 準反復の関係を描いたこと、および最適化のための RKC のステップ幅適応法を連続 Runge–Kutta 法を用いて構成したことである。しかし、RKC を最適化手法として活用している先行研究が見つかる [1]。この研究と本研究の関係は節末で説明する。

2.2 準備：Chebyshev–Runge–Kutta 法 (cf. [2])

定義 2.1 (線形安定性解析 (cf. [2])). 数値解法を Dahlquist のテスト方程式 $\dot{y} = \lambda y$, $y(0) = 1$ に適用したときの値を $R(h\lambda)$ と定めるとき、 R を安定性関数、集合 $\{z \in \mathbb{C} \mid |R(z)| \leq 1\}$ を安定領域という。

刻み幅 h は $z = h\lambda$ が安定領域に含まれるように取る必要がある。上記の λ はベクトル場の Jacobian の固有値に相当する。凸関数に対する勾配流は、その値が非負の実数であるため、負の実軸を広く含む安定領域が勾配流の離散化に適している。このような安定性関数は以

下のように構成できる。

$$R_s(z) := \frac{1}{T_s(w_0)} T_s(w_0 + w_1 z),$$

$$w_0 := 1 + \frac{\varepsilon}{s^2}, \quad w_1 := \frac{T_s(w_0)}{T'_s(w_0)}.$$

ここで T_s は s 次 Chebyshev 多項式であり、 $\varepsilon > 0$ は事前に与えられるパラメータである。対応する安定領域は図 1 のような形状をしている。

安定性関数 R_s を実現する数値解法は、 s 段の陽的 Euler 法の合成である因数分解法やその改善である Lebedev の方法、Chebyshev 多項式の 3 項漸化式を用いる Houwen–Sommeijer の方法 (HS 法) がある。

2.3 RKC の最適なパラメータと Chebyshev 準反復との関係

μ -強凸かつ L -平滑な 2 次関数 $f(x) = x^\top Ax/2 + b^\top x$ に関する勾配流 $\dot{x} = -\nabla f(x)$ に RKC を適用したとき、 $x^{(k+1)} - x^* = R_s(-hA)(x^{(k)} - x^*)$ が成立する。ここで $x^{(k)}$ は数値解、 x^* は最適解である。 s を固定し、 $\max_{\mu \leq \lambda \leq L} R_s(-h\lambda)$ を最小にする ε, h は次のように求める：

$$(w_1 h)^* = \frac{2}{L - \mu}, \quad \frac{\varepsilon^*}{s^2} = \frac{2\mu}{L - \mu}.$$

実はこのパラメータによる HS 法の RKC は Chebyshev 準反復の最初の s 回の反復と一致する。Chebyshev 準反復とは強凸 2 次関数に対して最適の最悪ケース収束レートをもつ最適化手法である。したがって、RKC は Chebyshev 準反復の数値解法的対応物と解釈できる。段数無限大の極限で RKC の収束率は Chebyshev 準反復と一致する：

$$\lim_{s \rightarrow \infty} (R_s(-hA))^{-1/s} = \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}. \quad (1)$$

2.4 連続 Runge–Kutta 法による曲線探索

本研究ではステップ幅の適応により RKC を 2 次関数以外に適用することを考える。通常の直線探索では失敗することがあるため、連続 Runge–Kutta 法を用いて「曲線探索」を行う。連続 Runge–Kutta 法は Runge–Kutta 法の出力を小さい計算コストで連続的に補間す

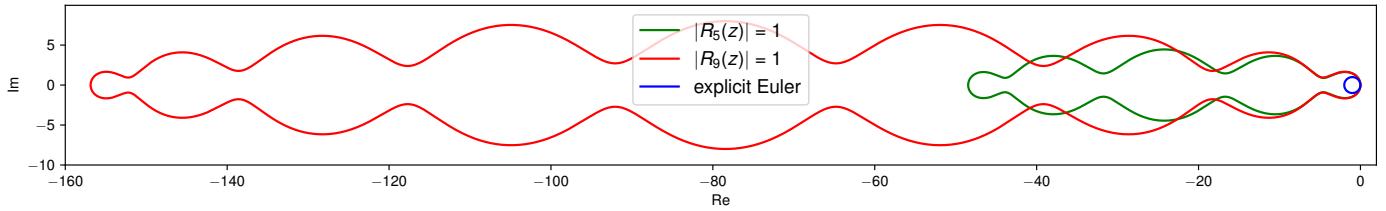


図 1. 陽的 Euler 法の安定領域と安定性関数に R_5, R_9 を持つ数値解法の安定領域.

る手法である. これにより RKC の 1 ステップの間を曲線で補完し, その上をバックトラックすることでステップ幅を適応させる (Algorithm 1). いくつかの最適化問題に本手法を適用した. Cahn–Hilliard 型ポテンシャルエネルギー最小化問題に適用した結果を図 2 に示す.

Algorithm 1 曲線探索付き RKC

Input: $f : \mathbb{R}^n \rightarrow \mathbb{R}, \nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n, x_0 \in \mathbb{R}^n, \varepsilon > 0, h_0 > 0, \text{tol} > 0$
1: $k \leftarrow 0, h \leftarrow h_0, x \leftarrow x_0$
2: **while** $\|\nabla f(x_k)\| \geq \text{tol}$ **do**
3: $f_b \leftarrow f(x)$
4: $p \leftarrow \nabla f(x)$
5: $x_{\text{temp}} \leftarrow \text{RKC}_{\varepsilon, h}(x)$
6: $u \leftarrow x_{\text{temp}}$
7: $\theta \leftarrow 1$
8: **while** $f(u) > f_b + 0.5\langle p, x_{\text{temp}} - x \rangle$ **do**
9: $\theta \leftarrow 0.7\theta$
10: $u \leftarrow x - (\theta - \theta^2)hp + \theta^2h(x_{\text{temp}} - x)$
11: **end while**
12: $x \leftarrow u$
13: $k \leftarrow k + 1$
14: **if** $\theta = 1$ **then**
15: $h \leftarrow 1.1h$
16: **else**
17: $h \leftarrow 0.7h$
18: **end if**
19: **end while**
Output: $x \in \mathbb{R}^n$

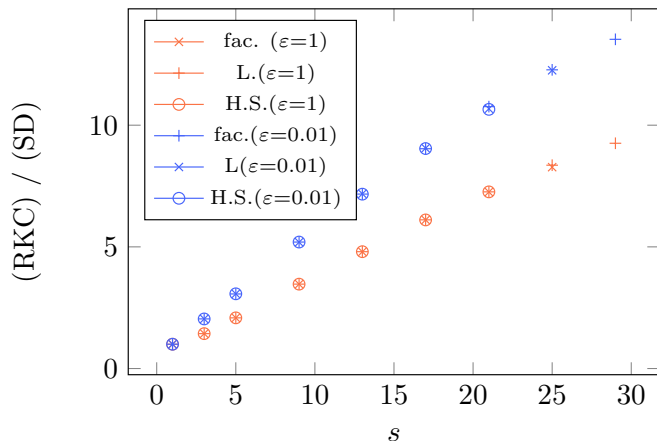


図 2. Cahn–Hilliard 型エネルギー最小化問題に Algorithm 1 を適用した結果. 縦軸は直線探索付き最急降下法 (SD) と Algorithm 1 の勾配計算回数との比である. fac. は因数分解法, L. は Lebedev の方法, H.S. は HS 法を指す.

2.5 先行研究との関係

先行研究 [1] では, HS 法による RKC を強凸 2 次関数に対する最適化手法として活用し, さらに強凸 2 次関数を摂動した非 2 次関数に対する応用も考えている. この研究では, 第 2.3 節とは異なり, 最適な RKC パラメータを, ε を固定し s, h について求めており, その結果, 式 (1) に対応するレートが準最適になっている. また, 本研究では先行研究とは異なり, ステップ幅適応により非 2 次関数への応用を提案している. 適応範囲が広いが, 理論保証は未だつけられていない.

3 ODE の本質的収束レート

連続系の収束レートは時間スケール変換でいくらかも加速できてしまう. これにより, ODE に収束レートが定義できず, 最適化手法との対応関係が考えられなくなってしまう問題があった. 本研究では数値解法の安定性を援用して, ODE に対して「本質的収束レート」を定義し, レートに不定性を除去した.

4 弱離散勾配による最適化手法の統一的記述

連続系の収束証明は Lyapunov 関数を用いて簡単に行える [4] が, 離散系の収束解析は各スキームに依存して個別に行われていた. 本研究では構造保存数値解法の離散勾配法を拡張し, 弱離散勾配を定義することで, 種々の ODE に対する様々な離散化を統一的に記述・解析した. 弱離散勾配によるスキームは既存の最適化手法や離散勾配法による手法を含んでいる.

References

- [1] A. Eftekhari, B. Vandereycken, G. Vilmart, and K. C. Zygalakis, “Explicit stabilised gradient descent for faster strongly convex optimisation,” *BIT*, vol. 61, no. 1, pp. 119–139, 2021 (cit. on pp. 1, 2).
- [2] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations. II*, ser. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 2010, vol. 14 (cit. on p. 1).
- [3] W. Su, S. Boyd, and E. J. Candès, “A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights,” *J. Mach. Learn. Res.*, vol. 17, Paper No. 153, 43, 2016 (cit. on p. 1).
- [4] A. C. Wilson, B. Recht, and M. I. Jordan, “A Lyapunov analysis of accelerated methods in optimization,” *J. Mach. Learn. Res.*, vol. 22, Paper No. 113, 34, 2021 (cit. on p. 2).