連続最適化および関連分野に関する夏季学校

連続最適化への応用に向けた
常微分方程式の数値解析入門

佐藤 峻

東京大学

2023 年 8 月 9 日

はじめの例

$$x^{(k+1)} = x^{(k)} - h_k \nabla f\left(x^{(k)}\right)$$

連続最適化の視点

最小化問題 $\min f(x)$ に対する最急降下法

- 目的：
  最適解の発見
- h_k：
  関数値等を基準に選ぶ
- キーワード：
  収束条件，収束レート，…

数値解析の視点

勾配流 $\dot{x} = -\nabla f(x)$ に対する陽的 Euler 法

- 目的：
  解軌道の忠実な再現
- h_k：
  誤差や安定性を基準に選ぶ
- キーワード：
  収束性，安定性，精度，…

目的は違うが，全く同じ更新式が使われている

$\longrightarrow$ 数値解析の視点は連続最適化に有用である (こともある)

連続最適化と微分方程式の数値解析

連続最適化問題

例：無制約最適化問題

$$\min_{x\in\mathbb{R}^d} f(x)$$

微分方程式

例：勾配流

$$\dot{x} = -\nabla f(x)$$

反復による
最適解の探索

連続極限

離散化

最適化手法

例：最急降下法

$$x^{(k+1)} = x^{(k)} - h_k \nabla f\left(x^{(k)}\right)$$

Nesterov の加速勾配法，Newton 法，...

対応

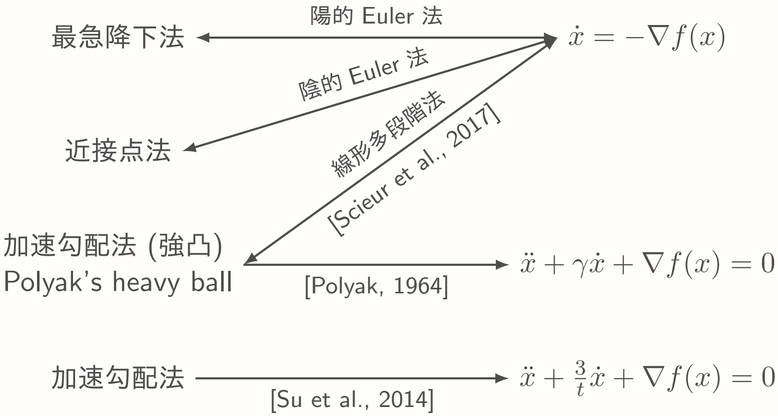
数値解法

例：陽的 Euler 法

$$x^{(k+1)} = x^{(k)} - h_k \nabla f\left(x^{(k)}\right)$$

Runge–Kutta 法，線形多段階法，...

最適化手法と常微分方程式 (ODE) の関係の例



右向きの矢印：連続極限
左向きの矢印：既存の数値解法としての解釈

ODE アプローチのご利益の例 1：新手法の構成

連続極限の方が扱いやすいことも多く，手法の設計に役立つ．

- 加速鏡像降下法 [Krichene et al., 2015]

$$\begin{array}{ccc} & \text{鏡像降下 ODE} \quad \text{加速勾配 ODE} & \\ & \searrow \quad \swarrow & \\ \text{加速鏡像降下法} & \xleftarrow{\text{離散化}} & \text{加速鏡像降下 ODE} \end{array}$$

- 凸と強凸に対する Nesterov の加速勾配法の統合 [Kim and Yang, 2023]

$$\begin{array}{ccc} & \text{加速勾配 ODE (凸)} \quad \text{加速勾配 ODE (強凸)} & \\ & \searrow \quad \swarrow & \\ \text{新たな加速勾配法} & \xleftarrow{\text{離散化}} & \text{加速勾配 ODE (統合)} \end{array}$$

ODE アプローチのご利益の例 2：「加速」の説明

[Scieur et al., 2017]

目的関数 $f\colon \mathbb{R}^d \to \mathbb{R}$ が μ-強凸 (かつ L-平滑) の場合 ($f^\star$：最適値)

勾配流 $\dot{x} = -\nabla f(x)$

$f(x(t)) - f^\star = \mathrm{O}\left(e^{-2\mu t}\right)$

陽的 Euler 法

$t_k = \mathrm{O}\left(\frac{k}{L}\right)$

最急降下法

$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(e^{-2\frac{\mu}{L}k}\right)$

線形多段階法

$t_k = \mathrm{O}\left(\frac{k}{\sqrt{\mu L}}\right)$

加速勾配法

$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(e^{-\sqrt{\frac{\mu}{L}}k}\right)$

「加速」の (一つの) 解釈：
安定性の向上による
ステップ幅制限の緩和

数値解析でよく知られている「安定性」の条件から，

- 陽的 Euler 法：$h_k = \mathrm{O}\left(\frac{1}{L}\right) \to t_k := \sum_{i=0}^{k-1} h_i = \mathrm{O}\left(\frac{k}{L}\right)$,
- 線形多段階法：$h_k = \mathrm{O}\left(\frac{1}{\sqrt{\mu L}}\right) \to t_k = \mathrm{O}\left(\frac{k}{\sqrt{\mu L}}\right)$.

今日の目標

常微分方程式の数値解析の基礎をざっくりと理解する.

連続最適化に役立ちつつある概念

- 各種の陽的数値解法
  陽的：次のステップの数値解が直接計算できる.
- 安定性
  安定：数値解 (あるいは混入した誤差) が発散しない.
- 構造保存数値解法
  微分方程式のもつ構造を再現する数値解法.

現状ではあまり有効活用されていない概念

- 各種の陰的数値解法
  陰的：次のステップの数値解を計算するために，非線形方程式を解く.
- 収束性 (精度)
  ステップ幅を小さくした際の厳密解への近づき方.

- E. Hairer, S. P. Nørsett, and G. Wanner. Solving ordinary differential equations I, Nonstiff problems, Springer-Verlag, 1993, xvi+528 pp.
- E. Hairer and G. Wanner. Solving ordinary differential equations II, Stiff and differential-algebraic problems, Springer-Verlag, 1996, xvi+614 pp.

コメント： 網羅的に ODE の数値解析の知見がまとめられている．数値実験例も豊富．和訳も「常微分方程式の数値解法 I」「常微分方程式の数値解法 II」という名前で出版されている．

- 三井 斌友．常微分方程式の数値解法，岩波書店，2003，146 pp.

コメント：常微分方程式の数値解析の基礎がコンパクトにまとめられている．演習問題に解答もついており，親切．

- E. Hairer, C. Lubich, G. Wanner. Geometric Numerical Integration, Springer-Verlag, 2006, xviii+644 pp.

コメント：ODE に対する構造保存数値解法 (後述) の定番教科書．

Table of contents

加速勾配法の連続極限 [Su et al., 2014]

Nesterov の加速勾配法

$$\begin{cases} x^{(k+1)} = y^{(k)} - s\nabla f\left(y^{(k)}\right) \\ y^{(k+1)} = x^{(k+1)} + \frac{k}{k+3}\left(x^{(k+1)} - x^{(k)}\right) \end{cases}$$

収束レート：

$$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(\frac{1}{k^2}\right)$$

$\rightarrow$

連続極限

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

収束レート：

$$f(x(t)) - f^\star = \mathrm{O}\left(\frac{1}{t^2}\right)$$

- 「加速」の直感的な理解
  - $\frac{3}{t}\dot{x}$ は「摩擦項」
  - 連続系の収束レートの証明は簡潔 (後述)
- Nesterov の加速勾配法の変種の導出
- 鏡像降下法との組合せ [Krichene et al., 2015]
- 強凸版との統合 [Kim and Yang, 2023]

連続極限の導出

Nesterov の加速勾配法

$$\begin{cases} x^{(k+1)} = y^{(k)} - s\nabla f\left(y^{(k)}\right) \\ y^{(k+1)} = x^{(k+1)} + \frac{k}{k+3}\left(x^{(k+1)} - x^{(k)}\right) \end{cases}$$

連続極限

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

↓ 第二式の時刻をずらす

$$\begin{cases} x^{(k+1)} = y^{(k)} - s\nabla f\left(y^{(k)}\right) \\ y^{(k)} = x^{(k)} + \frac{k-1}{k+2}\left(x^{(k)} - x^{(k-1)}\right) \end{cases}$$

↓ 二式の和

$$x^{(k+1)} = x^{(k)} + \frac{k-1}{k+2}\left(x^{(k)} - x^{(k-1)}\right) - s\nabla f\left(y^{(k)}\right)$$

連続極限の導出

Nesterov の加速勾配法

$$\begin{cases} x^{(k+1)} = y^{(k)} - s\nabla f\left(y^{(k)}\right) \\ y^{(k+1)} = x^{(k+1)} + \frac{k}{k+3}\left(x^{(k+1)} - x^{(k)}\right) \end{cases}$$

連続極限

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

$$\downarrow$$

$$x^{(k+1)} = x^{(k)} + \frac{k-1}{k+2}\left(x^{(k)} - x^{(k-1)}\right) - s\nabla f\left(y^{(k)}\right)$$

$$\downarrow \ \tfrac{k-1}{k+2} = 1 - \tfrac{3}{k+2}$$

$$\frac{x^{(k+1)} - 2x^{(k)} + x^{(k-1)}}{s} + \frac{3}{(k+2)\sqrt{s}} \frac{x^{(k)} - x^{(k-1)}}{\sqrt{s}} + \nabla f\left(y^{(k)}\right) = 0$$

連続極限の導出

Nesterov の加速勾配法

$$\begin{cases} x^{(k+1)} = y^{(k)} - s\nabla f\left(y^{(k)}\right) \\ y^{(k+1)} = x^{(k+1)} + \frac{k}{k+3}\left(x^{(k+1)} - x^{(k)}\right) \end{cases}$$

連続極限

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

$\downarrow$

$\uparrow \ \sqrt{s} \to 0$

$$\underbrace{\frac{x^{(k+1)} - 2x^{(k)} + x^{(k-1)}}{s}}_{\ddot{x}} + \underbrace{\frac{3}{(k+2)\sqrt{s}}}_{\frac{3}{t}}\underbrace{\frac{x^{(k)} - x^{(k-1)}}{\sqrt{s}}}_{\dot{x}} + \nabla f\Big(\underbrace{y^{(k)}}_{x}\Big) = 0$$

(この離散化は不自然)

$\boxed{x^{(k)} \approx x(k\sqrt{s}) \text{ という想定}}$

連続系の収束レート

以下の関数を用いて，収束レートを示す ($x^\star$：最適解)：

$$\mathcal{E}(t) := t^2(f(x(t)) - f^\star) + 2\left\|x(t) + \frac{t}{2}\dot{x}(t) - x^\star\right\|^2$$

1. $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(t) \le 0$：$\mathcal{E}(t)$ は単調非増加．

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(t) &= 2t(f(x) - f^\star) + t^2\langle\nabla f(x), \dot{x}\rangle + 4\left\langle x + \frac{t}{2}\dot{x} - x^\star, \underbrace{\frac{3}{2}\dot{x} + \frac{t}{2}\ddot{x}}_{= -\frac{t}{2}\nabla f(x)}\right\rangle \\ &= 2t(f(x) - f^\star) - 2t\langle\nabla f(x), x - x^\star\rangle \\ &\le 0\end{aligned}$$

← 凸関数は接線で下から抑えられる

(前頁の ODE)

2. 収束レートの証明：

$$f(x(t)) - f^\star \le \frac{\mathcal{E}(t)}{t^2} \le \frac{\mathcal{E}(0)}{t^2} = \frac{2\|x(0) - x^\star\|^2}{t^2}$$

再掲：加速勾配法の連続極限 [Su et al., 2014]

Nesterov の加速勾配法

$$\begin{cases} x^{(k+1)} = y^{(k)} - s\nabla f\left(y^{(k)}\right) \\ y^{(k+1)} = x^{(k+1)} + \frac{k}{k+3}\left(x^{(k+1)} - x^{(k)}\right) \end{cases}$$

収束レート：

$$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(\frac{1}{k^2}\right)$$

$\rightarrow$

連続極限

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

収束レート：

$$f(x(t)) - f^\star = \mathrm{O}\left(\frac{1}{t^2}\right)$$

- 「加速」の直感的な理解
  - $\frac{3}{t}\dot{x}$ は「摩擦項」
  - 連続系の収束レートの証明は簡潔
- Nesterov の加速勾配法の変種の導出
- 鏡像降下法との組合せ [Krichene et al., 2015]
- 強凸版との統合 [Kim and Yang, 2023]

Table of contents

常微分方程式 (ODE: Ordinary Differential Equation)

常微分方程式の初期値問題 ($x\colon [0,T) \to \mathbb{R}^d$, $\dot{x} \coloneqq \frac{\mathrm{d}x}{\mathrm{d}t}$)

$$\dot{x}(t) = g(t, x(t)) \quad (t \in (0,T)), \qquad x(0) = x_0.$$

$g(t,\cdot)$ が局所 Lipschitz 連続であれば，解は一意に存在する．

古典的な例

- 調和振動子：

$$\begin{cases} \dot{p} = kx, \\ m\dot{x} = p. \end{cases}$$

- Lotka–Volterra 方程式：

$$\begin{cases} \dot{x} = ax - bxy, \\ \dot{y} = cxy - dy. \end{cases}$$

近年の例

- 加速勾配法の連続極限 [Su et al., 2014]：

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0.$$

- Neural ODE [Chen et al., 2018]：

$$\dot{x} = f_{\mathrm{NN}}(\theta, x, t).$$

最も単純な数値解法：陽的 Euler 法

陽的 Euler 法：

$$\frac{x^{(k+1)} - x^{(k)}}{h_k} = g\left(t_k, x^{(k)}\right) \quad (k = 0, 1, \dots), \qquad x^{(0)} = x_0.$$

cf. ODE

$$\dot{x}(t) = g(t, x(t)) \quad (t \in (0, T)), \qquad x(0) = x_0.$$

適切な仮定の下，

$$\left\| x^{(k)} - x(t_k) \right\| \leq C h_{\max} \qquad \left(h_{\max} := \max_i h_i \right).$$

- ステップ幅 h_k が十分小さいとき，$x^{(k)}$ は $x(t_k)$ の良い近似．
- 右辺が $h_{\max}$ について 1 次なので，<u>1 次精度</u>という．

より高度な数値解法

陽的 Euler 法：

$$\frac{x^{(k+1)} - x^{(k)}}{h_k} = g\left(t_k, x^{(k)}\right) \quad (k = 0, 1, \dots), \qquad x^{(0)} = x_0.$$

1 ステップの中に
内部段を作る

過去の情報の利用
($x^{(k-1)}$ など)

Runge–Kutta 法

- Heun 法
- いわゆる Runge–Kutta 法
- Dormand–Prince の方法
  (MATLAB の ode45,
  scipy.integrate.solve_ivp 等)

線形多段階法

- Adams–Bashforth 公式
- Adams–Moulton 公式
- 後退微分公式
  (BDF: Backward
  Differentiation Formula)

これらの共通の一般化 (一般線形法) や，これらに属さない手法もある．

定義： Runge–Kutta 法

$\dot{x} = g(t, x)$ に対する s 段 Runge–Kutta 法：

$$\begin{cases} X_i^{(k)} = x^{(k)} + h \sum_{j=1}^{s} a_{ij} g\left(t_k + c_j h, X_j^{(k)}\right) \quad (i = 1, \ldots, s), \\ x^{(k+1)} = x^{(k)} + h \sum_{i=1}^{s} b_i g\left(t_k + c_i h, X_i^{(k)}\right). \end{cases}$$

$X_i^{(k)}$ ：内部段

a_{ij}, b_i, c_i ：パラメータ

例：陽的 Euler 法

$$\frac{x^{(k+1)} - x^{(k)}}{h} = g\left(t_k, x^{(k)}\right) \quad \longrightarrow \quad \begin{cases} X_1^{(k)} = x^{(k)} + h \cdot 0 \cdot g\left(t_k, X_1^{(k)}\right) \\ x^{(k+1)} = x^{(k)} + h \cdot 1 \cdot g\left(t_k, X_1^{(k)}\right) \end{cases}$$

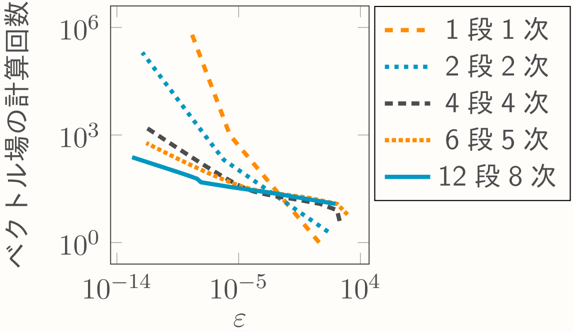
拡張のご利益 (の例)：計算量の改善

問題設定：時刻 $T > 0$ の厳密解 $x(T)$ の誤差 $\varepsilon > 0$ 以内の近似の計算
(簡単のため，ステップ幅 $h := T/K$ は固定．)

- 陽的 Euler 法：ベクトル場の計算回数 $\mathrm{O}\left(\frac{1}{\varepsilon}\right)$
  $\because$ 誤差評価 $\left\|x^{(K)} - x(T)\right\| \le Ch = C\frac{T}{K} = \varepsilon \quad \Rightarrow \quad K = \frac{CT}{\varepsilon}$
- p 次精度の Runge–Kutta 法：ベクトル場の計算回数 $\mathrm{O}\left(\left(\frac{1}{\varepsilon}\right)^{\frac{1}{p}}\right)$
  $\because$ 誤差評価 $\left\|x^{(K)} - x(T)\right\| \le C_p h^p = C_p\left(\frac{T}{K}\right)^p = \varepsilon \Rightarrow K = T\left(\frac{C_p}{\varepsilon}\right)^{\frac{1}{p}}$

数値例：
$\dot{x} = -x$, $T = 10$.

要求精度が小さいとき，
次数が高いほど有利．



Butcher 配列

Runge–Kutta 法のパラメータは，[Butcher, 1964] に従って以下のように表示することが多い：

$$
\begin{array}{c|ccc}
c_1 & a_{11} & \cdots & a_{1s} \\
\vdots & \vdots & & \vdots \\
c_s & a_{s1} & \cdots & a_{ss} \\
\hline
 & b_1 & \cdots & b_s
\end{array}
$$

再掲：Runge–Kutta 法

$\dot{x} = g(t, x)$ に対する s 段 Runge–Kutta 法：

$$
\begin{cases}
X_i^{(k)} = x^{(k)} + h \displaystyle\sum_{j=1}^{s} a_{ij} g\left(t_k + c_j h, X_j^{(k)}\right) \quad (i = 1, \ldots, s), \\
x^{(k+1)} = x^{(k)} + h \displaystyle\sum_{i=1}^{s} b_i g\left(t_k + c_i h, X_i^{(k)}\right).
\end{cases}
$$

陽的 Runge–Kutta 法と陰的 Runge–Kutta 法

定義：陽的 Runge–Kutta 法

Runge–Kutta 法が陽的 $\iff$ 任意の $i \le j$ に対して，$a_{ij} = 0$.
そうでない場合，陰的 Runge–Kutta 法と呼ばれる.

Heun 法 (陽的)：

$$X_2^{(k)} = x^{(k)} + h \cdot 0 \cdot g\left(t_k, X_1^{(k)}\right) + h \cdot 0 \cdot g\left(t_k + h, X_2^{(k)}\right)$$

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}$$

$\longrightarrow$ $X_1^{(k)}$ は陽的に計算可能

$\longrightarrow$ $X_2^{(k)}$ は陽的に計算可能

$$X_2^{(k)} = x^{(k)} + h \cdot 1 \cdot g\left(t_k, X_1^{(k)}\right) + h \cdot 0 \cdot g\left(t_k + h, X_2^{(k)}\right)$$

陰的台形則 (陰的):

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

$\longrightarrow$ $X_1^{(k)}$ は陽的に計算可能

$\longrightarrow$ $X_2^{(k)}$ は陽的に計算できない

$$X_2^{(k)} = x^{(k)} + h \cdot \frac{1}{2} \cdot g\left(t_k, X_1^{(k)}\right) + h \cdot \frac{1}{2} \cdot g\left(t_k + h, X_2^{(k)}\right)$$

Runge–Kutta 法の例

- 1 次の方法:

陽的 Euler 法

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

陰的 Euler 法

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$$

- 2 次の方法:

陰的台形則

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

陰的中点則

$$\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array}$$

Heun 法

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}$$

- いわゆる Runge–Kutta 法 (4 次)

$$\begin{array}{c|cccc} 0 & & & & \\ 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ 1 & 0 & 0 & 1 & \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array}$$

線形多段階法

定義：線形多段階法 (固定刻み幅)

$\dot{x} = g(t, x)$ に対する 線形 r 段階法：

$$\begin{aligned}\alpha_r x^{(k+r)} &+ \alpha_{r-1} x^{(k+r-1)} + \cdots + \alpha_0 x^{(k)} \\ &= h(\beta_r g_{k+r} + \beta_{r-1} g_{k+r-1} + \cdots + \beta_0 g_k), \\ g_k &:= g\left(t_k, x^{(k)}\right).\end{aligned}$$

α_i, β_i：パラメータ ($\alpha_r \neq 0$, $|\alpha_0| + |\beta_0| \neq 0$)
陽的 $\iff$ $\beta_r = 0$,　　陰的 $\iff$ $\beta_r \neq 0$

例：陽的 Euler 法

$$\frac{x^{(k+1)} - x^{(k)}}{h} = g\left(t_k, x^{(k)}\right)$$

$$\longrightarrow \quad 1 \cdot x^{(k+1)} + (-1) \cdot x^{(k)} = h\left(0 \cdot g\left(t_{k+1}, x^{(k+1)}\right) + 1 \cdot g\left(t_k, x^{(k)}\right)\right)$$

出発値

定義：線形多段階法 (固定刻み幅)

$$\sum_{i=0}^{r} \alpha_i x^{(k+i)} = h \sum_{i=0}^{r} \beta_i g\left(t_{k+i}, x^{(k+i)}\right).$$

線形多段階法を用いるには，初期値 $x^{(0)}$ だけではなく，
出発値 $x^{(1)}, \ldots, x^{(r-1)}$ も事前に用意しておく必要がある．

$x^{(i)} \approx x(ih)$ の計算法の例：

- 厳密解の Taylor 展開
- Runge–Kutta 法
- 低段階の多段階法

出発値が満たすべき条件は後述．

線形多段階法の例

定義：線形多段階法 (固定刻み幅)

$$\sum_{i=0}^{r} \alpha_i x^{(k+i)} = h \sum_{i=0}^{r} \beta_i g\Bigl(t_{k+i}, x^{(k+i)}\Bigr).$$

- Adams 型：積分表示の右辺を Lagrange 補間を用いて近似.

積分表示：
$$x(t_{k+r}) - x(t_{k+r-1}) = \int_{t_{k+r-1}}^{t_{k+r}} g(t, x(t))\mathrm{d}t.$$

  - Adams–Bashforth 公式：陽的 ($x^{(k)}, \ldots, x^{(k+r-1)}$ を用いる).
  - Adams–Moulton 公式：陰的 ($x^{(k)}, \ldots, x^{(k+r)}$ を用いる).
- 後退微分公式 (BDF)：ODE の左辺を高次の片側差分で近似.

$$\dot{x}(t_{k+r}) = g\Bigl(t_{k+r}, x^{(k+r)}\Bigr)$$

Table of contents

最急降下法の収束条件の数値解析的解釈

目的関数は L-平滑 (勾配が L-Lipschitz 連続) とする.

凸関数に対する最急降下法の収束条件 cf. [Nesterov, 2004, Thm. 2.1.14]

最急降下法は，$h_k < 2/L$ のとき $f\left(x^{(k)}\right) \to f^\star$ を満たす.

$\downarrow$ 数値解析の言葉で言い換え

勾配流に対する陽的 Euler 法 (= 最急降下法) は，
$h_k < 2/L$ のもとで平衡点に収束する.

数値解析の手法で同様の条件は出るか？

$\downarrow$

線形安定性解析

Dahlquist のテスト方程式と安定領域

定義：安定性と安定領域

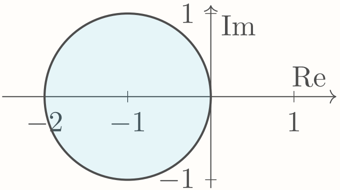
Dahlquist のテスト方程式 ($\lambda \in \mathbb{C}$ は定数)

$$\dot{x} = \lambda x, \qquad x(0) = 1$$

に対して 1 ステップ数値解法を適用した場合の出力を $R(h\lambda)$ とする．このとき，R を安定性関数，$S = \{z \in \mathbb{C} \mid |R(z)| \leq 1\}$ を安定領域という．

陽的 Euler 法の安定性関数は $R(z) = 1 + z$

安定領域：
$(|\lambda h + 1| \leq 1)$



安定領域の意味

定義：安定性と安定領域

Dahlquist のテスト方程式 ($\lambda \in \mathbb{C}$ は定数) $\dot{x} = \lambda x,\ x(0) = 1$ に対して 1 ステップ数値解法を適用した場合の出力を $R(h\lambda)$ とする．このとき，R を安定性関数，$S = \{z \in \mathbb{C} \mid |R(z)| \leq 1\}$ を安定領域という．

与えられたパラメータ $\lambda \in \mathbb{C}$ に対して，
ステップ幅 h で数値解法を適用したとき，

$$x^{(k)} = (R(h\lambda))^k x^{(0)}.$$

$\Rightarrow$ 「$\sup_k \left| x^{(k)} \right| < \infty \iff |R(h\lambda)| \leq 1$」

安定領域の意味

与えられたパラメータ $\lambda \in \mathbb{C}$ に対して，$h\lambda \in S$ を満たすようにステップ幅 h を取れば，数値解は有界に留まる．

安定領域の使い方の例

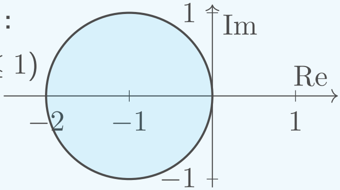
ODE：勾配流

- 勾配流 $\dot{x} = -\nabla f(x)$ の線形化：$\dot{\delta} = -\nabla^2 f(x)\delta$
- f が L-平滑な凸関数 $\Rightarrow$ $\nabla^2 f(x)$ の固有値は $[0, L]$ に含まれる.
- $\nabla^2 f(x)$ の対角化により，d 個の独立なスカラー線形方程式 (Dahlquist のテスト方程式) に分離できる.
- 分離後の各 ODE について，$\lambda \in [-L, 0]$.

数値解法：陽的 Euler 法の安定領域

安定領域：

$(|\lambda h + 1| \leq 1)$



安定領域の意味：

$$\underline{\forall \lambda \in [-L, 0],\ \lambda h \in S} \Rightarrow \text{安定}$$

$$\Updownarrow$$

$$h \leq \frac{2}{L}$$

(最急降下法のよく知られているステップ幅上限を再現).

Runge–Kutta (RK) 法の安定性関数

命題：RK 法の安定性関数 (cf. [Hairer and Wanner, 1996, IV, Prop. 3.2]

$A = (a_{ij})$, $b = (b_i)$ とおくと，RK 法の安定性関数 R は以下を満たす：

$$R(z) = \frac{\det(I - zA + z\mathbf{1}b^\top)}{\det(I - zA)}.$$

($I \in \mathbb{R}^{s\times s}$：単位行列，$\mathbf{1} \in \mathbb{R}^s$：全要素が 1 のベクトル)

- 陽的 RK の場合 (A が狭義下三角)，R は s 次多項式.
- 陰的 RK の場合，R は有理関数
  (分母，分子ともに s 次以下の多項式).

Dahlquist のテスト方程式に，RK 法を適用すると，直ちに

$$R(z) = 1 + zb^\top(I - zA)^{-1}\mathbf{1}$$

が得られる．上の命題は Schur 補行列を使えば容易に示せる (演習問題).

陽的 Runge–Kutta 法の安定領域

s 段 p 次陽的 Runge–Kutta 法の安定性関数

- 安定性関数は s 次多項式
- Dahlquist のテスト方程式の厳密解 $x(h) = \exp(\lambda h)$
  $\to R(z)$ の p 次以下の係数は $\exp(z)$ の Taylor 展開と一致.
- $p = s$ の場合 $(s \leq 4)$, $R(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^s}{s!}$.

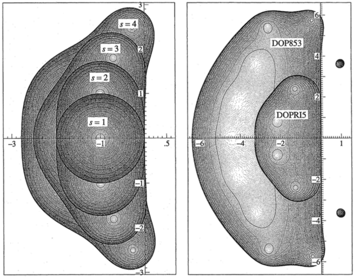


Fig. 2.1. Stability domains for explicit Runge-Kutta methods of order $p = s$

Fig. 2.2. Stability domains for DOPRI methods

[Hairer and Wanner, 1996] より引用.

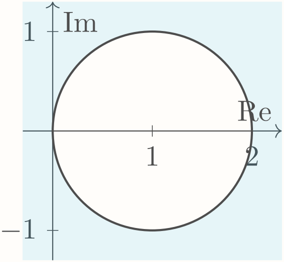
A 安定性

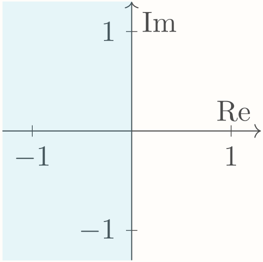
Dahlquist のテスト方程式 $\dot{x} = \lambda x$ の厳密解 $x(t) = \exp(\lambda t)$ が有界
$\iff \lambda \in \mathbb{C}^- := \{z \in \mathbb{C} \mid \mathrm{Re}\, z \leq 0\}$

定義：A 安定性

数値解法が A 安定 $\iff$ 安定性領域 S が $S \supseteq \mathbb{C}^-$ を満たす．

- A 安定な数値解法：
  陰的 Euler 法 (1 段 1 次)，Gauss 法 (s 段 $2s$ 次)，…





- 1 次以上の精度をもつ陽的 Runge–Kutta 法は A 安定にならない．

線形多段階法の安定性

線形多段階法には，よく考えられる安定性が 2 種類ある：

- 絶対安定性：Dahlquist のテスト方程式に対する安定性
- 零安定性：$\dot{x} = 0$ に対する安定性 $\leftarrow$ 多段階法特有の問題

線形 r 段階法は，$r+1$ 項間漸化式なので，
安定性の議論において，特性多項式が有用．

定義：特性多項式

第一特性多項式：$\rho(\zeta) \coloneqq \alpha_r \zeta^r + \alpha_{r-1}\zeta^{r-1} + \cdots + \alpha_0,$

第二特性多項式：$\sigma(\zeta) \coloneqq \beta_r \zeta^r + \beta_{r-1}\zeta^{r-1} + \cdots + \beta_0.$

線形多段階法を (ρ, σ) で表すこともある．

復習：線形 r 段階法 (固定刻み幅)

$$\sum_{i=0}^{r} \alpha_i x^{(k+i)} = h \sum_{i=0}^{r} \beta_i g\left(t_{k+i}, x^{(k+i)}\right).$$

線形多段階法の零安定性

定義：零安定性

線形多段階法 (ρ, σ) が零安定
$\Longleftrightarrow$ 第一特性多項式 ρ が以下の条件 (根条件) を満たす.

- 全ての根の絶対値が 1 以下；
- 根の絶対値がちょうど 1 であるとき，単根である.

零安定性の意味：
$\dot{x} = 0$ に線形多段階法 (ρ, σ) を適用した

$$\alpha_r x^{(k+r)} + \alpha_{r-1} x^{(k+r-1)} + \cdots + \alpha_0 x^{(k)} = 0$$

の解が有界に留まる.
ラフな説明：ρ が相異なる r 個の根 $\zeta_1, \ldots, \zeta_r$ をもつとき，

$$x^{(k)} = C_1 \zeta_1^k + C_2 \zeta_2^k + \cdots + C_r \zeta_r^k$$

$(C_1, \ldots, C_r$ は出発値から定まる).

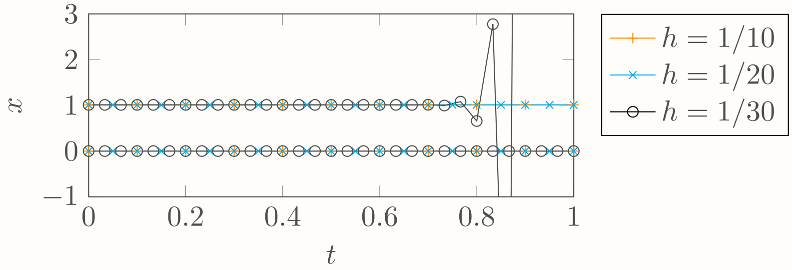
零安定でない線形多段階法

陽的 2 段階法 (3 次精度)

$$x^{(k+2)} + 4x^{(k+1)} - 5x^{(k)} = h\Big(4g\Big(t_{k+1}, x^{(k+1)}\Big) + 2g\Big(t_k, x^{(k)}\Big)\Big)$$

の第一特性多項式 $\rho(\zeta) = \zeta^2 + 4\zeta - 5$ の根は $1, -5 \Rightarrow$ 零安定でない.

数値実験 ($\dot{x} = 0$)：
2 通りの初期値 0, 1.01 に対して，$x^{(1)}$ をそれぞれ 0, 1.01 (厳密解) とした.
ステップ幅 h を小さくすると，数値計算結果は発散 (丸め誤差の拡大).



線形多段階法の絶対安定性

Dahlquist のテスト方程式 $\dot{x} = \lambda x$ に対する線形多段階法

$$\sum_{i=0}^{r} \alpha_i x^{(k+i)} = \lambda h \sum_{i=0}^{r} \beta_i x^{(k+i)}.$$

この漸化式の特性多項式：$\rho(\zeta) - \lambda h \sigma(\zeta)$.

定義：絶対安定性

線形多段階法 (ρ, σ) の絶対安定領域：

$$S := \{\, z \in \mathbb{C} \mid \pi(\zeta; z) := \rho(\zeta) - z\sigma(\zeta) \text{ が根条件を満たす} \,\}.$$

$S \supseteq \mathbb{C}^-$ のとき，A 安定という．

(復習) 根条件：

- 全ての根の絶対値が 1 以下；
- 根の絶対値がちょうど 1 であるとき，単根である．

加速勾配法に対応する線形多段階法

μ-強凸 L-平滑関数に対する Nesterov の加速勾配法

$$y^{(k+1)} = x^{(k)} - \frac{1}{L}\nabla f\left(x^{(k)}\right),$$

$$x^{(k+1)} = y^{(k+1)} + \gamma\left(y^{(k+1)} - y^{(k)}\right), \qquad \gamma := \frac{1-\sqrt{\frac{\mu}{L}}}{1+\sqrt{\frac{\mu}{L}}}.$$

y を消去すると,

$$x^{(k+1)} - (1+\gamma)x^{(k)} + \gamma x^{(k-1)} = \frac{1+\gamma}{L}\left(-\nabla f\left(x^{(k)}\right)\right) - \frac{\gamma}{L}\left(-\nabla f\left(x^{(k-1)}\right)\right).$$

$h = \frac{1}{L(1-\gamma)}$ とする (後述).

加速勾配法に対応する線形多段階法

$$x^{(k+2)} - (1+\gamma)x^{(k+1)} + \gamma x^{(k)} = h\left((1-\gamma^2)g\left(x^{(k+1)}\right) - \gamma(1-\gamma)g\left(x^{(k)}\right)\right)$$

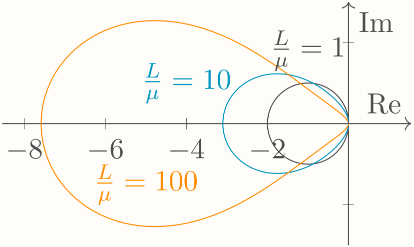
加速勾配法に対応する線形多段階法の絶対安定領域

加速勾配法に対応する線形多段階法

$$x^{(k+2)}-(1+\gamma)x^{(k+1)}+\gamma x^{(k)} = h\Big((1-\gamma^2)g\big(x^{(k+1)}\big) - \gamma(1-\gamma)g\big(x^{(k)}\big)\Big)$$

上の線形多段階法について $\left(\gamma = \left(1-\sqrt{\frac{\mu}{L}}\right)\Big/\left(1+\sqrt{\frac{\mu}{L}}\right)\right)$,

$$\rho(\zeta) = \zeta^2 - (1+\gamma)\zeta + \gamma, \qquad \sigma(\zeta) = (1-\gamma^2)\zeta - \gamma(1-\gamma),$$
$$\pi(\zeta; z) = \rho(\zeta) - z\sigma(\zeta) = \zeta^2 - (1+\gamma)(1+z(1-\gamma))\zeta + \gamma(1+z(1-\gamma)).$$



安定領域の形状は L/μ に依存.

安定領域の左端 $\approx -\frac{2}{3}\sqrt{\frac{L}{\mu}}$ (演習問題).

勾配流の場合は $\lambda \in [-L, 0]$ に相当
$\Rightarrow$ 刻み幅制限は $\mathrm{O}\big(1/\sqrt{\mu L}\big)$.

(再掲) 「加速」の説明 [Scieur et al., 2017]

目的関数 $f\colon \mathbb{R}^d \to \mathbb{R}$ が μ-強凸 (かつ L-平滑) の場合 ($f^\star$：最適値)

勾配流 $\dot{x} = -\nabla f(x)$

$$f(x(t)) - f^\star = \mathrm{O}\left(e^{-2\mu t}\right)$$

最急降下法 $t_k = \mathrm{O}\left(\frac{k}{L}\right)$ → 陽的 Euler 法

$$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(e^{-2\frac{\mu}{L}k}\right)$$

線形多段階法 $t_k = \mathrm{O}\left(\frac{k}{\sqrt{\mu L}}\right)$ → 加速勾配法

$$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(e^{-\sqrt{\frac{\mu}{L}}k}\right)$$

「加速」の (一つの) 解釈：
安定性の向上による
ステップ幅制限の緩和

数値解析でよく知られている「安定性」の条件から,

- 陽的 Euler 法：$h_k = \mathrm{O}\left(\frac{1}{L}\right) \to t_k \coloneqq \sum_{i=0}^{k-1} h_i = \mathrm{O}\left(\frac{k}{L}\right)$,
- 線形多段階法：$h_k = \mathrm{O}\left(\frac{1}{\sqrt{\mu L}}\right) \to t_k = \mathrm{O}\left(\frac{k}{\sqrt{\mu L}}\right)$.

Table of contents

自励系と非自励系

一般的な非自励系 $\dot{x} = g(t, x)$ に対して，自励系

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} g(s, x) \\ 1 \end{bmatrix}$$

を考えると，$s = t$ より両者は等価 $\Rightarrow$ (大抵の場合) 自励系を考えれば十分.

例：**Runge–Kutta** 法

上記の自励系に対する Runge–Kutta 法

$$X_i^{(k)} = x^{(k)} + h\sum_{j=1}^{s} a_{ij} g\left(S_j^{(k)}, X_j^{(k)}\right), \qquad S_i^{(k)} = s^{(k)} + h\sum_{j=1}^{s} a_{ij},$$

$$x^{(k+1)} = x^{(k)} + h\sum_{i=1}^{s} b_i g\left(S_i^{(k)}, X_i^{(k)}\right), \qquad s^{(k+1)} = s^{(k)} + h\sum_{i=1}^{s} b_i.$$

$c_i = \sum_{j=1}^{s} a_{ij}$, $\sum_{i=1}^{s} b_i = 1$ の下，前述の Runge–Kutta 法と一致.

局所誤差と大域誤差

局所誤差評価： $\left\| x^{(1)} - x(h) \right\| \leq Ch^{p+1}$

大域誤差評価：
ODE に沿って時間 h 進める写像 (flow map) $\varphi_h \colon \mathbb{R}^d \to \mathbb{R}^d$ を用いて，
誤差 $e^{(k)} := \left\| x^{(k)} - x(kh) \right\|$ を評価：

$$e^{(k+1)} \leq \underbrace{\left\| x^{(k+1)} - \varphi_h\left(x^{(k)}\right) \right\|}_{\text{局所誤差}} + \underbrace{\left\| \varphi_h\left(x^{(k)}\right) - \varphi_h(x(kh)) \right\|}_{\text{ODE の性質}}$$

$$\leq Ch^{p+1} + \exp(Lh)e^{(k)}$$

$\Rightarrow e^{(k)} \leq C\exp(Lkh)kh^{p+1} \leq C\exp(LT)Th^p$ ($\because kh \leq T$).

ODE の性質 (演習問題)

ODE を定める写像 g が L-Lipschitz 連続のとき，

$$\|\varphi_h(x) - \varphi_h(y)\| \leq \exp(Lh)\|x - y\| \qquad (x, y \in \mathbb{R}^d).$$

Runge–Kutta 法の精度

定義：Runge–Kutta 法の精度

Runge–Kutta 法が p 次精度である
$\iff$ 十分滑らかな任意の ODE に対して，以下の局所誤差評価が成立：

$$\left\| x^{(1)} - x(h) \right\| \leq C h^{p+1}.$$

自励系 $\dot{x} = g(x)$ に対する Runge–Kutta 法

$$\begin{cases} X_i^{(k)} = x^{(k)} + h \displaystyle\sum_{j=1}^{s} a_{ij} g\left(X_j^{(n)}\right) \quad (i = 1, \ldots, s), \\ x^{(k+1)} = x^{(k)} + h \displaystyle\sum_{i=1}^{s} b_i g\left(X_i^{(k)}\right). \end{cases}$$

例：陽的 Euler 法 (1 次精度)

以下の展開より，$\left\| x^{(1)} - x(h) \right\| \le Ch^2$ が成立：

数値解：
$$\begin{aligned} x^{(1)} &= x^{(0)} + hg(x^{(0)}) \\ &= \underline{x_0 + hg(x_0)}. \end{aligned}$$

厳密解：
$$\begin{aligned} x(h) &= x(0) + h\frac{\mathrm{d}x}{\mathrm{d}t}(0) + \frac{h^2}{2}\frac{\mathrm{d}^2x}{\mathrm{d}t^2}(0) + \cdots \\ &= \underline{x(0) + hg(x_0)} + \frac{h^2}{2}g'(x_0)g(x_0) + \cdots . \end{aligned}$$

ここで，以下の関係式を用いた：

$$\begin{aligned} &\frac{\mathrm{d}x}{\mathrm{d}t}(t) = g(x(t)), \\ &\frac{\mathrm{d}^2x}{\mathrm{d}t^2}(t) = \frac{\mathrm{d}}{\mathrm{d}t}g\left(x(t)\right) = g'(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}(t) = g'(x(t))g(x(t)). \end{aligned}$$

p 次精度の Runge–Kutta 法を扱うには，$\dfrac{\mathrm{d}^kx}{\mathrm{d}t^k}(0)$ $(k = 1, \ldots, p)$ が必要.

高階導関数の計算

注意：

- $x(t)$ は引数を省略して x と書く．
- $g^j_{i_1 i_2 \cdots i_n} := \frac{\partial^n g^j}{\partial x^{i_1} \partial x^{i_2} \ldots \partial x^{i_n}}$ とする．ここで $g(x) = (g^1(x), \cdots, g^d(x))^\mathsf{T}$．
- Einstein の和の規約を利用する．

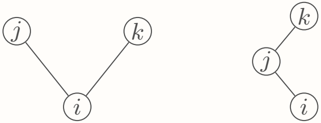
$$
\begin{aligned}
\frac{\mathrm{d}x^i}{\mathrm{d}t} &= g^i(x) \\
\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} &= \frac{\mathrm{d}}{\mathrm{d}t} g^i(x) = g^i_j(x) \frac{\mathrm{d}x^j}{\mathrm{d}t} = g^i_j(x) g^j(x) \\
\frac{\mathrm{d}^3 x^i}{\mathrm{d}t^3} &= \frac{\mathrm{d}}{\mathrm{d}t} \left(g^i_j(x) g^j(x) \right) = g^i_{jk}(x) g^j(x) g^k(x) + g^i_j(x) g^j_k(x) g^k(x) \\
\frac{\mathrm{d}^4 x^i}{\mathrm{d}t^4} &= \frac{\mathrm{d}}{\mathrm{d}t} \left(g^i_{jk}(x) g^j(x) g^k(x) + g^i_j(x) g^j_k(x) g^k(x) \right) = \cdots
\end{aligned}
$$

容易に計算できるが，煩雑．

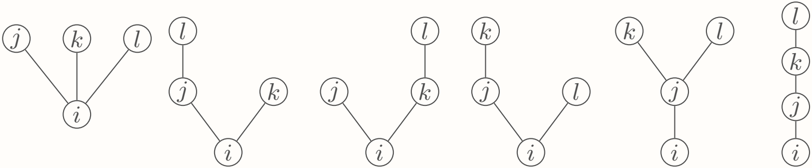
処方箋：根つき木

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = \underline{g^i_j(x) g^j(x)}.$$

$$\frac{\mathrm{d}^3 x^i}{\mathrm{d}t^3} = \underline{g^i_{jk}(x) g^j(x) g^k(x)} + \underline{g^i_j(x) g^j_k(x) g^k(x)}$$



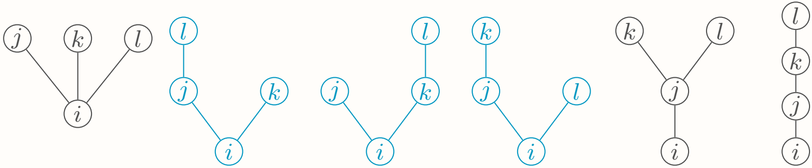
$$\frac{\mathrm{d}^4 x^i}{\mathrm{d}t^4} = g^i_{jkl} g^j g^k g^l + g^i_{jk} g^j_l g^l g^k + g^i_{jk} g^j g^k_l g^l + g^i_{jl} g^j_k g^k g^l + g^i_j g^j_{kl} g^k g^l + g^i_j g^j_k g^k_l g^l$$



処方箋：根つき木

これらの項は同一：単なるインデックスの付け替え

$$\frac{\mathrm{d}^4 x^i}{\mathrm{d}t^4} = g^i_{jkl} g^j g^k g^l + g^i_{jk} g^j_l g^l g^k + g^i_{jk} g^j g^k_l g^l + g^i_{jl} g^j_k g^k g^l + g^i_j g^j_{kl} g^k g^l + g^i_j g^j_k g^k_l g^l$$



これらのラベル付き木は形状が同じ

$$\downarrow$$

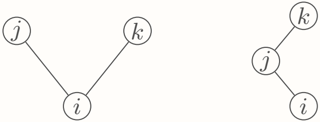
$x(t)$ の高階導関数は根つき木で整理できる．

定義：ラベルつき木

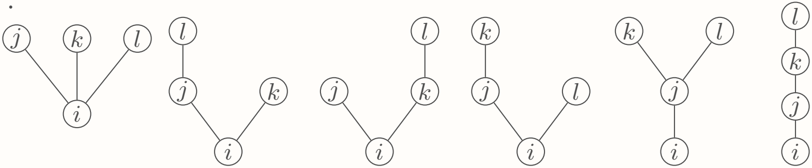
全順序集合 S_q が，最小の要素 i をもち，$|S_q| = q$ を満たすとする．このとき，q 次の (根つき) ラベルつき木とは，写像 $\tau : S_q \setminus \{i\} \to S_q$ であり，任意の $z \in S_q$ に対して，$\tau(z) < z$ を満たすものである．次数 q のラベルつき木全体のなす集合を $\mathcal{LT}_q$ で表す．

$\mathcal{LT}_3$:

$(S_3 = \{i < j < k\})$



$\mathcal{LT}_4$:



$(S_4 = \{i < j < k < l\})$

根つき木：定義

定義：根つき木

根つき木の集合 $\mathcal{T}$ を以下のように再帰的に定義する：

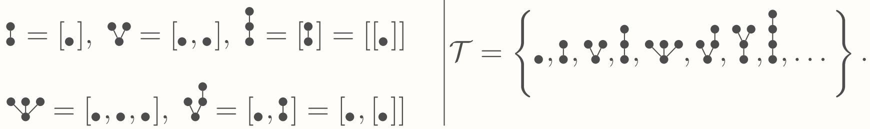
- 頂点を 1 つだけもつグラフ $\bullet$ は $\mathcal{T}$ に属す；
- $\tau_1, \ldots, \tau_m \in \mathcal{T}$ であるとき，$\tau_1, \ldots, \tau_m$ の根を新たに用意した一つの頂点に繋げてできるグラフも $\mathcal{T}$ に属す．この操作を

$$\tau = [\tau_1, \ldots, \tau_m]$$

と表す．ここで，新たに導入した頂点が τ の根である．

注意：$[\tau_1, \ldots, \tau_m]$ は $\tau_1, \ldots, \tau_m$ の並べ方に依存しない．

例：



根つき木：要素的微分

定義：要素的微分

木 $\tau \in \mathcal{T}$ に対して，要素的微分とは，$G(\bullet)(x) = g(x)$ と

$$G^i([\tau_1, \ldots, \tau_m])(x) = g^i_{j_1,\ldots,j_m}(x) G^{j_1}(\tau_1)(x) \cdots G^{j_m}(\tau_m)(x)$$

により再帰的に定義される写像 $G(\tau)\colon \mathbb{R}^d \to \mathbb{R}^d$ である．

例：

$$G^i()(x) = G^i([\bullet])(x) = g^i_j(x) G^j(\bullet)(x) = g^i_j(x) g^j(x)$$

$$G^i()(x) = G^i([\bullet, \bullet])(x) = g^i_{jk}(x) G^j(\bullet)(x) G^k(\bullet)(x) = g^i_{jk}(x) g^j(x) g^k(x)$$

$$G^i()(x) = G^i([])(x) = g^i_j(x) G^j()(x) = g^i_j(x) g^j_k(x) g^k(x)$$

「根を除いた頂点の個数」 $=$ 「微分の階数」

根つき木：$x(t)$ の高階導関数

定理： 厳密解 $x(t)$ の q 階導関数 cf. [Hairer et al., 1993, II, Thm. 2.6]

$$\frac{\mathrm{d}^q x}{\mathrm{d}t^q}(0) = \sum_{|\tau|=q} \alpha(\tau) G(\tau)(x_0).$$

ここで，

- $|\tau|$ は，$\tau \in \mathcal{T}$ の頂点の個数を表す；
- $\alpha(\tau)$ は，τ と同値なラベルつき木の個数を表す．

τ								
$\|\tau\|$	1	2	3	3	4	4	4	4
$\alpha(\tau)$	1	1	1	1	1	3	1	1

Runge–Kutta 法の Taylor 展開に向けて (1/3)

Runge–Kutta 法の内部段 $X_i^{(1)}$ (と出力 $y^{(1)}$) の h に関する Taylor 展開:

$$
\begin{aligned}
X_i^{(1)}(0) &= x_0 \\
\frac{\mathrm{d}}{\mathrm{d}h} X_i^{(1)}(0) &= \sum_{j=1}^{s} a_{ij} g\left(X_j^{(1)}(h)\right) \\
&\qquad + h \sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}(h)\right)\left(\frac{\mathrm{d}}{\mathrm{d}h} X_j^{(1)}(h)\right)\Bigg|_{h=0} \\
&= \sum_{j=1}^{s} a_{ij} g\left(x_0\right)
\end{aligned}
$$

$$
\begin{cases}
X_i^{(1)} = x_0 + h \sum_{j=1}^{s} a_{ij} g\left(X_j^{(1)}\right) \quad (i = 1, \ldots, s), \\
x^{(1)} = x_0 + h \sum_{i=1}^{s} b_i g\left(X_i^{(1)}\right).
\end{cases}
$$

引数 h は省略する.

$$\frac{\mathrm{d}^2}{\mathrm{d}h^2} X_i^{(1)}(h) = \frac{\mathrm{d}}{\mathrm{d}h}\left(\sum_{j=1}^{s} a_{ij} g\left(X_j^{(1)}\right) + h \sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}}{\mathrm{d}h} X_j^{(1)}\right)\right)$$

$$= 2\sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}}{\mathrm{d}h} X_j^{(1)}\right) + h \sum_{j=1}^{s} a_{ij} g''\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}}{\mathrm{d}h} X_j^{(1)}, \frac{\mathrm{d}}{\mathrm{d}h} X_j^{(1)}\right)$$

$$+ h \sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}^2}{\mathrm{d}h^2} X_j^{(1)}\right)$$

$$\Downarrow$$

$$\frac{\mathrm{d}^2}{\mathrm{d}h^2} X_i^{(1)}(0) = 2\sum_{j=1}^{s} a_{ij} g'\left(x_0\right)\left(\sum_{k=1}^{s} a_{jk} g\left(x_0\right)\right)$$

$$= 2\sum_{j=1}^{s}\sum_{k=1}^{s} a_{ij} a_{jk} g'(g)$$

Runge–Kutta 法の Taylor 展開に向けて (3/3)

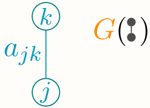
$$
\begin{aligned}
\frac{\mathrm{d}^3}{\mathrm{d}h^3} X_i^{(1)}(h) &= \frac{\mathrm{d}}{\mathrm{d}h}\left(2\sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}\right)\frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)} + h\sum_{j=1}^{s} a_{ij} g''\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}, \frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}\right)\right. \\
&\qquad\qquad \left. + h\sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}\right)\frac{\mathrm{d}^2}{\mathrm{d}h^2}X_j^{(1)}\right) \\
&= 3\sum_{j=1}^{s} a_{ij} g''\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}, \frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}\right) + 3\sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}\right)\frac{\mathrm{d}^2}{\mathrm{d}h^2}X_j^{(1)} \\
&\quad + h\sum_{j=1}^{s} a_{ij} g'''\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}, \frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}, \frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}\right) \\
&\quad + 3h\sum_{j=1}^{s} a_{ij} g''\left(X_j^{(1)}\right)\left(\frac{\mathrm{d}^2}{\mathrm{d}h^2}X_j^{(1)}, \frac{\mathrm{d}}{\mathrm{d}h}X_j^{(1)}\right) + h\sum_{j=1}^{s} a_{ij} g'\left(X_j^{(1)}\right)\frac{\mathrm{d}^3}{\mathrm{d}h^3}X_j^{(1)}
\end{aligned}
$$

$$\Downarrow$$

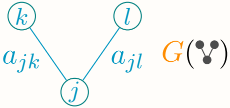
$$
\begin{aligned}
\frac{\mathrm{d}^3}{\mathrm{d}h^3} X_i^{(1)}(0) &= 3\sum_{j=1}^{s} a_{ij}\left(g''(x_0)\left(\sum_{k=1}^{s} a_{jk} g(x_0), \sum_{l=1}^{s} a_{jl} g(x_0)\right) + g'(x_0)\frac{\mathrm{d}^2}{\mathrm{d}h^2}X_j^{(1)}(0)\right) \\
&= 3\sum_{j,k,l} a_{ij}a_{jk}a_{jl} g''(g,g) + 3\sum_{j,k,l} a_{ij}a_{jk}a_{kl} g'(g'(g)).
\end{aligned}
$$

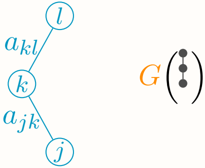
根つき木による整理

$$\frac{\mathrm{d}^2}{\mathrm{d}h} X_i^{(1)} = 2 \sum_{j,k} a_{ij} a_{jk} g'(g)$$



$$\frac{\mathrm{d}^3}{\mathrm{d}h^3} X_i^{(1)}(0) = 3 \sum_{j,k,l} a_{ij} a_{jk} a_{jl} g''(g, g) + 3 \sum_{j,k,l} a_{ij} a_{jk} a_{kl} g'(g'(g))$$





$\vdots$

定理： RK 法にまつわる q 階導関数 cf. [Hairer et al., 1993, II, Thm. 2.11]

$$\frac{\mathrm{d}^q}{\mathrm{d}h^q} X_i^{(1)}(0) = \sum_{|\tau|=q} \alpha(\tau)\tau! \sum_{j=1}^{s} a_{ij}\Phi_j(\tau)G(\tau)(x_0),$$

$$\frac{\mathrm{d}^q}{\mathrm{d}h^q} y^{(1)}(0) = \sum_{|\tau|=q} \alpha(\tau)\tau! \sum_{i=1}^{s} b_i\Phi_i(\tau)G(\tau)(x_0),$$

- 木 τ の階乗 $\tau!$ は， $\bullet! = 1$ と以下の式で再帰的に定義される：

$$[\tau_1, \dots, \tau_m]! = |[\tau_1, \dots, \tau_m]| \, (\tau_1!) \cdots (\tau_m!);$$

- Φ_j は， $\Phi_j(\bullet) = 1$ と以下の式で再帰的に定義される：

$$\Phi_j([\tau_1, \dots, \tau_m]) = \sum_{k_1, \cdots, k_m} a_{jk_1} a_{jk_2} \cdots a_{jk_m} \Phi_{k_1}(\tau_1) \cdots \Phi_{k_m}(\tau_m).$$

B 級数 (Butcher 級数)

定義：B 級数

写像 $\delta\colon \mathcal{T} \cup \{\emptyset\} \to \mathbb{R}$ に対して，以下の形式的級数を B 級数という：

$$B(\delta, x_0) = \delta(\emptyset)x_0 + \sum_{\tau\in\mathcal{T}} \frac{h^{|\tau|}}{\sigma(\tau)}\delta(\tau)G(\tau)(x_0)$$

対称度 $\sigma(\tau)$ は，$\sigma(\bullet) = 1$ と以下の式で定義される：

$$\sigma([\tau_1, \dots, \tau_m]) = \sigma(\tau_1)\cdots\sigma(\tau_m)\mu_1!\mu_2!\cdots .$$

ここで，整数 $\mu_1, \mu_2, \dots$ は $\tau_1, \dots, \tau_m$ の中の同じ木の個数である.

例：

$$\sigma(\;) = \sigma([\bullet]) = 1, \qquad \sigma(\;) = \sigma([\bullet, \bullet]) = 1\cdot 1\cdot 2!$$

$$\sigma(\;) = \sigma([\;]) = 1 \qquad \sigma(\;) = \sigma([\bullet, \bullet, \bullet]) = 1\cdot 1\cdot 1\cdot 3!$$

$$\sigma(\;) = \sigma([\bullet, \;]) = 1\cdot 1 \qquad \sigma(\;) = \sigma([\;]) = 2$$

厳密解と Runge–Kutta 法の出力を表す B 級数

厳密解の Taylor 展開：

$$x(h) = 1 \cdot x_0 + \sum_{\tau \in \mathcal{T}} \frac{h^{|\tau|}}{\sigma(\tau)} \frac{1}{\tau!} G(\tau)(x_0)$$

Runge–Kutta 法の出力の Taylor 展開：

$$x^{(1)}(h) = 1 \cdot x_0 + \sum_{\tau \in \mathcal{T}} \frac{h^{|\tau|}}{\sigma(\tau)} \sum_{i=1}^{s} b_i \Phi_i(\tau) G(\tau)(x_0)$$

$$B(\delta, x_0) = \delta(\emptyset) x_0 + \sum_{\tau \in \mathcal{T}} \frac{h^{|\tau|}}{\sigma(\tau)} \delta(\tau) G(\tau)(x_0)$$

Runge–Kutta 法の次数条件

定理：次数条件 cf. [Hairer et al., 1993, II, Thm. 2.13]

Runge–Kutta 法が p 次精度 $\iff$

$$\sum_{i=1}^{s} b_i \Phi_i(\tau) = \frac{1}{\tau!} \qquad \text{for } |\tau| \le p.$$

厳密解：

$$x(h) = 1 \cdot x_0 + \sum_{\tau \in \mathcal{T}} \frac{h^{|\tau|}}{\sigma(\tau)} \frac{1}{\tau!} G(\tau)(x_0),$$

数値解：

$$x^{(1)}(h) = 1 \cdot x_0 + \sum_{\tau \in \mathcal{T}} \frac{h^{|\tau|}}{\sigma(\tau)} \sum_{i=1}^{s} b_i \Phi_i(\tau) G(\tau)(x_0),$$

Table: 次数条件の個数

次数 p	1	2	3	4	5	6	7	8	9	10
条件の数	1	2	4	8	17	37	85	200	486	1205

次数条件の具体例

$$\sum_{i=1}^{s} b_i \Phi_i(\tau) = \frac{1}{\tau!} \qquad \text{for } |\tau| \le p.$$

$$\sum_{i} b_i = 1$$

$$\sum_{i,j} b_i a_{ij} = \frac{1}{2}$$

$$\sum_{i,j,k} b_i a_{ij} a_{ik} = \frac{1}{3}$$

$$\sum_{i,j,k} b_i a_{ij} a_{jk} = \frac{1}{6}$$

$$\sum_{i,j,k,l} b_i a_{ij} a_{ik} a_{il} = \frac{1}{4}$$

$$\sum_{i,j,k,l} b_i a_{ij} a_{ik} a_{kl} = \frac{1}{8}$$

$$\sum_{i,j,k,l} b_i a_{ij} a_{jk} a_{jl} = \frac{1}{12}$$

$$\sum_{i,j,k,l} b_i a_{ij} a_{jk} a_{kl} = \frac{1}{24}$$

余談 1：高次精度の陽的 Runge–Kutta 法

p：次数，s：段数

- $p = s$ を満たす陽的 Runge–Kutta 法は，$p \leq 4$ で存在する．
- $p \geq 5$ 次精度の陽的 Runge–Kutta 法は，$s \geq p+1$ 段必要．
- $p \geq 7$ 次精度の陽的 Runge–Kutta 法は，$s \geq p+2$ 段必要．
- $p \geq 8$ 次精度の陽的 Runge–Kutta 法は，$s \geq p+3$ 段必要．

- [Hairer, 1978] が，17 段 10 次の陽的 Runge–Kutta 法を提案．
  注：決めるべきパラメータ 153 個，満たすべき次数条件 1205 個．
- 40 年以上を経て，[Zhang, 2019] は，
  16 段 10 次の陽的 Runge–Kutta 法が構成できたと主張している．
  ただし，これは次数条件を最小二乗問題として数値的に解いた結果．
  (BFGS 法と，いろいろな実装上の工夫の賜物とのこと．)

- 任意の正整数 p に対して，$s = p(p-1)/2 + 1$ 段の
  p 次陽的 Runge–Kutta 法が存在 (非実用的) [Zhang, 2019, Prop. 1.28].

余談 2：陰的 Runge–Kutta 法

選点法 (陰的 Runge–Kutta 法とも解釈できる)

$c_1, \ldots, c_s$ を相異なる実数とする．s 次選点多項式 u を

$$u(t_k) = x^{(k)}, \qquad \dot{u}(t_k + c_i h) = g(u(t_k + c_i h)) \quad (i = 1, \ldots, s)$$

から定め，$x^{(k+1)} = u(t_k + h)$ を出力とする．

定理：選点法の次数 cf. [Hairer et al., 2006, II, Thm. 1.5]

選点法の精度 p は，$c_1, \ldots, c_s$ を用いる数値積分公式の次数と一致する．

数値積分：

$$\int_0^1 f(x)\mathrm{d}x \approx \frac{1}{s}\sum_{i=1}^{s} b_i f(c_i)$$

Gauss 法：shifted Legendre 多項式の零点を c_i として用いると，$p = 2s$.

Table of contents

線形多段階法の局所誤差

線形多段階法における局所誤差： $\left\|x^{(r)} - x(rh)\right\|$
ただし，$x^{(1)}, \dots, x^{(r-1)}$ は正確な出発値 $x^{(i)} = x(ih)$ $(i = 1, \dots, r-1)$.

復習：線形 r 段階法 (固定刻み幅)

$$\sum_{i=0}^{r} \alpha_i x^{(k+i)} = h \sum_{i=0}^{r} \beta_i g\left(t_{k+i}, x^{(k+i)}\right).$$

定義と，「正確な出発値」の仮定より，(簡潔性のため，自励系とする)

$$\underline{\sum_{i=0}^{r}\Big(\alpha_i x(ih) - h\beta_i \overset{\dot{x}(ih)}{g(x(ih))}\Big)} = \alpha_r\left(x(rh) - x^{(r)}\right) - h\beta_r\left(g(x(rh)) - g\left(x^{(r)}\right)\right)$$

$$= \left(\alpha_r I - h\beta_r \frac{\partial g}{\partial x}(\eta)\right)\left(x(rh) - x^{(r)}\right)$$

$$\|$$

$$\boxed{\mathcal{L}(x;h) := \sum_{i=0}^{r}(\alpha_i x(ih) - h\beta_i \dot{x}(ih))}$$

(最後の等号は平均値の定理より従う)

線形多段階法の次数の求め方

前ページで確認したこと：

$$\mathcal{L}(x;h) \coloneqq \sum_{i=0}^{r} (\alpha_i x(ih) - h\beta_i \dot{x}(ih)) \approx \alpha_r \left(x(rh) - x^{(r)} \right).$$

p 次精度 $\left(\left\| x^{(r)} - x(rh) \right\| \leq Ch^{p+1}\right)$

$\iff$ 十分滑らかな $x\colon \mathbb{R} \to \mathbb{R}^d$ について，$\|\mathcal{L}(x;h)\| \leq C' h^{p+1}$.

$\iff$ p 次以下の多項式 P について，$\mathcal{L}(P;h) = 0$.

定理：線形多段階法の次数 [Hairer et al., 1993, III, Thm. 2.4]

線形多段階法 (ρ, σ) が p 次精度

$$\iff \quad \sum_{i=0}^{r} \alpha_i = 0 \quad \text{かつ} \quad \sum_{i=0}^{r} \alpha_i i^q = q \sum_{i=0}^{r} \beta_i i^{q-1} \ (q = 1, \ldots, p).$$

加速勾配法に対応する線形多段階法の精度 (1/2)

μ-強凸 L-平滑関数に対する Nesterov の加速勾配法の書き換え：

$$\underbrace{1}_{\alpha_2}\, x^{(k+1)} + \underbrace{(-1-\gamma)}_{\alpha_1}\, x^{(k)} + \underbrace{\gamma}_{\alpha_0}\, x^{(k-1)} = \underbrace{\frac{1+\gamma}{L}}_{h\beta_1}\left(-\nabla f\left(x^{(k)}\right)\right) + \underbrace{\frac{-\gamma}{L}}_{h\beta_0}\left(-\nabla f\left(x^{(k-1)}\right)\right).$$

- $\displaystyle\sum_{i=0}^{2} \alpha_i = \gamma + (-1-\gamma) + 1 = 0.$
- $q = 1$ の条件：

$$0 = \sum_{i=0}^{2} \alpha_i i - \sum_{i=0}^{2} \beta_i = 1 - \gamma - \frac{1}{hL} \quad \Rightarrow \quad h = \frac{1}{L(1-\gamma)}.$$

定理：線形多段階法の次数条件 [Hairer et al., 1993, III, Thm. 2.4]

$$p \text{ 次} \iff \sum_{i=0}^{r} \alpha_i = 0 \quad \text{かつ} \quad \sum_{i=0}^{r} \alpha_i i^q = q \sum_{i=0}^{r} \beta_i i^{q-1} \ (q = 1, \ldots, p).$$

加速勾配法に対応する線形多段階法の精度 (2/2)

加速勾配法に対応する線形多段階法

$$x^{(k+2)} - (1+\gamma)x^{(k+1)} + \gamma x^{(k)} = h\Big((1-\gamma^2)g\big(x^{(k+1)}\big) - \gamma(1-\gamma)g\big(x^{(k)}\big)\Big)$$

定理：線形多段階法の次数条件 [Hairer et al., 1993, III, Thm. 2.4]

$$p \text{ 次} \iff \sum_{i=0}^{r} \alpha_i = 0 \quad \text{かつ} \quad \sum_{i=0}^{r} \alpha_i i^q = q \sum_{i=0}^{r} \beta_i i^{q-1} \ (q = 1, \ldots, p).$$

$q = 2$ の条件は成立しない：

$$\sum_{i=0}^{2} \alpha_i i^2 = \gamma \cdot 0^2 + (-1-\gamma) \cdot 1^2 + 2^2 = 3 - \gamma,$$

$$2 \sum_{i=0}^{2} \beta_i i = 2\big(-\gamma(1-\gamma) \cdot 0 + \big(1-\gamma^2\big)\big) = 2\big(1-\gamma^2\big)$$

線形多段階法の収束性 (大域誤差)

定義：線形多段階法の収束性

線形多段階法が p 次収束
$\iff$ ある正の数 $h_0 \in \mathbb{R}$ が存在し,$h \leq h_0$ に対して，出発値が

$$\left\| x^{(i)} - x(ih) \right\| \leq C_0 h^p \qquad (i = 0, 1, \ldots, r-1)$$

を満たすとき，以下が成立する：

$$\left\| x^{(k)} - x(kh) \right\| \leq C h^p \qquad (k = r, r+1, \ldots).$$

定理：収束性の必要十分条件 [Hairer et al., 1993, III, Thm. 4.2 and 4.5]

線形多段階法が p 次収束 $\iff$ 線形多段階法が p 次精度かつ零安定

Dahlquist の障壁

- 線形多段階法が p 次精度であるための条件は $p+1$ 個の線形方程式.
- 線形 r 段階法のパラメータの自由度は $2r+1$ (陽的に限ると $2r$) 個.

$\rightarrow$ 陰的では $p=2r$, 陽的では $p=2r-1$ が達成可能?
$\rightarrow$ 達成可能だが，零安定性が成立しない.

定理：Dahlquist の第 1 障壁 [Hairer et al., 1993, III, Thm. 3.5]

零安定な線形 r 段階法の次数 p は以下の関係を満たす：

- r が偶数のとき，$p \leq r+2$.
- r が奇数のとき，$p \leq r+1$.
- $\frac{\beta_r}{\alpha_r} \leq 0$ (陽的な場合を含む) のとき，$p \leq r$.

定理：Dahlquist の第 2 障壁 [Hairer and Wanner, 1996, V, Thm. 1.4]

A 安定な線形多段階法は高々 2 次精度である.

Table of contents

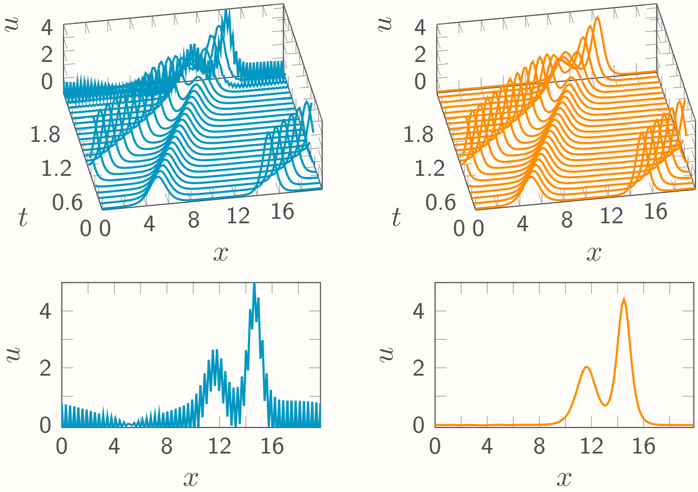
構造保存数値解法

例：非線形波動のシミュレーション (KdV 方程式)

汎用解法　　構造保存数値解法

時間発展

$T = 2.07$



構造保存数値解法

汎用解法

Runge–Kutta 法，線形多段階法，…

+ 全ての方程式に利用可能
+ 豊富な理論 (精度，安定性)
− 難しい問題に対しては非効率的

特殊解法

Störmer 法，Verlet 法，…

+ 優れたパフォーマンス (特に長時間積分)
− 「方程式」に強く依存

構造保存数値解法 (GNI: Geometric Numerical Integration)

- 方程式の「重要な」性質を抽出
  symplecticity (symplectic 法), 保存性/散逸性 (離散勾配法,…)，…
- 優れた長時間挙動
- 適用範囲の拡大：離散勾配法の場合
  各論 → Hamilton 系 [Gonzalez, 1996]
  → (任意の) 保存/散逸系 [McLachlan et al., 1999]

散逸的数値解法

散逸系

$$\dot{x} = S(x)\nabla V(x)$$

- $S : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $V : \mathbb{R}^d \to \mathbb{R}$
- $S(x)$ が半負定値 $(\forall x \in \mathbb{R}^d)$ $\Rightarrow$ V は散逸：$\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) \le 0$
- 勾配流，散逸的 PDE の空間離散化

散逸的数値解法

$V\left(x^{(k+1)}\right) \le V\left(x^{(k)}\right)$ を満たす数値解法

- 各種の構成方法：離散勾配法，射影法，…
- 優れた長時間挙動
- 理論解析のサポート

離散勾配法

勾配流 $$\dot{x} = S(x)\nabla V(x)$$

散逸則 $$\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) = \langle \nabla V(x), \dot{x} \rangle = \langle \nabla V(x), S(x)\nabla V(x) \rangle \leq 0$$

連鎖律

半負定値性

離散勾配法 [Gonzalez, 1996],[McLachlan et al., 1999]

スキーム $$\frac{x^{(k+1)} - x^{(k)}}{h} = \overline{S}\left(x^{(k+1)}, x^{(k)}\right)\overline{\nabla}V\left(x^{(k+1)}, x^{(k)}\right)$$

散逸則 $$\frac{V\left(x^{(k+1)}\right) - V\left(x^{(k)}\right)}{h} = \left\langle \overline{\nabla}V, \frac{x^{(k+1)} - x^{(k)}}{h} \right\rangle = \langle \overline{\nabla}V, \overline{S}\,\overline{\nabla}V \rangle = 0$$

離散連鎖律

半負定値性

離散勾配の定義と構成例

離散勾配

C^1 級関数 $V : \mathbb{R}^d \to \mathbb{R}$ に対して，
以下の性質を満たす連続写像 $\overline{\nabla} V : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ を離散勾配という：

1. $\langle \overline{\nabla} V(x,y), x-y \rangle = V(x) - V(y)$；
2. $\overline{\nabla} V(x,x) = \nabla V(x)$.

- Gonzalez の離散勾配 [Gonzalez, 1996]

$$\overline{\nabla}_{\mathrm{G}} V(x,y) = \nabla V\left(\frac{x+y}{2}\right) + \frac{V(x) - V(y) - \left\langle \nabla V\left(\frac{x+y}{2}\right), x-y \right\rangle}{\|x-y\|_2^2}(x-y)$$

- Average Vector Field (AVF) [Quispel and McLaren, 2008]

$$\overline{\nabla}_{\mathrm{AVF}} V(x,y) = \int_0^1 \nabla V(\xi x + (1-\xi) y) \mathrm{d}\xi$$

- Itoh–Abe の離散勾配 [Itoh and Abe, 1988] (次のページ)

Itoh–Abe の離散勾配 [Itoh and Abe, 1988]

勾配流 $\dot{x} = -\nabla V(x)$ に対する Itoh–Abe の離散勾配法

$$
\begin{bmatrix}
\dfrac{x_1^{(k+1)} - x_1^{(k)}}{h} \\[2ex]
\dfrac{x_2^{(k+1)} - x_2^{(k)}}{h} \\[2ex]
\vdots \\[2ex]
\dfrac{x_d^{(k+1)} - x_d^{(k)}}{h}
\end{bmatrix}
= -
\begin{bmatrix}
\dfrac{V\left(x_1^{(k+1)}, x_2^{(k)}, \ldots, x_d^{(k)}\right) - V\left(x_1^{(k)}, x_2^{(k)}, \ldots, x_d^{(k)}\right)}{x_1^{(k+1)} - x_1^{(k)}} \\[2ex]
\dfrac{V\left(x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k)}, \ldots, x_d^{(k)}\right) - V\left(x_1^{(k+1)}, x_2^{(k)}, \ldots, x_d^{(k)}\right)}{x_2^{(k+1)} - x_2^{(k)}} \\[2ex]
\vdots \\[2ex]
\dfrac{V\left(x_1^{(k+1)}, \ldots, x_{d-1}^{(k+1)}, x_d^{(k+1)}\right) - V\left(x_1^{(k+1)}, \ldots, x_{d-1}^{(k+1)}, x_d^{(k)}\right)}{x_d^{(k+1)} - x_d^{(k)}}
\end{bmatrix}
$$

- 第一式に含まれる未知量は $x_1^{(k+1)}$ のみ．
  1 変数の非線形方程式を解けばよい．
- $x_1^{(k+1)}$ を得たら，第二式の未知量は $x_2^{(k+1)}$ のみ．以下同様．
- 1 変数非線形方程式を d 回解けばよい．

構造保存数値解法の軽量化

離散勾配法の弱点：計算コスト
陰的非線形スキーム, i.e., 毎ステップ非線形方程式を解く必要あり

<u>高速化のための工夫</u>

多段線形化：多段エネルギーを散逸

- 離散変分法 (PDE 版の離散勾配法) の陰的線形多段化 [Matsuo and Furihata, 2001]

補助変数を入れて陰的線形化：修正エネルギーを散逸

- 非線形 Schrödinger 方程式に対する Besse の方法 [Besse, 2004]
- Invariant Energy Quadratization 法 [Yang and Han, 2017]
- Scalar Auxiliary Variable 法 [Shen et al., 2018]

補助変数を入れて 1 変数の非線形方程式に帰着：元のエネルギーを散逸

- Lagrange Multiplier 法 [Cheng et al., 2020]

離散勾配法の最適化への応用 [Ehrhardt et al., 2018]

離散勾配法のおさらい

$$\frac{x^{(k+1)} - x^{(k)}}{h} = -\overline{\nabla} f\left(x^{(k+1)}, x^{(k)}\right)$$

- 離散勾配 $\overline{\nabla} f$ を巧妙に用いて散逸性を再現
- 離散勾配は必ず $x^{(k+1)}$ に依存するため，陰的解法
- (非線形方程式の解が存在する限り) $f\left(x^{(k+1)}\right) \leq f\left(x^{(k)}\right)$

最適化への応用：

- 各種の標準的な離散勾配法の可解性
- 最適化手法としての収束レートの証明
- 関連研究
  - 画像関係の最適化問題への応用 [Ringholm et al., 2018]
  - 非凸非平滑関数の最適化への応用 [Riis et al., 2022]

Table of contents

連続最適化と微分方程式の数値解析

最適化に適したダイナミクス

最適化問題

例：無制約最適化問題

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

微分方程式

例：勾配流

$$\dot{x} = -\nabla f(x)$$

反復による最適解の探索

連続極限

離散化

最適化に適した離散化手法

最適化手法

例：最急降下法

$$x^{(k+1)} = x^{(k)} - h_k \nabla f\left(x^{(k)}\right)$$

Nesterov の加速勾配法，Newton 法，…

対応

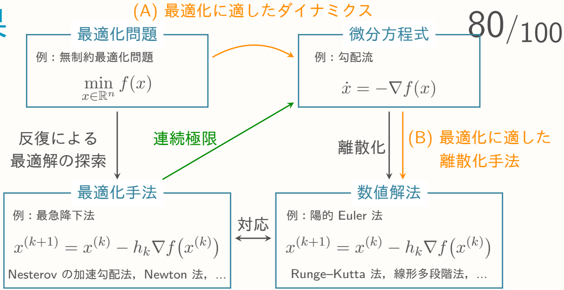
数値解法

例：陽的 Euler 法

$$x^{(k+1)} = x^{(k)} - h_k \nabla f\left(x^{(k)}\right)$$

Runge–Kutta 法，線形多段階法，…

我々の周辺で得た成果



(B) 構造保存数値解法の適用 [Onuma and Sato, 2023]

Lagrange Multiplier 法 [Cheng et al., 2020] の勾配流への適用

(A) 本質的な収束レート [Ushiyama et al., 2022b]

連続時間ダイナミクスにおける収束レートの不定性を除去した本質的収束レート

(B) 安定な陽解法の適用 [Ushiyama et al., 2022a]

(B) 弱い離散勾配による手法の構成と統一的収束証明 [Ushiyama et al., 2023]

離散勾配の定義を最適化用に拡張，収束レートの証明

(B) 加速勾配法の可変刻み線形多段階法解釈 [Nozawa, 2022]

加速勾配法が可変刻み線形多段階法とみなせることの指摘，手法の拡張

構造保存数値解法の適用 (1/4)：LM 法

仮定：関数 $V\colon \mathbb{R}^d \to \mathbb{R}$ が $V(x) = \frac{1}{2}\langle x, Qx\rangle + f(x)$ とかける

勾配流に対する Lagrange Multiplier (LM) 法 [Cheng et al., 2020]

$$\begin{cases} \frac{x^{(k+1)} - x^{(k)}}{h} = -\left(Q\frac{x^{(k+1)} + x^{(k)}}{2} + \eta^{(k)}\nabla f\left(x^{(k)}\right)\right) \\ f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) = \eta^{(k)}\left\langle \nabla f\left(x^{(k)}\right), x^{(k+1)} - x^{(k)}\right\rangle \end{cases}$$

- 未知量 $x^{(k+1)}, \eta^{(k)}$ は (n 次元線形方程式を 2 回と) 1 次元非線形方程式を 1 回解くだけで計算可能
- (非線形方程式の解が存在する限り) $V\left(x^{(k+1)}\right) \leq V\left(x^{(k)}\right)$
- Cahn–Hilliard 方程式などの散逸量をもつ PDE に適用

→ 最適化に使えそう！

問題点：
非線形方程式の可解性が未解決.
→ 本研究：可解性 (ここでは Q が零行列の場合だけを扱う)

構造保存数値解法の適用 (2/4)：可解性

Lagrange Multiplier 法の特殊ケース

$$\begin{cases} \frac{x^{(k+1)}-x^{(k)}}{h} = -\eta^{(k)}\nabla f\left(x^{(k)}\right) \\ f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) = \eta^{(k)}\left\langle \nabla f\left(x^{(k)}\right), x^{(k+1)} - x^{(k)} \right\rangle \end{cases}$$

解くべき非線形方程式：

$$F_h(\eta) := f\left(x^{(k)} - \eta h \nabla f\left(x^{(k)}\right)\right) - f\left(x^{(k)}\right) + h\eta^2 \left\| \nabla f\left(x^{(k)}\right) \right\|^2 = 0$$

定理：可解性

任意の $h > 0$ に対して，ある $\eta \in \mathbb{R}$ が存在し，$F_h(\eta) = 0$ と $\eta \geq (1 + Lh/2)^{-1}$ を満たす．
さらに，f が凸の場合，$\eta \in [(1 + Lh/2)^{-1}, 1]$ を満たす解が存在する．

構造保存数値解法の適用 (3/4)：計算量の削減

Lagrange Multiplier 法の特殊ケース

$$\begin{cases} \frac{x^{(k+1)} - x^{(k)}}{h} = -\eta^{(k)} \nabla f\left(x^{(k)}\right) \\ f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) = \eta^{(k)} \left\langle \nabla f\left(x^{(k)}\right), x^{(k+1)} - x^{(k)} \right\rangle \end{cases}$$

提案手法：不等式への緩和

$$\begin{cases} \frac{x^{(k+1)} - x^{(k)}}{h} = -\eta^{(k)} \nabla f\left(x^{(k)}\right) \\ f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) \leq \eta^{(k)} \left\langle \nabla f\left(x^{(k)}\right), x^{(k+1)} - x^{(k)} \right\rangle \end{cases}$$

- 離散散逸則が成立.
- 満たすべき不等式 $F_h(\eta) \leq 0$ について：
  $\eta \in [0, (1 + Lh/2)^{-1}]$ のとき，$F_h(\eta) \leq 0$
  $\rightarrow$ バックトラッキングで求解可能.
- 特殊なステップ幅基準を採用した最急降下法にともみなせる.

構造保存数値解法の適用 (4/4)：提案手法の性質

定理：収束レート

f の条件	収束レート	勾配流の収束レート
	$\min_{0\le i\le k}\left\|\nabla f\left(x^{(k)}\right)\right\| = \mathrm{O}\left(k^{-1/2}\right)$	$\mathrm{O}\left(t^{-1/2}\right)$
凸	$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(k^{-1}\right)$	$\mathrm{O}\left(t^{-1}\right)$
μ-強凸	$f\left(x^{(k)}\right) - f^\star = \mathrm{O}\left(\exp\left(\frac{-2\mu kh}{(Lh+2)^2}\right)\right)$	$\mathrm{O}(\exp(-2\mu t))$

- 散逸則のおかげで証明は比較的シンプル
- 任意の $h > 0$ に対して収束が保証されるが，実際の挙動は h に依存
  $\to$ h を適応的に変更する手法も提案.
- 数値実験では，通常の最急降下法と同程度のパフォーマンス (当然?)

事実

関数 $f\colon \mathbb{R}^d \to \mathbb{R}^d$ が L-平滑な凸関数であるとき，
勾配流 $\dot{x} = -\nabla f(x)$ の任意の解 x について，

$$f(x(t)) - f^\star = \mathrm{O}\left(\frac{1}{t}\right).$$

時間スケーリング $t = \alpha(\tau)$ により，$\widetilde{x}(\tau) = x(\alpha(\tau))$ とすると...

$$\text{ODE：} \quad \frac{\mathrm{d}}{\mathrm{d}\tau}\widetilde{x}(\tau) = -\alpha'(\tau)\nabla f(\widetilde{x}(\tau))$$

$$\text{収束レート：} \quad f(\widetilde{x}(\tau)) - f^\star = \mathrm{O}\left(\frac{1}{\alpha(\tau)}\right)$$

任意の収束レートが達成可能...？

本質的収束レート (2/5)：観察

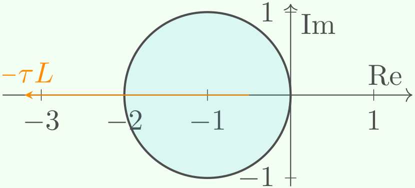
収束レートが不定であれば,

- 「最適な」連続時間ダイナミクスという問いは無意味
- 離散時間ダイナミクスにおける最適レートの理論と不整合

$\longrightarrow$ 我々のアプローチ：
「最終的に時間離散化する」ことを前提に「本質的な収束レート」を定義

観察

例：$\mathrm{O}\left(1/\tau^2\right)$ を達成する ODE $\frac{\mathrm{d}}{\mathrm{d}\tau}\widetilde{x}(\tau) = -\tau \nabla f(\widetilde{x}(\tau))$
線形化した場合の最小固有値 $= -\tau L$：



固有値が左に動いていく
$\rightarrow$ ステップ幅を小さくする必要あり

アイデア

- 時間スケーリングでうつりあう ODE は同値類でまとめる
- 「本質的な収束レート」は「同値類」に対して定義する

定義：適切な代表元と収束レート

漸近的に最適解に向かうベクトル場の集合 $\mathcal{G}$ 上の同値関係 $\sim$ を

$$g \sim \widetilde{g} \iff \exists\,\text{時間スケーリング}\ \alpha\ \text{s.t.}\ \alpha'(t)g(\alpha(t),x) = \widetilde{g}(t,x)$$

で定め，g の同値類を $[g]$ と書く．$g \in \mathcal{G}$ に対して，$g_0 \in [g]$ が $t \to \infty$ で $\rho\left(\frac{\partial g_0(t,x)}{\partial x}\right) = \Theta(1)$ を満たすとき，g_0 を $[g]$ の適切な代表元という．適切な代表元 g_0 に対応する解 x が $f(x(t)) - f^\star = \Theta(\beta(t))$ を満たすとき，β を $[g]$ の収束レートという．

- $[\dot{x} = -\nabla f(x)]$ の収束レートは $1/t$
- $\left[\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0\right]$ の収束レートは $1/t^2$

本質的収束レート (4/5)：主定理

定理

仮定：

(A0) 数値解法の安定領域は有界.

(A1) 任意の k において，$h_k \frac{\partial g}{\partial x}\left(t_{k-1}, x^{(k-1)}\right)$ の全ての固有値が安定領域に含まれる ($t_k := \sum_{i=1}^{k} h_i$).

(A2) ある $h_{\max} \in \mathbb{R}_{>0}$ が存在し，任意の k で $h_k \leq h_{\max}$ が成立.

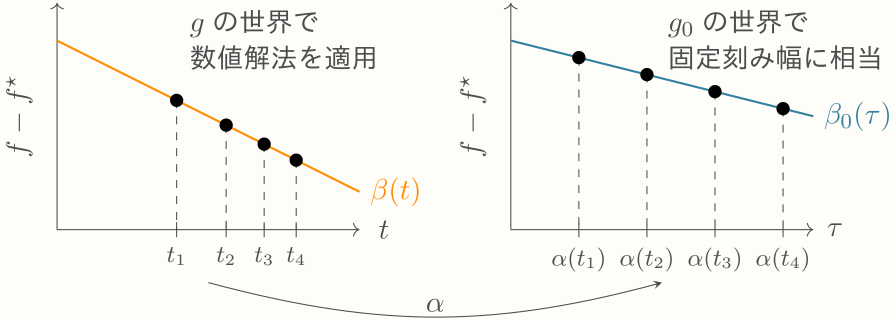
$g \in \mathcal{G}$ に対して，g_0 が $[g]$ の適切な代表元であり，α が g と g_0 の間の時間スケーリングとする ($g(x,t) = \alpha'(t) g_0(\alpha(t), x)$)．このとき，任意の k に対して $0 < \rho\left(\frac{\partial g}{\partial x}\left(t_{k-1}, x^{(k-1)}\right)\right) < \infty$ ならば，$\alpha(t_k) = \mathrm{O}(k)$ が成立.

仮定 (A0) について：
陽的解法を想定すると，この仮定は妥当.
例：全ての陽的 Runge–Kutta 法の安定領域は有界

本質的収束レート (5/5)：主定理の意味

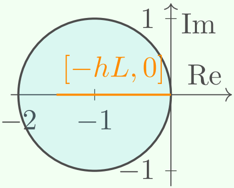
定理

(A0), (A1), (A2) を仮定する．
$g \in \mathcal{G}$ に対して，g_0 が $[g]$ の適切な代表元であり，α が g と g_0 の間の時間スケーリングとする ($g(t,x) = \alpha'(t) g_0(\alpha(t), x)$)．このとき，任意の k に対して $0 < \rho\left(\frac{\partial g}{\partial x}\left(t_{k-1}, x^{(k-1)}\right)\right) < \infty$ ならば，$\alpha(t_k) = \mathrm{O}(k)$ が成立．



安定な陽解法の適用 (1/5)：安定領域とステップ幅

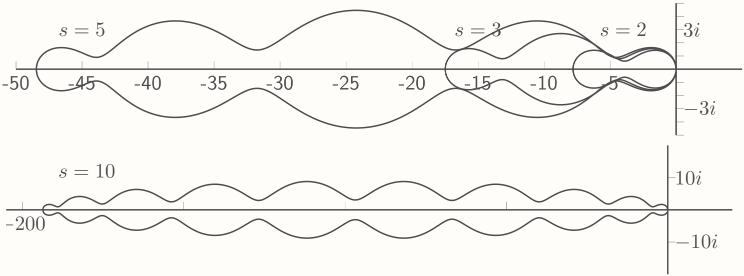
安定領域：
$(|\lambda h + 1| \leq 1)$



L 平滑凸関数に対する
勾配流では $\lambda \in [-L, 0]$

$\to h \in \left[0, \frac{2}{L}\right]$ で安定
(最急降下法の収束条件と一致)

$\to$ 実軸の負の方向に広い安定領域をもつ陽解法は，最適化に使えそう.



ある s 段陽的 Runge–Kutta 法の安定領域

安定な陽解法の適用 (2/5)：安定領域の広い RK 法

通常のアプローチ：
数値解法の構成 (精度などを基準とする) $\to$ 安定領域を調べる

広い安定領域をもつ解法を構成するアプローチ：
安定領域を作る $\to$ それに合うような数値解法を作る.

具体例 (前ページの図)： RKC (Runge–Kutta–Chebyshev) 法
s 次 Chebyshev 多項式 T_s を「うまく」シフト，スケーリング：

$$R_s(z) := \frac{T_s(w_0 + w_1 z)}{T_s(w_0)} \qquad w_0 := 1 + \frac{\eta}{s^2} \qquad w_1 := \frac{T_s(w_0)}{T_s'(w_0)}$$

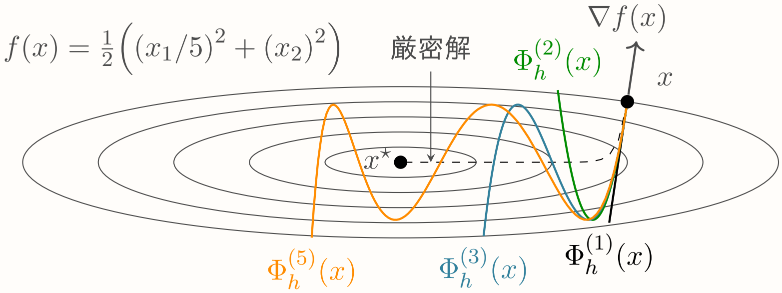
- 安定領域の左端 $\approx -2T_s'(1) = -2s^2$
- 陽的 Euler 法 s 個の合成で実現可能

	最大ステップ幅	勾配の評価回数	比
陽的 Euler	2	1	2
s 次の場合	$2s^2$	s	$2s$

安定な陽解法の適用 (3/5)：s 段 RKC 法のイメージ

$\Phi_h^{(s)}(x)$：初期値 x に対して s 段 RKC 法を 1 回適用した際の出力.

$s = 1, 2, 3, 5$ のそれぞれで，h を動かした曲線：



- s を大きくしても (精度は上がっていないので) 厳密解には近づかない.
- s が大きい RKC 法は非常に大きな h を使える ($2s^2/L$ 程度).

安定な陽解法の適用 (4/5)：RKC 法の実装について

同じ安定性関数をもつ複数の Runge–Kutta 法が存在

- 微分方程式が線形であれば，出力は同じ (安定性関数のみで定まる)
- 所望の安定性関数が s 次多項式 $\Rightarrow$ s 段陽的 Runge–Kutta 法

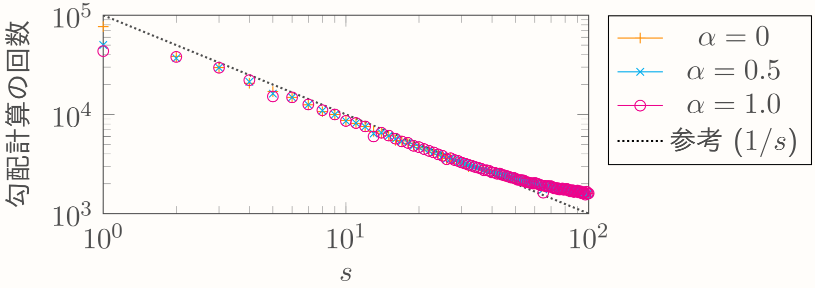
実装の方法：

- 陽的 Euler 法の合成：
  Chebyshev 多項式の根を用いる．根の並べ方に注意が必要．
- 2 段 RK 法の合成：
  Chebyshev 多項式の根のペアを用いる．ペアの並べ方に注意が必要．
- Chebyshev 多項式の漸化式を用いる方法：
  - Chebyshev 多項式の 3 項間漸化式を用いて定める．
  - 内部段も安定に計算でき，計算手順に任意性がない．
  - 強凸 2 次関数の最適化への応用 [Eftekhari et al., 2021]
  - Chebyshev iterative method (最適化), Chebyshev 加速法 (線形計算) と関係がある．

目的関数 (強凸) ($d = 128$, $L \approx 512 + \alpha$, $\mu \approx 0.0759$, $\kappa \approx 6760$)：

$$f(x) = \frac{d}{2}\left(x_1^2 + \sum_{i=2}^{d}(x_i - x_{i-1})^2 + x_d^2\right) + \alpha \log\left(\sum_{i=1}^{d}\exp(x_i)\right).$$

勾配ノルムが 10^{-10} 以下になるまでの勾配計算の回数：



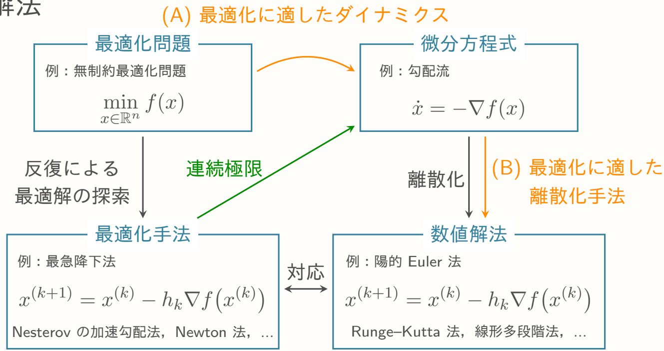
強凸ではない (凸でもない) 場合の数値実験でもよく動く.

Table of contents

まとめ

常微分方程式の数値解析の基礎をざっくりと説明した.

- Runge–Kutta 法
- 線形多段階法
- 安定性
- 収束性 (精度)
- 構造保存数値解法



[Besse, 2004] Besse, C. (2004).
A relaxation scheme for the nonlinear Schrödinger equation.
SIAM Journal on Numerical Analysis, 42(3):934–952.

[Butcher, 1964] Butcher, J. C. (1964).
Implicit Runge-Kutta processes.
Mathematics of Computation, 18:50–64.

[Chen et al., 2018] Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018).
Neural ordinary differential equations.
In *Advances in Neural Information Processing Systems*, volume 31.

[Cheng et al., 2020] Cheng, Q., Liu, C., and Shen, J. (2020).
A new Lagrange multiplier approach for gradient flows.
Computer Methods in Applied Mechanics and Engineering, 367:113070.

[Eftekhari et al., 2021] Eftekhari, A., Vandereycken, B., Vilmart, G., and Zygalakis, K. C. (2021).
Explicit stabilised gradient descent for faster strongly convex optimisation.
BIT Numerical Mathematics, 61:119–139.

[Ehrhardt et al., 2018] Ehrhardt, M. J., Riis, E. S., Ringholm, T., and Schönlieb, C.-B. (2018).
A geometric integration approach to smooth optimisation: Foundations of the discrete gradient method.
arXiv, arXiv:1805.06444.

[Gonzalez, 1996] Gonzalez, O. (1996).
Time integration and discrete Hamiltonian systems.
Journal of Nonlinear Science, 6:449–467.

[Hairer, 1978] Hairer, E. (1978).
A Runge–Kutta method of order 10.
IMA Journal of Applied Mathematics, 21(1):47–59.

参考文献 II

[Hairer et al., 2006] Hairer, E., Lubich, C., and Wanner, G. (2006).
Geometric numerical integration, Structure-preserving algorithms for ordinary differential equations, volume 31 of *Springer Series in Computational Mathematics*.
Springer-Verlag, Berlin, second edition.

[Hairer et al., 1993] Hairer, E., Nørsett, S. P., and Wanner, G. (1993).
Solving ordinary differential equations I, Nonstiff problems, volume 8 of *Springer Series in Computational Mathematics*.
Springer-Verlag, Berlin, second edition.

[Hairer and Wanner, 1996] Hairer, E. and Wanner, G. (1996).
Solving ordinary differential equations II, Stiff and differential-algebraic problems, volume 14 of *Springer Series in Computational Mathematics*.
Springer-Verlag, Berlin.

[Itoh and Abe, 1988] Itoh, T. and Abe, K. (1988).
Hamiltonian-conserving discrete canonical equations based on variational difference quotients.
Journal of Computational Physics, 76(1):85–102.

[Kim and Yang, 2023] Kim, J. and Yang, I. (2023).
Unifying Nesterov's accelerated gradient methods for convex and strongly convex objective functions.
In *Proceedings of the 40th International Conference on Machine Learning*, volume 202.

[Krichene et al., 2015] Krichene, W., Bayen, A., and Bartlett, P. L. (2015).
Accelerated mirror descent in continuous and discrete time.
In *Advances in Neural Information Processing Systems*, volume 28.

[Matsuo and Furihata, 2001] Matsuo, T. and Furihata, D. (2001).
Dissipative or conservative finite-difference schemes for complex-valued nonlinear partial differential equations.
Journal of Computational Physics, 171(2):425–447.

[McLachlan et al., 1999] McLachlan, R. I., Quispel, G. R. W., and Robidoux, N. (1999).
Geometric integration using discrete gradients.
Philosophical Transactions of the Royal Society A, 357:1021–1045.

[Nesterov, 2004] Nesterov, Y. (2004).
Introductory lectures on convex optimization: A basic course, volume 87.
Springer Science & Business Media.

[Nozawa, 2022] Nozawa, R. (2022).
Analysis and construction of continuous optimization algorithms via variable step size linear multistep methods (in Japanese).
Bachelor's thesis, The University of Tokyo.

[Onuma and Sato, 2023] Onuma, K. and Sato, S. (2023).
Existence results on Lagrange multiplier approach for gradient flows and application to optimization.
to appear in Japan Journal of Industrial and Applied Mathematics.

[Polyak, 1964] Polyak, B. T. (1964).
Some methods of speeding up the convergence of iteration methods.
USSR Computational Mathematics and Mathematical Physics, 4(5):1–17.

[Quispel and McLaren, 2008] Quispel, G. R. W. and McLaren, D. I. (2008).
A new class of energy-preserving numerical integration methods.
Journal of Physics A: Mathematical and Theoretical, 41:045206.

[Riis et al., 2022] Riis, E. S., Ehrhardt, M. J., Quispel, G., and Schönlieb, C.-B. (2022).
A geometric integration approach to nonsmooth, nonconvex optimisation.
Foundations of Computational Mathematics, 22:1351–1394.

[Ringholm et al., 2018] Ringholm, T., Lazic, J., and Schönlieb, C.-B. (2018).
Variational image regularization with Euler's elastica using a discrete gradient scheme.
SIAM Journal on Imaging Sciences, 11(4):2665–2691.

[Scieur et al., 2017] Scieur, D., Roulet, V., Bach, F., and d'Aspremont, A. (2017).
Integration methods and optimization algorithms.
In *Advances in Neural Information Processing Systems*, volume 30.

[Shen et al., 2018] Shen, J., Xu, J., and Yang, J. (2018).
The scalar auxiliary variable (SAV) approach for gradient flows.
Journal of Computational Physics, 353:407–416.

参考文献 IV

[Su et al., 2014] Su, W., Boyd, S., and Candes, E. (2014).
A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights.
In *Advances in Neural Information Processing Systems*, volume 27.

[Ushiyama et al., 2022a] Ushiyama, K., Sato, S., and Matsuo, T. (2022a).
Deriving efficient optimization methods based on stable explicit numerical methods.
JSIAM Letters, 14:29–32.

[Ushiyama et al., 2022b] Ushiyama, K., Sato, S., and Matsuo, T. (2022b).
Essential convergence rate of ordinary differential equations appearing in optimization.
JSIAM Letters, 14:119–122.

[Ushiyama et al., 2023] Ushiyama, K., Sato, S., and Matsuo, T. (2023).
A new unified framework for designing convex optimization methods with prescribed theoretical convergence estimates: A numerical analysis approach.
arXiv, arXiv:2302.07404.

[Yang and Han, 2017] Yang, X. and Han, D. (2017).
Linearly first- and second-order, unconditionally energy stable schemes for the phase field crystal model.
Journal of Computational Physics, 330:1116–1134.

[Zhang, 2019] Zhang, D. K. (2019).
Discovering new Runge-Kutta methods using unstructured numerical search.
arXiv, arXiv:1911.00318.