RIMS 共同研究 (公開型)
数値解析が拓く次世代情報社会〜エッジから富岳まで〜

連続最適化に対する数値解析学的アプローチ

佐藤 峻，牛山寛生，松尾宇泰

東京大学

2022 年 10 月 12 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，...

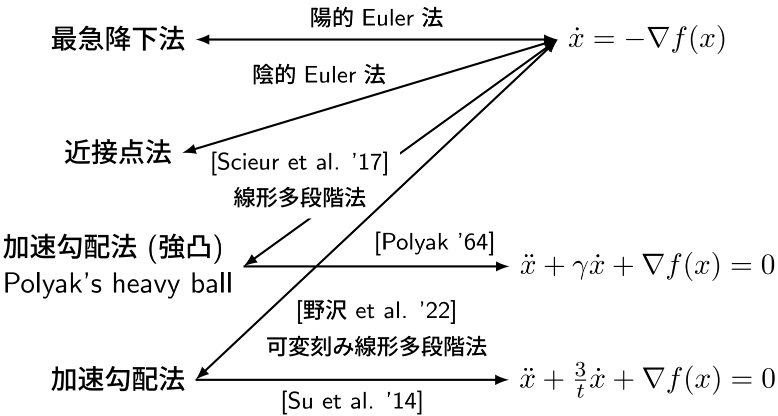
数値解析の視点

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
$\rightarrow$ 数値解析の視点は連続最適化に有用である (と期待される)

最適化手法と微分方程式の関係の例



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

連続最適化と微分方程式の数値解析

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，…

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

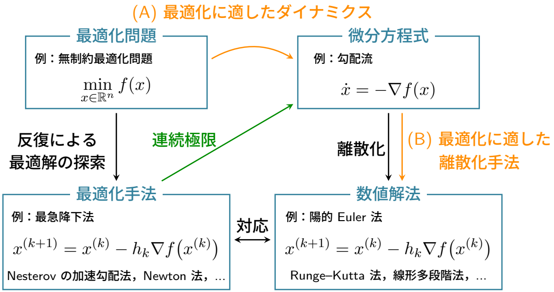
数値解法

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–S. '22]
Lagrange 未定乗数法 アプローチ [Cheng–Liu–Shen '20] の勾配流への適用

(A) **本質的な収束レート** [Ushiyama–S.–Matsuo '22]
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–S.–Matsuo '22]：今日の話題

(B) **弱い離散勾配を用いた手法の構成と統一的収束証明** [牛山–佐藤–松尾 '22]
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** [野沢–松尾–佐藤 '22]
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

Table of contents

1. 最適化における一次法と最適レート
2. 【本研究】 安定な陽的解法
3. 既存研究との関係
4. まとめ

準備

扱う問題：無制約連続最適化問題

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

仮定：

- f が L-平滑 (L-smooth) ：勾配 ∇f が L-Lipschitz 連続．
- $\operatorname{argmin} f \neq \emptyset$：最適解 $x^\star$ が存在 ($f^\star := f(x^\star)$)

よく使う用語や記法：

- $\langle \cdot, \cdot \rangle$, $\|\cdot\|$：$\mathbb{R}^n$ の通常の内積と 2-ノルム．
- (関数 f の) 凸性：

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \qquad (\forall\, x, y \in \mathbb{R}^n, \lambda \in [0,1])$$

- μ-強凸性：$f(x) - \frac{\mu}{2}\|x\|^2$ が凸関数．
  (μ-強凸 $\Rightarrow$ 凸)

最適化手法における「一次法」の枠組

直感的な説明

一次法：現時点までの勾配のみを用いて反復を行う手法の総称

一次法の例：

- **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 (強凸)** [Nesterov '83]：

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

- **非線形共役勾配法**：

$$\begin{cases} d^{(k)} = -\nabla f\left(x^{(k)}\right) + \beta_k d^{(k-1)} \\ x^{(k+1)} = x^{(k)} + h_k d^{(k)} \\ \beta_{k+1} = \dfrac{\left\|\nabla f\left(x^{(k+1)}\right)\right\|^2}{\left\|\nabla f\left(x^{(k)}\right)\right\|^2} \end{cases}$$

最適化手法における「一次法」の枠組

直感的な説明

一次法：現時点までの勾配のみを用いて反復を行う手法の総称

一次法の例：

- 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 (強凸) [Nesterov '83]：

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

一次法ではない例：Newton 法

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

最適化手法における「一次法」の枠組

直感的な説明

一次法：現時点までの勾配のみを用いて反復を行う手法の総称

一次法の例：

- 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 (強凸) [Nesterov '83]：

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

定義 (一次法：first order methods) cf. [Nesterov '04]

一次法 $\iff$ 点列 $\left\{x^{(k)}\right\}$ が以下の性質を満たす：

$$
x^{(k+1)} - x^{(0)} \in \operatorname{span}\left\{\nabla f\left(x^{(0)}\right), \nabla f\left(x^{(1)}\right), \ldots, \nabla f\left(x^{(k)}\right)\right\} \quad (\forall k \geq 0)
$$

μ-強凸関数に対する最適レート

最急降下法は，$h_k = 2/(\mu + L)$ の下で以下を満たす ($\kappa := L/\mu$：条件数)：

$$f\left(x^{(k)}\right) - f^\star \leq \frac{L}{2}\left(\frac{\kappa - 1}{\kappa + 1}\right)^{2k}\left\|x^{(0)} - x^\star\right\|^2.$$

この収束レートは満足できるものか？

μ-強凸関数に対する最適レート

最急降下法は，$h_k = 2/(\mu + L)$ の下で以下を満たす ($\kappa := L/\mu$：条件数)：

$$f\left(x^{(k)}\right) - f^\star \leq \frac{L}{2}\left(\frac{\kappa - 1}{\kappa + 1}\right)^{2k}\left\|x^{(0)} - x^\star\right\|^2.$$

この収束レートは満足できるものか？

$\rightarrow$ No! 下界が知られている (達成する手法も存在)：

μ-強凸関数に対する一次法の最適レート

適切な仮定の下，

(2 次) cf. [Nesterov '04] $$f\left(x^{(k)}\right) - f^\star \geq \frac{\mu}{2}\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^{2k}\left\|x^{(0)} - x^\star\right\|^2,$$

(一般) [Drori–Taylor '22] $$f\left(x^{(k)}\right) - f^\star \geq \frac{\mu}{2}\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa}}\right)^{2k}\left\|x^{(0)} - x^\star\right\|^2.$$

Table of contents

1. 最適化における一次法と最適レート
2. 【本研究】 安定な陽的解法
3. 既存研究との関係
4. まとめ

最急降下法の収束条件の数値解析的解釈

凸関数に対する最急降下法の収束条件 cf. [Nesterov '04, 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？ $\to$ 線形安定性解析

観察

- 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．

Dahlquist のテスト方程式と安定領域 cf. [Hairer–Wanner '96]

安定領域

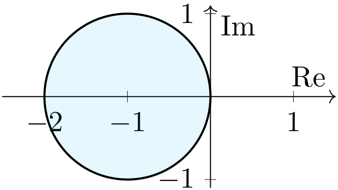
Dahlquist のテスト方程式 ($\lambda \in \mathbb{C}$ は定数)

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

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**，$\{z \in \mathbb{C} \mid |R(z)| \le 1\}$ を**安定領域**という．

陽的 Euler 法の安定性関数は $R(z) = 1 + z$

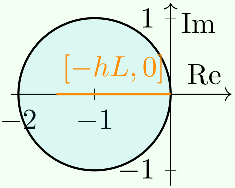
安定領域：
($|\lambda h + 1| \le 1$)



前頁の設定では $\lambda \in [-L, 0]$
$\to h \in [0, 2/L]$ で安定
(最急降下法の収束条件と一致)

より広い安定領域をもつ数値解法へ

安定領域：
$(|\lambda h + 1| \leq 1)$



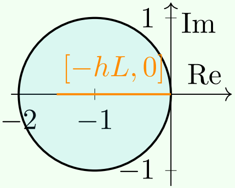
今の設定では $\lambda \in [-L, 0]$
$\to h \in [0, 2/L]$ で安定
(最急降下法の収束条件と一致)

$\to$ より広い安定領域をもつ数値解法は，最適化に適していそう．

より広い安定領域をもつ数値解法へ

安定領域：

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



今の設定では $\lambda \in [-L, 0]$
$\rightarrow h \in [0, 2/L]$ で安定
(最急降下法の収束条件と一致)

$\rightarrow$ より広い安定領域をもつ数値解法は，最適化に適していそう．

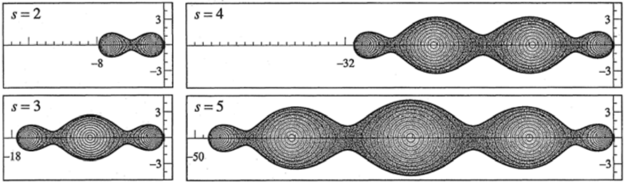


Fig. 2.13. Stability domains for *damped* Chebyshev stability functions, $\varepsilon = 0.05$

[Hairer–Wanner '96] より引用．

広い安定領域をもつ陽的解法の構成

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)
s 次 Chebyshev 多項式 T_s を「うまく」シフト，スケーリング：

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

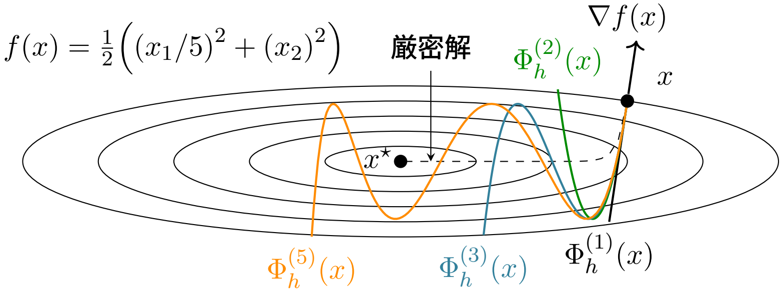
- 安定領域の左端 $\approx -2T_s'(1) = -2s^2$
- 陽的 Euler 法 s 個の合成で実現可能

	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	$2s$

s 段 RKC (Runge–Kutta–Chebyshev) 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)．

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：

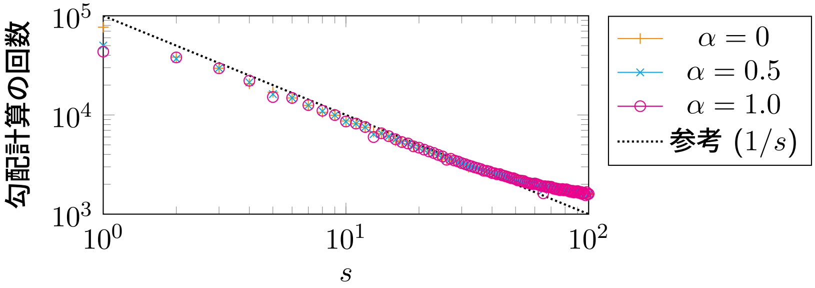
- 陽的 Euler 法の合成 [Saul'ev '60],[Guillou–Lago '61]：
  - Chebyshev 多項式の根を用いる．
  - 根を用いる順番に任意性があり，注意が必要．
- 2 段 RK 法の合成 [Lebedev '89]：
  - Chebyshev 多項式の根を 2 つづつペアにして用いる．
  - やはり順番には任意性があり，注意が必要．
- Chebyshev 多項式の漸化式を用いる方法 [van der Houwer–Sommeijer '80]：
  - Chebyshev 多項式の 3 項間漸化式を用いて定める．
  - 内部段も安定に計算でき，計算手順に任意性がない.

数値実験

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：

K. Ushiyama, S., T. Matsuo: Deriving efficient optimization methods based on stable explicit numerical methods, JSIAM Letters, 14:29–32, 2022.

Table of contents

1. 最適化における一次法と最適レート
2. 【本研究】 安定な陽的解法
3. 既存研究との関係
4. まとめ

関連する既存研究

目的関数が強凸 2 次関数の場合 (=線形方程式の場合) の研究

- RKC 法を勾配流に適用 [Eftekhari–Vandereycken–Vilmart–Zygalakis '21]:
  - [van der Houwer–Sommeijer '80] の実装を利用
  - スキームパラメータを上手く選ぶと，$s \to \infty$ で (ほぼ) 最適レート
  - 「強凸 2 次関数+小さい凸関数」の場合への拡張
- Chebyshev iterative method：
  - 強凸 2 次関数専用の最適化手法
  - 最適なレートを達成
- 線形計算における Chebyshev 加速法 (cf. [杉原–室田 '09])

(再掲) μ-強凸関数に対する一次法の最適レート

適切な仮定の下

(2 次) cf. [Nesterov '04]
$$f\left(x^{(k)}\right) - f^\star \geq \frac{\mu}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k}\left\|x^{(0)} - x^\star\right\|^2,$$

(一般) [Drori–Taylor '22]
$$f\left(x^{(k)}\right) - f^\star \geq \frac{\mu}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}}\right)^{2k}\left\|x^{(0)} - x^\star\right\|^2.$$

強凸 2 次関数に対する RKC 法 (1/2) [Eftekhari et al. '21]

RKC 法 (漸化式を利用した実装)

事前に定めるパラメータ：s, η, h.
上のパラメータから事前に計算しておく量：

$$w_0 := 1 + \frac{\eta}{s^2}, \quad w_1 := \frac{T_s(w_0)}{T_s'(w_0)}, \quad \nu_i := \frac{2w_0 T_{i-1}(w_0)}{T_i(w_0)} \quad (i = 1, \ldots, s)$$

反復：

$$X_0^{(k)} = x^{(k)}, \qquad X_1^{(k)} = X_0^{(k)} - h\frac{w_1}{w_0}\nabla f\left(X_0^{(k)}\right),$$

$$X_i^{(k)} = -h\nu_i \frac{w_1}{w_0}\nabla f\left(X_{i-1}^{(k)}\right) + \nu_i X_{i-1}^{(k)} - (\nu_i - 1)X_{i-2}^{(k)} \quad (i = 2, 3, \ldots, s),$$

$$x^{(k+1)} = X_s^{(k)}.$$

強凸 2 次関数に対する RKC 法 (2/2) [Eftekhari et al. '21]

<u>方針</u> η を固定して，s と h を最適に選ぶ：

$$s = \left\lceil \sqrt{\frac{(\kappa-1)\eta}{2}} \right\rceil, \qquad h = \frac{w_0 - 1}{w_1 \mu}.$$

- $C(\eta) := \frac{1}{T_s(w_0)}$ とすると (1 ステップで勾配を s 回計算していることに注意)，

$$f\left(x^{(k+1)}\right) - f^\star \le (C(\eta))^{2s}\left(f\left(x^{(k)}\right) - f^\star\right).$$

- 収束レートを表す $C(\eta)$ について，

$$\lim_{\eta\to\infty} C(\eta) = \underbrace{\boxed{\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}}}_{\text{最適レート}} + \mathrm{O}\left(\kappa^{-3/2}\right).$$

s を固定してやりなおせば少し改善できる (次ページ)．

簡単な改善

方針 s を固定して，η と h を最適に選ぶ：

$$\eta = \frac{2s^2}{\kappa - 1}, \qquad h = \frac{w_0 - 1}{w_1 \mu}.$$

- $C(s) := \frac{1}{T_s(w_0)}$ とすると (1 ステップで勾配を s 回計算していることに注意)，

$$f\left(x^{(k+1)}\right) - f^\star \leq (C(s))^{2s}\left(f\left(x^{(k)}\right) - f^\star\right).$$

- 収束レートを表す $C(s)$ について，

$$\lim_{s \to \infty} C(s) = \underbrace{\boxed{\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}}}_{\text{最適レート}}.$$

前述の数値実験はこの実装に基づく．

Chebyshev iterative method (1/2)[1]

ある程度一般的な 1 次法

$$x^{(k)} = x^{(k-1)} + \sum_{i=0}^{k-2} c_i^{(k-1)} \left(x^{(i+1)} - x^{(i)} \right) + c_{k-1}^{(k-1)} \nabla f \left(x^{(k-1)} \right)$$

の中で，2 次関数に対する収束レートが最適になるようなパラメータの決定
$\iff$ 上の手法に関係する多項式 P について，

$$\min_{P} \max_{\lambda \in [\mu, L]} |P(\lambda)|.$$

$\rightarrow$ P を Chebyshev 多項式を用いて定めるのが最適．

ここから遡ってパラメータを定めたものが Chebyshev iterative method．
(RKC 法のときと似た導出になっている．)

[1] cf. "F. Pedregosa, Residual Polynomials and the Chebyshev method, URL: http://fa.bianp.net/blog/2020/polyopt/"

Chebyshev iterative method (2/2)[2]

Chebyshev iterative method

事前に計算しておく量：

$$\rho = \frac{\kappa - 1}{\kappa + 1}, \qquad \nu_1 = 2, \qquad x^{(1)} = -\frac{2}{L + \mu} \nabla f\left(x^{(0)}\right).$$

反復 ($k = 2, \dots$)**：**

$$\nu_k = \left(1 - \frac{\rho^2}{4} \nu_{k-1}\right)^{-1},$$

$$x^{(k)} = x^{(k-1)} + (\nu_k - 1)\left(x^{(k-1)} - x^{(k-2)}\right) - \nu_k \frac{2}{L + \mu} \nabla f\left(x^{(k-1)}\right).$$

cf. RKC 法の内部段の計算 (を整理したもの)：

$$X_i^{(k)} = X_{i-1}^{(k)} + (\nu_i - 1)\left(X_{i-1}^{(k)} - X_{i-2}^{(k)}\right) - \nu_i h \frac{w_1}{w_0} \nabla f\left(X_{i-1}^{(k)}\right).$$

[2]cf. "F. Pedregosa, Residual Polynomials and the Chebyshev method, URL: http://fa.bianp.net/blog/2020/polyopt/"

RKC 法と Chebyshev iterative method

RKC 法

陽的 Euler 法の合成

2 段法の合成

漸化式を用いた実装

η を固定した最適スキーム

s を固定した最適スキーム

$s \to \infty$ かつ $h \to \infty$ として「1 回」で終わらせる

Chebyshev iterative method

s 回ごとにリスタート

- 両者「最善」を目指した結果なので，同じ漸化式が生まれるのはある意味で必然．
- 2 次関数であり，μ と L が精密に分かっている場合には，リスタートは無益．
- RKC 法をうまく使うことで，2 次関数以外の関数に拡張できることを期待．

Table of contents

1. 最適化における一次法と最適レート
2. 【本研究】 安定な陽的解法
3. 既存研究との関係
4. まとめ

まとめ

(B) 構造保存数値解法の適用 [Onuma–S. '22]

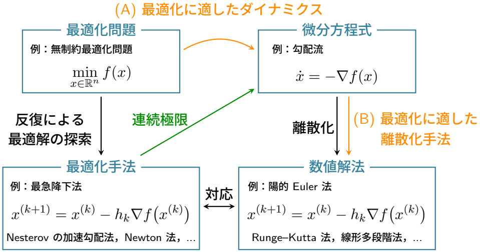
(A) 本質的な収束レート [Ushiyama–S.–Matsuo '22]

(B) 安定な陽解法 (RKC 法) の適用 [Ushiyama–S.–Matsuo '22]：今日の話題

- 線形安定性解析の観点から，勾配流に適した数値解法 (RKC 法) を選択
- 既存手法との関係の整理

(B) 弱い離散勾配を用いた手法の構成と統一的収束証明 [牛山–佐藤–松尾 '22]

(B) 加速勾配法の可変刻み線形多段階法解釈 [野沢–松尾–佐藤 '22]



今後の課題：

(B) 一般の非線形の場合の RKC 法の収束レートの証明

(A) より良いダイナミクス (ODE) の提案

(B) より複雑な最適化 ODE に対する離散化の検討