非線形方程式の解の精度保証 (+ 自動微分)

柏木 雅英
kashi@waseda.jp
http://verifiedby.me/

早稲田大学 基幹理工学部 応用数理学科

非線形方程式

非線形方程式

$$f(x) = 0, \quad f : \mathbb{R}^n \to \mathbb{R}^n$$

の解を精度保証することを考える。

- 解が explicit に $x^* = \cdots$ と数式で書けるなら、それを区間演算で計算すればよい。
- しかし、ほとんどの場合解の explicit な表記は無く、反復法で求めざるを得ない。
- 解が何らかの反復

  $$x_{i+1} = g(x_i)$$

  で計算されるとき、その計算に混入する誤差を区間演算で把握したとしても解の精度保証にはならない。個々の x_i の誤差を把握することは出来ても、収束先の包含を得ることはできない。
- そもそも非線形方程式は、解が存在するかどうかすら分からないので、非線形方程式の解の精度保証をすることは解の存在保証をすることと同じ。

二分法

$n=1$、すなわち、$f(x)=0, \quad f:\mathbb{R}^1 \to \mathbb{R}^1$ という1変数の問題なら、いわゆる二分法で簡単に解を精度保証付きで求めることができる。

定理

$f(x)=0, \quad f:\mathbb{R}\to\mathbb{R}$ を連続関数とする。区間 $I=[a,b]$ に対して、

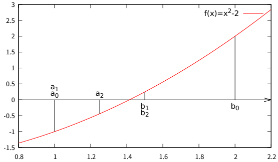
- $f(a)<0$ かつ $f(b)>0$
- $f(a)>0$ かつ $f(b)<0$

のいずれかが成立するならば、区間 I に方程式 $f(x)=0$ の解が存在する(唯一とは限らない)。

証明は中間値の定理より。$f(a)$ 及び $f(b)$ の計算の誤差で符号判定を誤ってはまずいので、区間演算を行って、

- $\overline{f([a,a])}<0$ かつ $\underline{f([b,b])}>0$
- $\underline{f([a,a])}>0$ かつ $\overline{f([b,b])}<0$

のいずれかの成立を判定する。



Krawczyk 法 (Krawczyk(1969), Kahan(1968))

区間演算を用いて、非線形方程式 $f(x) = 0, \quad f : \mathbb{R}^n \to \mathbb{R}^n$ の解の存在を保証する。

Krawczyk 法

$I \subset \mathbb{R}^n$ は区間ベクトル (候補者集合)、$c = \mathrm{mid}(I)$、$R \simeq f'(c)^{-1}$、E は単位行列とし、

$$K(I) = c - Rf(c) + (E - Rf'(I))(I - c)$$

としたとき、$K(I) \subset \mathrm{int}(I)$ ならば I に $f(x) = 0$ の解が唯一存在する。
($\mathrm{int}(I)$: I の内部)

証明

$g(x) = x - Rf(x)$ に対して平均値形式と縮小写像原理を適用する。

Krawczyk 法は $f'(I)$ (区間 I における f の全てのヤコビ行列を包含する区間行列) を持つ $\Longrightarrow$ 自動微分が必要

Krawczyk 法を導く (1)

縮小写像の原理

X を $d(x,y)$ を距離とする完備距離空間とし、$g: X \to X$ を全ての $x, y \in X$ に対して

$$d(g(x), g(y)) \le \alpha d(x,y)$$

を満たすような写像とする。但し、$\alpha < 1$ とする。(このような g を縮小写像と言い、実数 α を縮小定数と言う。) このとき、以下が成立する:

1. $g(x) = x$ を満たす点 x^* (不動点と言う) が X に唯一存在する。
2. x_0 を X 内の任意の点とし、点列 $\{x_k\}$ を $x_{k+1} = g(x_k)$ で定めると、x_k は x^* に収束し、その収束速度は $d(x_k, x^*) \le \dfrac{\alpha^k}{1-\alpha} d(x_1, x_0)$ で表される。
3. X 内の任意の点 x_0 に対して、$d(x_0, x^*) \le \dfrac{d(g(x_0), x_0)}{1-\alpha}$ が成立。

Krawczyk 法を導くため、縮小写像原理を以下のように読み替える。

X	:	$\mathbb{R}^n$ に含まれる区間ベクトル (閉集合)
$d(x,y)$	:	$\|x - y\|$ 。但しノルムは $\|x\| = \max \lvert x_i \rvert$ (最大値ノルム)

Krawczyk 法を導く (2)

非線形方程式

$$f(x) = 0, \quad f : \mathbb{R}^n \to \mathbb{R}^n$$

の解が区間ベクトル $I \subset \mathbb{R}^n$ に存在することを保証する。そのため、$c = \mathrm{mid}(I)$ として、行列 $R \simeq f'(c)^{-1}$ を用いて、写像

$$g(x) = x - Rf(x)$$

を考える。R が逆行列を持つならば

$$f(x) = 0 \Longleftrightarrow g(x) = x$$

である。g が I において縮小写像原理の仮定を満たすことを示すには、

1. $\{g(x) \mid x \in I\} \subset I$ (g が I から I への写像であること)
2. g が I 上で縮小写像であること。

の 2 つを示せばよい。

Krawczyk 法を導く (3)

条件 1

$g(x) = x - Rf(x)$ に対して平均値形式を適用して、(E は単位行列)

$$\begin{aligned} K(I) &= g(c) + g'(I)(I - c) \\ &= c - Rf(c) + (E - Rf'(I))(I - c) \end{aligned}$$

のような写像 K (Krawczyk 写像という) について、$K(I) \subset I$ を確認する。

(単に $g(I) = I - Rf(I)$ を計算して $g(I) \subset I$ を確認しようしてもうまくいかないことに注意)

条件 2

平均値形式により $\forall x, y \in I$ について $g(x) \in g(y) + g'(I)(x - y)$ なので、区間行列のノルムを $\|g'(I)\| = \max_{X \in g'(I)} \|X\|$ と定義すると

$$\|g(x) - g(y)\| \leq \|g'(I)\| \|x - y\|$$

が成立するので、$\|g'(I)\| = \|E - Rf'(I)\| < 1$ を確認する。

Krawczyk 法を導く (4)

注意 1

「R が逆行列を持つならば $f(x)=0 \Longleftrightarrow g(x)=x$」と書いたが、
$\|E-Rf'(I)\|<1$ ならば自動的に R は逆行列を持つので改めて確認する必要はない。

証明: R が非正則とすると、ある行列 $A \in f'(I)$ に対して RA も非正則となり、よって $RAx=0$ となるベクトル $x \neq 0$ が存在する。すると、$(E-RA)x=x$ となるので $\|E-RA\| \geq 1$ となり、これは矛盾。

注意 2

$c-Rf(c)+(E-Rf'(I))(I-c) \subset \mathrm{int}(I)$ ならば、自動的に $\|E-Rf'(I)\|<1$ が言えるので、改めて確認する必要はない。

略証: ベクトル $x \in \mathbb{R}^n$ に対する scaled maximum norm を、$\|x\|_u = \max\{|x_i|/u_i | 1 \leq i \leq n\}$ で定める。但し scaling vector u は全成分が正。これで誘導される行列ノルムは

$\|A\|_u = \sup\limits_{\|x\|_u=1} \|Ax\|_u = \|\mathrm{mag}(A)u\|_u$ 。このとき、

$c-Rf(c)+(E-Rf'(I))(I-c) \subset \mathrm{int}(I) \Rightarrow \mathrm{mag}(E-Rf'(I))\mathrm{rad}(I) < \mathrm{rad}(I)$ なので、

$u=\mathrm{rad}(I)$ と定めると、$\mathrm{mag}(E-Rf'(I))u < u$ より $\|\mathrm{mag}(E-Rf'(I))u\|_u < 1$ 。これは、

$\|E-Rf'(I)\|_u < 1$ を示している。

Krawczyk 法の例

例題

方程式 $\begin{cases} x_1^2 + x_2^2 - 1 = 0 \\ x_1 - x_2 = 0 \end{cases}$ の解が区間 $I = \begin{pmatrix} [0.6, 0.8] \\ [0.6, 0.8] \end{pmatrix}$ にあることを示す。

計算例

ヤコビ行列 $f'(x) = \begin{pmatrix} 2x_1 & 2x_2 \\ 1 & -1 \end{pmatrix}$ はここでは手計算。$c = \begin{pmatrix} 0.7 \\ 0.7 \end{pmatrix}$、$f'(c) = \begin{pmatrix} 1.4 & 1.4 \\ 1 & -1 \end{pmatrix}$ で、$f'(c)$ の近似 R は、$R = \begin{pmatrix} 0.4 & 0.5 \\ 0.4 & -0.5 \end{pmatrix}$ とした。このとき、$f'(I) = \begin{pmatrix} [1.2, 1.6] & [1.2, 1.6] \\ 1 & -1 \end{pmatrix}$ となり、

$$c - Rf(c) + (E - Rf'(I))(I - c) = \begin{pmatrix} [0.68, 0.736] \\ [0.68, 0.736] \end{pmatrix} \subset \mathrm{int} \begin{pmatrix} [0.6, 0.8] \\ [0.6, 0.8] \end{pmatrix}$$

なので、解の一意存在が保証された。

- 計算過程が計算機向きで、ほぼ機械的に行えるのが嬉しい。
- 区間ヤコビ行列 $f'(I)$ の計算を「内部演算を区間演算に置き換えた自動微分」で行うことによって、ほぼ完全に機械的に実行可能になる。

Krawczyk 法の応用

近似解 c を元にした解の存在保証

$R \simeq f'(c)^{-1}$、$r = 2\|Rf(c)\|$ (Newton 法の修正量の 2 倍) とし、

$$I = c + r \begin{pmatrix} [-1,1] \\ \vdots \\ [-1,1] \end{pmatrix}$$

を候補者集合として Krawczyk 法を使うとよい。

区間ベクトル I 内の全解探索

- 区間ベクトル I での解の存在定理 (Krawczyk 法)
- 区間ベクトル I での解の非存在定理 (例えば $f(I) \not\ni 0$ なら I に解なし)

の 2 つの定理を両方試し、両方共失敗したならば区間を分割する、という作業をどちらかが成立するまで再帰的に繰り返す。

線形方程式に対して同様に縮小写像原理を使う

線形方程式 $Ax = b$ に対して、$f(x) = Ax - b = 0$、$R \simeq A^{-1}$ として、写像 $g : \mathbb{R}^n \to \mathbb{R}^n$ (定義域は $\mathbb{R}^n$ 全域とする) を

$$g(x) = x - R(Ax - b)$$

と定める。このとき、$\|g'(x)\| = \|E - RA\| < 1$ ならば g は縮小写像となる。よって、近似解 x_0 に対して、

$$\|x_0 - x^*\| \le \frac{\|R(Ax_0 - b)\|}{1 - \|E - RA\|} \le \frac{\|R\|\|Ax_0 - b\|}{1 - \|E - RA\|}$$

のような誤差評価が成り立つ (x^* は真解)。これは、線形方程式に対して非常によく知られた誤差評価の式。

自動微分 (1)

自動微分

関数 $f : \mathbb{R}^n \to \mathbb{R}^m$ (の計算手順) と、ある $a \in \mathbb{R}^n$ が与えられたとき、$f(a)$ を計算すると同時に $f'(a)$ の計算を行うアルゴリズム。

- そのような値は、いわゆる数値微分 (例えば
$$\frac{\partial f_j}{\partial x_i}(a) \simeq \frac{f_j(a_1, \cdots, a_i + \varepsilon, \cdots, a_n) - f_j(a_1, \cdots, a_i, \cdots, a_n)}{\varepsilon}$$
とする) でも計算できるが、自動微分は数値微分と違って離散化誤差を含まない (誤差があるとすれば丸め誤差のみ)。また、数値微分は ε が大きすぎれば離散化誤差が増大し、小さすぎれば桁落ちによる誤差が増大するという困難を抱えているが、そのような問題は発生しない。
- Bottom Up 型 (forward mode) と Top Down 型 (backward mode) がある。
- 近年、機械学習のアルゴリズムを支える技術として自動微分が再注目されている。
- 平均値形式や Krawczyk 法に現れる $f'(I)$ の計算に最適。

自動微分 (2)

$f:\mathbb{R}\to\mathbb{R}$ (1 変数関数) と $a\in\mathbb{R}$ に対して、$f(a)$ を計算すると同時に $f'(a)$ を Bottom-Up 型で計算することを考える。

方針

関数 $f(x)$ の計算中に現れる全ての中間変数 $f_1, f_2, \cdots$ について、$(f_i(a), \frac{df_i}{dx}(a))$ の値のペアを保持する。

アルゴリズム

$x=(a,1)$ と初期化。以下、加減乗除と数学関数 h について、

- $(v_1,d_1)\pm(v_2,d_2)\to(v_1\pm v_2,\ d_1\pm d_2)$
- $(v_1,d_1)\times(v_2,d_2)\to(v_1\times v_2,\ v_2d_1+v_1d_2)$
- $(v_1,d_1)/(v_2,d_2)\to(v_1/v_2,\ (1/v_2)d_1+(-v_1/(v_2)^2)d_2)$
- $h(v_1,d_1)\to(h(v_1),\ h'(v_1)d_1)$

の演算規則で計算。ただし式中の定数 c は、$(c,\ 0)$ とみなす。最終的に得られた値を (v_f,d_f) とすると、$v_f=f(a), d_f=f'(a)$ となっている。

自動微分 (3)

例

$f(x) = (x^2+1)(x-3)+2x$ について、$f(2)$ と $f'(2)$ の値を計算してみる。

- $x : (2, 1)$
- $x^2 : (2, 1) \times (2, 1) = (4, 4)$
- $x^2 + 1 : (4, 4) + 1 = (5, 4)$
- $x - 3 : (2, 1) - 3 = (-1, 1)$
- $(x^2 + 1)(x - 3) : (5, 4) \times (-1, 1) = (-5, 1)$
- $2x : 2 \times (2, 1) = (4, 2)$
- $(x^2 + 1)(x - 3) + 2x : (-5, 1) + (4, 2) = (-1, 3)$

実際、$f(2) = -1$ で、$f'(x) = 3x^2 - 6x + 3$ なので $f'(2) = 3$ であることが確かめられる。

自動微分 (4)

$f : \mathbb{R}^n \to \mathbb{R}$ (n 変数関数) と $a \in \mathbb{R}^n$ に対して、$f(a)$ を計算すると同時に $\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)$ を計算することを考える (Bottom Up 型)。

方針

関数 $f(x)$ の計算中に現れる全ての中間変数 $f_1, f_2, \cdots$ について、$(f_i(a), \frac{\partial f_i}{\partial x_1}(a), \cdots, \frac{\partial f_i}{\partial x_n}(a))$ の $n+1$ 個の値を保持する。

アルゴリズム

以下、保持する値を、$v \in \mathbb{R}$、$d \in \mathbb{R}^n$ のペアとして (v, d) のように書く。$x_i = (a_i, e_i)$ $(1 \le i \le n)$ と初期化 (e_i は単位ベクトル)。以下、加減乗除と数学関数 h について、

- $(v_1, d_1) \pm (v_2, d_2) \to (v_1 \pm v_2,\ d_1 \pm d_2)$
- $(v_1, d_1) \times (v_2, d_2) \to (v_1 \times v_2,\ v_2 d_1 + v_1 d_2)$
- $(v_1, d_1)/(v_2, d_2) \to (v_1/v_2,\ (1/v_2)d_1 + (-v_1/(v_2)^2)d_2)$
- $h(v_1, d_1) \to (h(v_1),\ h'(v_1)d_1)$

の演算規則で計算。ただし式中の定数 c は、$(c,\ 0)$ とみなす。最終的に得られた値を (v_f, d_f) とすると、$v_f = f(a)$, $d_f = f'(a)$ となっている。

自動微分 (5)

例

$f(x,y) = (x-y)(x^2+y)+xy$ について、$a=(2,1)$ として、$f(a)$、$\frac{\partial f}{\partial x}(a)$、$\frac{\partial f}{\partial y}(a)$ を計算してみる。

- $x : (2,1,0)$
- $y : (1,0,1)$
- $x-y : (2,1,0)-(1,0,1)=(1,1,-1)$
- $x^2 : (2,1,0)\times(2,1,0)=(4,4,0)$
- $x^2+y : (4,4,0)+(1,0,1)=(5,4,1)$
- $(x-y)(x^2+y) : (1,1,-1)\times(5,4,1)=(5,9,-4)$
- $xy : (2,1,0)\times(1,0,1)=(2,1,2)$
- $(x-y)(x^2+y)+xy : (5,9,-4)+(2,1,2)=(7,10,-2)$

実際、$f(2,1)=7$ で、$\frac{\partial f}{\partial x}=3x^2-2yx+2y$ なので $\frac{\partial f}{\partial x}(2,1)=10$、
$\frac{\partial f}{\partial y}=-2y+2x-x^2$ なので $\frac{\partial f}{\partial y}(2,1)=-2$ であることが確かめられる。

自動微分 (6)

(注意)

- 二項演算の演算規則は、その演算を $g(x,y)$ として
$$g((v_1,d_1),(v_2,d_2)) \to (g(v_1,v_2), \frac{\partial g}{\partial x}(v_1,v_2)d_1 + \frac{\partial g}{\partial y}(v_1,v_2)d_2)$$
のように作られている。
- n 変数関数の自動微分は、「ある i を固定し、x_i のみを変数、残りの変数を定数とみなすことによって f を自動微分で計算」を $1 \le i \le n$ に対して繰り返すことによっても実行できる。この場合、1 変数用のプログラムがそのまま使えるが、$f(a)$ という同じ値の計算を n 回繰り返すことになり計算量が増えてしまう。
- Top Down 型 (backward mode) の自動微分というアルゴリズムもあるが、本稿では説明を省略する。一般に、$f : \mathbb{R}^n \to \mathbb{R}^m$ の自動微分の計算量は、自動微分を行わない場合の f の計算量に比べて、
  - Bottom Up なら、n 倍程度で、m によらない。
  - Top Down なら、m 倍程度で、n によらない。

  ということが知られている。Krawczyk 法で扱われる関数は $f : \mathbb{R}^n \to \mathbb{R}^n$ なのでどちらでも計算量は変わらず、そのためプログラムが簡単で済む Bottom Up 法がよく使われる。

自動微分 + 区間演算

自動微分の内部に使われる数値の型を区間型にすることによって、区間入力に対する微分 (ヤコビ行列) を計算することができるようになる。

例: $f(x) = \begin{cases} x_1^2 + x_2^2 - 1 \\ x_1 - x_2 \end{cases}$ の区間 $I = \begin{pmatrix} [0.6, 0.8] \\ [0.6, 0.8] \end{pmatrix}$ における区間ヤコビ行列を計算する。

- $x_1 : ([0.6, 0.8], 1, 0)$
- $x_2 : ([0.6, 0.8], 0, 1)$
- $x_1^2 : ([0.6, 0.8], 1, 0) \times ([0.6, 0.8], 1, 0) = ([0.36, 0.64], [1.2, 1.6], 0)$
- $x_2^2 : ([0.6, 0.8], 0, 1) \times ([0.6, 0.8], 0, 1) = ([0.36, 0.64], 0, [1.2, 1.6])$
- $x_1^2 + x_2^2 : ([0.36, 0.64], [1.2, 1.6], 0) + ([0.36, 0.64], 0, [1.2, 1.6]) = ([0.72, 1.28], [1.2, 1.6], [1.2, 1.6])$
- $x_1^2 + x_2^2 - 1 : ([0.72, 1.28], [1.2, 1.6], [1.2, 1.6]) - 1 = ([-0.28, 0.28], [1.2, 1.6], [1.2, 1.6])$
- $x_1 - x_2 : ([0.6, 0.8], 1, 0) - ([0.6, 0.8], 0, 1) = ([-0.2, 0.2], 1, -1)$

と進むと、ヤコビ行列 $f'(I) = \begin{pmatrix} [1.2, 1.6] & [1.2, 1.6] \\ 1 & -1 \end{pmatrix}$ が得られる。これを使うと真に機械的な Krawczyk 法となる。