数値計算の品質保証法

関根 晃太

2023 年 4 月 19 日

前書き

皆さんはご飯を安心して食べられますか? もちろん，安心・安全であることが当たり前すぎて，毎日，安全性をほとんど気にせずに美味しく食べていると思います．では，なぜ安心・安全であることが当たり前なのでしょうか? それは，農家や，出荷業者，加工業者，スーパーマッケートなどの多くの人が安心・安全であることを保証してくれているからではないでしょうか? もちろん中には出荷できない商品もあったかもしれません．しかし，品質管理に基づく厳しい基準を設け，基準を満たしていないと判断された商品は出荷しないため，安心・安全ではない商品は皆さんの食卓には届きません．即ち，多くの人が品質を保証をしてくれているから，毎日，安心してご飯を食べることができます．もし，出荷できない物品なんて滅多にないのだから，サボってしまえという考えであれば，人件費などが抑えられて価格は安くなるかもしれません．しかし，本当にサボってしまったらご飯を実際に口にする私たちは安心・安全にご飯を食べられなくなってしまいます．

では，数値計算はどうでしょうか．数値計算においては，この本を手に取って読もうする皆さんは「業者側」の立場です．すなわち，ただコンピュータで計算するだけでなく，得られたの結果が「絶対に間違いがない」と自信をもって提供しなければなりません．しかし，実際には数値計算を行う上では多くの誤差が含まれます．

例えば，丸め誤差です．現在のコンピュータは実数を IEEE 754 標準で定められた浮動小数点数に近似して保持します．そのため，10 進数の 1.1 のような簡単な値でさえ，2 進数では循環小数になってしまうため，コンピュータ内では近似して保持されます．また，浮動小数点数同士の演算結果も浮動小数点数で表せない場合，切り捨てや切り上げによって近似することで浮動小数点数にして保持します．この誤差を「丸め誤差」といいます．そのため，浮動小数点数の演算は非常に高速である反面，演算を経て得られた結果は丸め誤差を含む近似値となってしまいます．

他にも数値計算では，有限次元の非線形方程式を解く方法として Newton 法といった反復解法があります．Newton 法では，正確な実数計算を無限回繰り返すことで真の解に到達することが保証されます．しかし，コンピュータ上では，実数の正確な計算ができないだけに留まらず，Newton 法の反復を無限回繰り返すことも不可能です．そのため，数値計算アルゴリズムに含まれる誤差にも注意しなければなりません．このような無限回繰り返す計算を途中でやめてしまう誤差は「打ち切り誤差」と呼ばれます．

さらに，偏微分方程式の解は求めたい読者もいると思います．しかし偏微分方程式の解は一般的には無限次元になります．ここでもコンピュータが苦手とする「無限」が出てくるために，偏微分方程式の解を数値計算で求める際には，無限次元の問題を有限要素法などを用いて有限次元の非線形方程式に近似して求めます．この際の誤差を離散化誤差と呼ばれます．

すなわち，非線形偏微分方程式の解を数値計算で求めようとすると，丸め誤差，打切り誤差，離散化誤差がすべて含まれてしまいます．もちろん，近似でも真の解の性質を捉えている場合も多くあると思います．しかし，問題なのは真の解の性質が捉えられているか，それとも，まるっきり違うものになってしまっているのか判断がつかない点です．このままでは，得られた結果を「絶対に間違いがない正しい結果だ!」と自信をもって提供することができません．そのために厳しい品質基準を設け，基準を満たした結果を検証済み (検品済み) として安心・安全に提供できる状況

にしたいのです．そして，厳しい品質基準を設けることが，この本の目標になります．

では，数値計算において品質基準を設けるにはどうすれば良いでしょうか? 食品の場合はある意味で「正解」が決められており，正解から具体的な数値で外れたものは出荷しないと判断できます．例えば,「農薬の残留量が 0.01ppm 以上だったら出荷は不可」という厳しい基準を設けられます．しかし，数値計算の場合は正解がそもそもわかりません．(もっというと正解が存在するかどうかすらわからないです．) 正解がわからなければ，得られた結果がどれくらい正しいか品質を保証をすることも一見すると不可能に感じます．その一見すると不可能な問題を可能にしようとすることが本書の目的です．そのため，数値計算の品質保証はまだまだ発展途上であり,「こんな問題もすぐに解けないのか!」とか,「面倒だ!」と言って匙を投げてしまいたくなる時もあるかもしれません．その際には，匙を投げてしまう前に「こんな問題はこうやって品質保証できるぞ!」,「こうやればもっと簡単だぞ!」といった改善法を非才な著者に投げつけて下さい．

本書の最低限の知識は章によって異なりますが，共通して線形代数，微積分，数値計算，プログラミングの知識は前提としております．第 5 章以降では，関数解析を展開していますが，ほとんどの定理に証明をつけました．線形代数，微積分，数値計算，プログラミングの知識しか持っていない読者にとっては，第 5 章以降の内容は一見すると抽象な内容で理解しにくいかもしれません．実際に著者自身も修士 2 年生までは電気電子工学科に所属しており，数学や数値計算とは大きく異なった勉強をしていたため，線形代数，微積分，数値計算，プログラミングの知識すら怪しい状況でした．しかし，博士課程以降なんとか独学で勉強をし，今では少しくらいは感覚がつかめたかな?っと思えるようになってきました．本書籍でも第 5 章がわかりにくい場合，まずは定理 5.4.2 の使い方のみ理解していただけたら幸いです．

目 次

第 1 章　浮動小数点数と区間演算 **7**
1.1　5 次の代数方程式の解の品質保証を手計算でやってみよう 7
1.2　コンピュータで扱う小数: 浮動小数点数とは? 10
1.3　1 回の浮動小数点数同士の演算における品質保証 12
1.4　上端下端型の区間演算 15
1.5　浮動小数点数で成分を持つ点行列同士の行列積 19
1.6　区間演算の落とし穴 20
1.7　中心半径型の区間と区間演算 22

第 2 章　区間ベクトル/行列の演算に対する品質保証法 **27**
2.1　記号の準備 27
2.2　区間行列積の計算方法 28

第 3 章　連立一次方程式の解の品質保証法 **33**
3.1　標準的な品質保証法 33
3.2　成分毎評価 34
3.3　(発展) さらに評価を極めるには? 38

第 4 章　固有値問題の固有値に対する品質保証法 **43**
4.1　全固有値に対する一般的な品質保証法 43
4.2　(発展) さらに評価を極めるには? 46

第 5 章　方程式の品質保証のための基本定理 **51**
5.1　Banach 空間 51
5.1.1　Banach 空間の定義と性質 51
5.1.2　ノルム空間の位相構造と Banach 空間の閉部分空間 55
5.2　作用素の基礎 58
5.2.1　作用素とは?写像との違いに注意しながら... 58
5.2.2　線形作用素 61
5.2.3　有界な線形作用素 63
5.2.4　定義域が X の全体となる有界な線形作用素全体の集合 $\mathcal{B}(X,Y)$ 64
5.3　非線形解析の基礎 72
5.3.1　Fréchet 微分 72
5.3.2　Bochner 積分 75
5.3.3　閉区間上の Banach 空間値関数に対する微分積分学の基本定理 77
5.4　方程式の解の品質保証の準備 81
5.4.1　Banach の不動点定理 81
5.4.2　方程式の解の品質保証のための基本定理 84

第 6 章　射影を用いた無限次元線形問題の解法 **89**
6.1　直和と射影 〜計算できる部分とできない部分の分離〜 89
6.2　射影を用いた Banach 空間上のガウスの消去法 92
6.3　準直交射影と $\|L^{-1}\|_{\mathcal{B}(X)}$ の評価方法 97

第 7 章　無限次元非線形問題の解法のエッセンス〜A が全単射の場合〜 **101**
7.1　$F'[\hat{u}]$ の全単射性の確認方法 101
7.2　系 5.4.2 の定数 η の計算方法 103
7.3　系 5.4.2 の定数 K の計算方法 104

第1章　浮動小数点数と区間演算

1.1　5次の代数方程式の解の品質保証を手計算でやってみよう

まず，数値計算を用いた方程式の品質保証法の感覚をつかんでもらうために，5次の代数方程式

$$\text{Find } x \in \mathbb{R} \text{ s.t. } -5x^5 + 5x^4 + 5x^3 + 6x^2 + 6x + 5 = 0 \tag{1.1}$$

の解を考えてみましょう．ここで，$\mathbb{R}$ は実数全体の集合を意味し，s.t. は such that の略語です．すなわち,「方程式 (1.1) を満たす解を実数全体から探せ」という問題です．5次の代数方程式の解は因数分解によって解ける場合もあるし，そもそも実数に解がない場合もあります．そして，数値計算を用いて方程式 (1.1) の近似解を求める手法は多く知られています.「答えがすぐにわからない」,「そもそも答えが存在するかもわからない」,「数値計算法で近似的に求められる」ので，はじめて品質保証を行うには持ってこいです.

先に答えを見せてしまうと，この方程式 (1.1) に対して私の手元のコンピュータで実際に品質保証まで含めて解いてみると結果は区間

$$[2.006595606321886, 2.006595606321888]$$

内に真の解が必ず一意に存在することを保証できました．すなわち，真の解の上位15桁は「2.00659560632188」であることがわかります．この結果はコンピュータを使って計算しましたが，本章では流れをつかんでもらうために手計算で計算していきたいと思います.

では，方程式 (1.1) に対して，手計算で数値計算の品質保証の流れをつかんでみましょう．まず，今回の例で使う重要な定理を紹介します．定理を紹介したのちに，一歩ずつ定理の使い方を解説しますので，定理だけで理解する必要はありません.

定理 1.1.1. $\hat{x} \in \mathbb{R}$ を与えられているとする ($\hat{x}$ はコンピュータで求めた近似解をイメージしており，なんでも構いません!)．$f : \mathbb{R} \to \mathbb{R}$ を与えられた関数とする．(問題となる方程式を $f(x) = 0$ と書き直したときの関数 f です!)． 関数 f は $\hat{x}$ で微分可能とし $f'[\hat{x}]$ と表記する．$f'[\hat{x}]$ は逆を持つと仮定する．η を

$$|f'[\hat{x}]^{-1} f(\hat{x})| \leq \eta$$

を満たす定数とする．また，関数 f は $\hat{x} + v,\ \forall v \in [-2\eta, 2\eta]$ で微分可能とする.

定数 K を不等式

$$|f'[\hat{x}]^{-1} \left(f'[\hat{x}] - f'[\hat{x} + v] \right)| \leq K,\ v \in [-2\eta, 2\eta]$$

を満たす定数とする．もし $2K \leq 1$ ならば，$f(x) = 0$ を満たす解を x^* とすると

$$|x^* - \hat{x}| \leq \frac{\eta}{1 - K} =: \rho$$

となり，$[\hat{x} - \rho, \hat{x} + \rho]$ 内に x^* は存在する．その上，$[\hat{x} - 2\eta, \hat{x} + 2\eta]$ 内で一意である.

第 5 章の目標が，この定理 1.1.1 をものすごく一般化した定理 5.4.2 の証明ですので，ここでは証明をつけずに利用だけしたい思います．数値計算の品質保証では上のような定理 1.1.1 の十分条件を満たすか確認をしていきます．これからひとつづつ設定/計算/十分条件の確認を行っていきますので，定理 1.1.1 のどこにあたる部分か，迷わないにならないように定理を見比べながら読み進めて下さい．

まず，品質を保証したい近似解 $\hat{x}$ を決めます．今回は手計算で話を進めるために，

$$\text{近似解 } \hat{x} = 2$$

とします．どんな近似解が来ても品質保証を行い判断をしなければならないために，最初の近似解はなんでも良いです．もちろん，この近似解が良ければ品質が良いものと判断されます．

次に，方程式 (1.1) を $f(x) = 0$ の形に書き直します．すなわち，

$$f(x) = -5x^5 + 5x^4 + 5x^3 + 6x^2 + 6x + 5$$

とすれば良いですね．また，f の微分も必要になります．f の y における微分は

$$f'[y] = -25y^4 + 20y^3 + 15y^2 + 12y + 6$$

となります．もちろん $y \in \mathbb{R}$ と取れるので，f は実数全体で微分可能です．

続いて，定数 η を計算する準備として，近似解 $\hat{x}$ における残差 $f(\hat{x})$ と微分値 $f'[\hat{x}]$ を計算します:

残差 $f(\hat{x})$:

$$f(\hat{x}) = -5 \times 2^5 + 5 \times 2^4 + 5 \times 2^3 + 6 \times 2^2 + 6 \times 2 + 5 = 1$$

近似解 $\hat{x}$ における f の微分値 $f'[\hat{x}]$

$$f'[\hat{x}] = -25 \times 2^4 + 20 \times 2^3 + 15 \times 2^2 + 12 \times 2 + 6 = -150$$

次に，近似解 $\hat{x}$ における f の微分値が逆を持つことを確認します．今回は $f'[\hat{x}] \neq 0$ であるため $f'[\hat{x}]$ は逆を持ちます．

次に Newton 法の修正量 $f'[\hat{x}]^{-1} f(\hat{x})$ の絶対値を計算し，η を計算します．

$$|f'[\hat{x}]^{-1} f(\hat{x})| = \frac{1}{150} =: \eta$$

次に，定数 K の計算です．関数 g を

$$\begin{aligned} g(y) =& 1 - \left(-\frac{1}{150}\right)\left(-25y^4 + 20y^3 + 15y^2 + 12y + 6\right),\ \forall y \in \left[2 - \frac{1}{75}, 2 + \frac{1}{75}\right] \\ =& -\frac{1}{150}(y-2)\left(25y^3 + 30y^2 + 45y + 78\right),\ \forall y \in \left[2 - \frac{1}{75}, 2 + \frac{1}{75}\right] \end{aligned}$$

とすると，定数 K は

$$|g(y)| \leq K,\ y \in [\hat{x} - 2\eta, \hat{x} + 2\eta]$$

となるため，関数 g の定義域を $[\hat{x} - 2\eta, \hat{x} + 2\eta]$ としたときの最大値と最小値を求めれば良いです．一応，関数の最大値と最小値の求め方は高校生の学習範囲ですが，値が綺麗ではないので最大値と最小値を求めるのは大変です．そのため結果だけ書くと

$$-\frac{4170083}{94921875} \leq g(y) \leq \frac{4065457}{94921875},\ \forall y \in [\hat{x} - 2\eta, \hat{x} + 2\eta]$$

となります．よって，定数 K は

$$K = \frac{4170083}{94921875}$$

とすれば良いです．

定数 K が求まったら十分条件

$$2K \leq 1$$

の確認です．もし，ここで 1 以下でなかった場合は品質保証が失敗になり，そもそも近似解の近くに解が存在するかどうかわかりません．今回の問題では，

$$2K = \frac{8340166}{94921875} \leq 1$$

となるので，近似解の近くに解が存在することが証明できます．これで定理 1.1.1 の十分条件はクリアです! その上で，真の解を x^* と書くと，真の解と近似解の誤差は

$$|x^* - \hat{x}| \leq \frac{\eta}{1-K}$$

となります．今回の場合は

$$|x^* - \hat{x}| \leq \frac{\frac{1}{150}}{1 - \frac{4170083}{94921875}} = \frac{1265625}{181503584} \leq 0.007$$

となります．また，区間

$$[1.993, 2.007]$$

内に真の解が存在していることも証明されました．これで，定理 1.1.1 の使い方の説明はおしまいです．定理を使っている最中に，真の解がわかっている必要がなく，その上，最後まで行っても真の解はわかりません．しかし，最初に設定した近似解 $\hat{x} = 2$ の近く (具体的には絶対誤差 $|x^* - \hat{x}| \leq 0.007$) に解が存在することがわかりました．

ついでに，この誤差が大きいか，小さいかわかりやすいように誤差率もどきの上界も計算してみましょう．誤差率の定義は

$$\text{誤差率} = \left| \frac{\text{真値} - \text{近似値}}{\text{真値}} \right| \times 100[\%]$$

です．しかし，真の解 x^* がわかっていないので分子の真値はわかりません．そこで，数値計算の品質保証では，近似値の近くに真値があることから誤差率もどき

$$\text{誤差率もどき} = \left| \frac{\text{真値} - \text{近似値}}{\text{近似値}} \right| \times 100[\%]$$

をつかいます．先ほどの例では誤差率もどき上界は

$$\text{誤差率もどき} = \left| \frac{x^* - \hat{x}}{\hat{x}} \right| \times 100[\%] \leq \frac{0.007}{2} \times 100 = 0.35[\%]$$

となります．誤差率もどきの上界 0.35%が大きいか小さいかは，用途によって変わります．もちろん要求する精度がもっと必要な場合は近似解を 2 としてはダメであり，もっと高品質な近似解を設定しなければなりません．

本節では定理 1.1.1 の

- $f'[\hat{x}]$ が逆を持つこと
- 定数 η の計算
- 定数 K の計算
- 十分条件 $2K \leq 1$ のチェック
- 絶対誤差の上界 ρ

は手計算し，確認していきました．ポイントは，近似解にどんな誤差が入っていようが関係なく，与えられた何でもいい近似解に対して検証できる点です．しかし，もっと複雑な問題や大規模な問題に対応するためには，手計算では限界があるためにコンピュータの利用は必須になります．そのために，定数 η や K をコンピュータで正確に行わなければなりません．そこで，本章では残りの節ではコピュータを用いてある定数 η や K を正確に求める方法として，ある種の集合演算である「浮動小数点数における区間演算」を導入します．区間演算によって，浮動小数点数を扱う際に発生する誤差を考慮しながら計算することが可能になります．それでは，浮動小数点数の正確な定義からはじめましょう．

1.2 コンピュータで扱う小数: 浮動小数点数とは?

コンピュータ上の計算と聞くと万能そうに聞こえますが，多くの場合，取り扱う数値を近似してしまいます．例えば，円周率 π をコンピュータで扱うにはどうすればよいか?と考えてみます．まず，記号として π のように保持することができます．しかし，π を記号としてそのまま処理できる演算は CPU 上には実装されていないため，数値計算とは別分野の数式処理としてソフトウェア的にしか計算できません (数式処理ソフトの例として Mathematica などがあります．)．それでは，π を小数で保持しようとしても，3.141592... と無限の桁が続くため，コンピュータのメモリーも無限に必要になります．π を使って演算したいのに，π をそのまま保持することはできないので，現状では,「有限桁に打ち切った小数」としてコンピュータで扱います．さらには，コンピュータは内部では 0 と 1 の 2 進数で表される決められた一定の長さのビット列でデータを保持します．そのため,「決められた一定の長さのビット列で表現する有限桁に打ち切った小数」の取り扱い方に関する議論が必要です．色々な規則が開発されているが，今，現在，多くの PC に実装されている CPU には IEEE 754 標準という規格に基づいています．IEEE 754 標準では，浮動小数点数に関する技術規格が制定されています．IEEE 754 標準の詳細を述べる前に，品質を扱うからこそ知っておいてほしい重要なことがあります．上の例で扱った π は「無理数なんだから，コンピュータで実際に持てなくても仕方ない」って思っている読者に是非聞いてほしいことです．例えば，IEEE 754 標準では十進数で 1.1 は厳密には表現できません．すなわち，何かしらのプログラミング言語で，浮動小数点数として 1.1 を保持しようとすると，1.1 を近似して保持をします．そのために，計算する以前にすでに誤差が発生してしまいます．

IEEE 754 標準には，single と double というフォーマットがあります．single は 32 bit の長さで近似して保持し，double は 64 bit の長さで近似して保持します．詳細は以下の規格書をみて下さい:

`http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4610935`

本章では，IEEE 754 規格の簡単な紹介と，品質保証に必要となる丸め変更について説明します．前述したとおり IEEE 754 規格では *single* と *double* という 2 進数の浮動小数点数のフォーマット

$$(-1)^s \times 2^e \times m$$

が規格化されています．ここで，s, e, m はそれぞれ

a) 符号ビット s：1 bit (0 あるいは 1)

b) 指数部 e：$e_{\min} \leq e \leq e_{\max}$ となる整数

c) 仮数部 m：正規化された形式 $d_0.d_1d_2\cdots d_{p-1}$ で表現される．ただし，d_i は 0 or 1.

です．$e_{\max}, e_{\min}, p$ は *single* と *double* で変化し，表 1.1 のように定められています (ただし，$e_{\min} := 1 - e_{\max}$ です)．

表 1.1: 指数部と仮数部

Parameter	single	double
p	24 bit	53 bit
$e_{\max}$	127	1023

また，図 1.1 に *Single* の場合の表現を図示します．ポイントは，正規化数と呼ばれる状態では常に $d_0 = 1$ としておき，d_1 から記憶させています．これにより 1 bit 分だけお得に記憶できるため,「ケチビット」と呼ばれます．

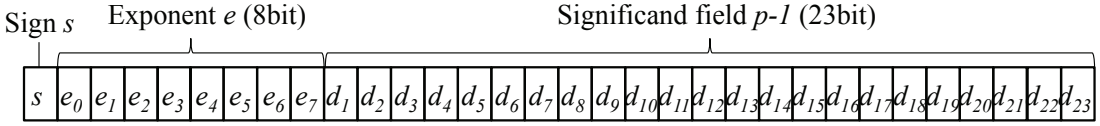

図 1.1: *Single* の場合の表現方法

問題

10 進数 7.25 を浮動小数点数に変換せよ．ただし，p は (d_0 も含め)6 bit とし，e は 3 bit とする．

解答

$$\begin{aligned} 7.25 &= (111.01)_2 \\ &= (-1)^0 \times 2^2 \times (1.1101)_2 \end{aligned}$$

ここで，$(\cdot)_2$ は 2 進数を意味する．そのため，解答は

$$0 \;\; 010 \;\; 11010$$

となる．

仮数部 $p = 3$ bit とすると，上記の例題にある科数部のビット 11010 は表現できないことに気が付くと思います．

$\mathbb{F}$ を IEEE 754 規格で定義された浮動小数点数の集合とします．もちろん，$\mathbb{F} \subset \mathbb{R}$ であることに注意してください．そのとき，IEEE 754 標準における切り捨てや切り上げなどの丸めは，$+,\ -,\ *,\ /,\ \sqrt{\cdot}$ の5つの演算結果に限り，丸めが定義されています．ここで注意点として，入力や出力，初等関数 sin などの丸めに関しては定義されていないので注意してください．実際に，$\circ \in \{+, -, \times, /\}$ としたとき，以下の例えば3つの丸めモードがあります (ちなみに，2008 年に制定された IEEE 754 規格では全部で 5 つの丸めモードがあります．):

a) 最近点丸め: $\tilde{\circ} : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ とし，最も近い浮動小数点数に丸める
b) 上向き丸め: $\overline{\circ} : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ とし，$\inf\{x \in \mathbb{F} \mid x \geq a \circ b\}$
c) 下向き丸め: $\underline{\circ} : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ とし，$\sup\{x \in \mathbb{F} \mid x \leq a \circ b\}$

例えば，C99 準拠の C 言語コンパイラでは fenv.h を使用することで，丸めモードの変更を行う fesetround 関数が利用できます．ただし，コンパイラの最適化により計算順序が変更されたり，IEEE 754 標準に準拠しない演算を行われ，意図せずに精度保証付き数値計算として成立しなくなる場合があります．そのために最適化の抑制として，C 言語で定められている volatile 属性を変数に付加することで最適化を抑制したり，浮動小数点数に関する最適化オプション (-fp-model など) がコンパイラオプションとしてある場合は最も厳格に準拠 (strict) するような設定する必要があります．

問題
次のプログラムについて問に答えよ:
1) 誤差が発生する行を答えよ．
2) 丸めモードの変更により動作が保証されている行を答えよ．

```
1   #include <stdio.h>
2   #include <math.h>
3
4   int main(void){
5     double a = 0.1;
6     double b = 1.1;
7     double c;
8
9     c = a + b;
10
11    printf("%lf\n", c);
12  }
```

解答
(1) 5, 6, 9, 11 行目
(2) 9 行目 (四則演算と平方根のみ丸めモードの変更による動作が保証されている)

1.3 1回の浮動小数点数同士の演算における品質保証

ここでは 1 回の浮動小数点数同士の演算における結果の品質保証法を紹介します．具体的には，$a, b \in \mathbb{F}$，すなわち，a も b も誤差なく浮動小数点数で記述されているとしたときの四則演算 ($a+b$, $a-b$, $a*b$, a/b) の品質保証について取り上げます．

品質の保証を行う際の信念は,「1bit たりともさぼらずに必ず数学的に正しい結果」を返すことです．しかしながら，$a+b$ という結果は多くの場合，実数 $\mathbb{R}$ になってしまいコンピュータでは表現できません．そこで，コンピュータで表現できる $a+b$ で表せる数学的に正しい結果は，2 つの浮動小数点数 $c_l, c_u \in \mathbb{F}$ を用いて

$$c_l \leq a+b \leq c_u$$

のように結果を挟み込むことです．別の言い方をすれば,「区間 $[c_l, c_u]$ の中に $a+b$ の結果が含まれている」あるいは「$a+b \in [c_l, c_u]$」といえます．もちろん c_l と c_u の幅が大きければ，使い物にならない結果かもしれません．

では，次に c_l と c_u をどのように作ればよいか，考えてみましょう．まず，演算子 $+$ を用いて丸めも含めた演算をもう一度おさらいすると

$$\begin{aligned}
a+b \in \mathbb{R} &: \text{数学における足し算．結果は浮動小数点数とは限らない} \\
a \tilde{+} b \in \mathbb{F} &: a+b \text{に最も近い浮動小数点数．} \\
a \overline{+} b \in \mathbb{F} &: a+b \text{よりも大きい浮動小数点数のうち最も小さい数．} \\
a \underline{+} b \in \mathbb{F} &: a+b \text{よりも小さい浮動小数点数のうち最も大きい数．}
\end{aligned}$$

でした．IEEE 754 標準では，$+$ 以外にも $-, *, /$ と $\sqrt{\cdot}$ のみ丸めモードの変更による動作の保証がされておりましたね．よって，$c_l = a \underline{+} b$, $c_u = a \overline{+} b$ とすれば,

$$a+b \in [a \underline{+} b, a \overline{+} b] = \{c \in \mathbb{R} \mid a \underline{+} b \leq c \leq a \overline{+} b\}$$

となり，$a+b$ の真の結果はわからないですが，コンピュータで数学的に正しい結果 $[a \underline{+} b, a \overline{+} b]$ を得ることができます．

同様に他の四則演算についても

$$\begin{aligned}
a-b &\in [a \underline{-} b, a \overline{-} b] \\
a \cdot b &\in [a \underline{\cdot} b, a \overline{\cdot} b] \\
a/b &\in [a \underline{/} b, a \overline{/} b]
\end{aligned}$$

のように 2 つの浮動小数点数の丸めモードを用いて真の結果を含む区間を得ることができます．

問題

プログラム

```
#include <stdio.h>
#include <math.h>

int main(void){
  double a = 0.1;
  double b = 1.1;
  double c;

  c = a + b;

  printf("%lf\n", c);
}
```

の a と b を品質保証を行う計算に変えよ (c については次節で取り扱う). ただし，丸めモードは以下をような fenv.h を使用すること:

```
#include <fenv.h>
#include <math.h>

fesetround(FE_UPWARD);        //  これ以降の四則演算を上向き丸めモードへの変更
fesetround(FE_DOWNWARD);      // これ以降の四則演算を下向き丸めモードへの変更
fesetround(FE_TONEAREST);     // これ以降の四則演算を最近点丸めモードへの変更
```

解答例
0.1 と書いてしまうと誤差が発生するために，1.0/10.0 のように倍精度浮動小数点数で誤差なく記述できる値の計算に変形してから，丸めモードを用いて上端と下端を計算すればよい (1.1 についても同様である). 具体的には以下のようなプログラムになる.

```
#include <stdio.h>
#include <fenv.h>
#include <math.h>

int main(void){

  fesetround(FE_DOWNWARD);
    double al = 1.0/10.0;     // 下向き丸めで 1/10 を計算
  fesetround(FE_UPWARD);
    double au = 1.0/10.0;     // 上向き丸めで 1/10 を計算

  fesetround(FE_DOWNWARD);
    double bl = 11.0/10.0;     // 下向き丸めで 11/10 を計算
  fesetround(FE_UPWARD);
    double bu = 11.0/10.0;     // 上向き丸めで 11/10 を計算

  double c;
  c = a + b;  // ここはどうする?っていうのが次の話!

  printf("%lf\n", c);
}
```

上記の解答を見てみると，それぞれ $0.1 \in [a_l, a_u]$, $1.1 \in [b_l, b_u]$ となる区間が得られます. ただし，よく見ると同じ計算 (例えば 1.0/10.0) が出てきます. もちろん，丸めモードを変えているので，結果はかわります. しかし，プログラムにはコンパイラの最適化オプションによっては，賢いコンパイラによって 1.0/10.0 は同じ計算だとみなされてしまい

```
  fesetround(FE_DOWNWARD);
    double al = 1.0/10.0;     // 下向き丸めで 1/10 を計算
  fesetround(FE_UPWARD);
```

```
  double au = al;

fesetround(FE_DOWNWARD);
  double bl = 11.0/10.0;     // 下向き丸めで 11/10 を計算
fesetround(FE_UPWARD);
  double bu = bl;
```

のように意図とは違うプログラムに置き換えてしまうかもしれませんので，注意してください．

1.4 上端下端型の区間演算

前節の問題の解答例では,$a = 0.1 \in [a_l, a_u]$, $b = 1.1 \in [b_l, b_u]$ のように端点が浮動小数点数となる区間を用いて，0.1 や 1.1 をコンピュータで数学的に正しく保持しました．しかし，その先のプログラムでは $c = a + b$ を計算しています．このままでは，区間同士の $+$ の意味や規則がわからないため，コンパイルエラーになってしまいます．本節では，区間同士の四則演算を定義することで，演算結果の品質を保証することが目標です．

まず，丸め誤差が入らない状況における区間同士の演算を考えています．$a \in [a_l, a_u]$, $b \in [b_l, b_u]$ としたときの $a + b$ の結果は集合

$$\{a + b \in \mathbb{R} \mid \forall a \in [a_l, a_u],\ \forall b \in [b_l, b_u]\}$$

で表されます．もちろん足し算以外の四則演算 $(-, \cdot, /)$ でも同じです．例えば，$a = 0.1 \in [a_l, a_u]$, $b = 1.1 \in [b_l, b_u]$ のとき，$a + b = 1.2$ は

$$1.2 \in \{a + b \in \mathbb{R} \mid \forall a \in [a_l, a_u],\ \forall b \in [b_l, b_u]\}$$

のように集合内に元の足し算結果が必ず含まれます．すなわち，この右辺の集合を区間同士の演算結果として考えれば，数学的に正しい結果を得ることが可能です．すなわち，区間同士の四則演算は

$$\begin{aligned}
[a_l, a_u] + [b_l, b_u] &= \{a + b \in \mathbb{R} \mid \forall a \in [a_l, a_u],\ \forall b \in [b_l, b_u]\} \\
[a_l, a_u] - [b_l, b_u] &= \{a - b \in \mathbb{R} \mid \forall a \in [a_l, a_u],\ \forall b \in [b_l, b_u]\} \\
[a_l, a_u] \cdot [b_l, b_u] &= \{a \cdot b \in \mathbb{R} \mid \forall a \in [a_l, a_u],\ \forall b \in [b_l, b_u]\} \\
[a_l, a_u] / [b_l, b_u] &= \{a / b \in \mathbb{R} \mid \forall a \in [a_l, a_u],\ \forall b \in [b_l, b_u]\}
\end{aligned}$$

のように定義します．その上で，右辺の集合は閉区間になり，端点 $a_l, a_u, b_l, b_u \in \mathbb{R}$ のみで書き表せることが知られています:

- 加算：$\{a + b \mid a \in [a_l, a_u], b \in [b_l, b_u]\} = [a_l + b_l,\ a_u + b_u]$
- 減算：$\{a - b \mid a \in [a_l, a_u], b \in [b_l, b_u]\} = [a_l - b_u,\ a_u - b_l]$
- 乗算：$\{a \cdot b \mid a \in [a_l, a_u], b \in [b_l, b_u]\} =$
  $[\min\{a_l \cdot b_l,\ a_u \cdot b_l,\ a_l \cdot b_u,\ a_u \cdot b_u\},\ \max\{a_l \cdot b_l,\ a_u \cdot b_l,\ a_l \cdot b_u,\ a_u \cdot b_u\}]$
- 除算：$\{a/b \mid a \in [a_l, a_u], b \in [b_l, b_u]\} =$
  $[\min\{a_l/b_l,\ a_u/b_l,\ a_l/b_u,\ a_u/b_u\},\ \max\{a_l/b_l,\ a_u/b_l,\ a_l/b_u,\ a_u/b_u\}]$
  ただし $0 \notin [b_l,\ b_u]$

この演算で計算した区間の中には，真の答えが必ず含まれています．

しかし，この区間の端点にはも浮動小数点数で表さなければコンピュータで取り扱うことができません．注意して頂きたい点は，端点 a_l, a_u, b_l, b_u は，既に浮動小数点数として計算機で保持できていることが前提です．演算前の端点 a_l, a_u, b_l, b_u が浮動小数点数でない場合，演算以前に品質保証の取り扱いにミスがあるので注意して下さい．ここで気にしなければならない誤差は，演算結果を構成する端点 $a_l, a_u, b_l, b_u \in \mathbb{F}$ 同士の演算 (例えば，$a_l + b_l$ など) の誤差です．そのために，例えば，$a_l, a_u, b_l, b_u \in \mathbb{F}$ における足し算では

$$[a_l, a_u] + [b_l, b_u] = [a_l + b_l,\ a_u + b_u] \subset [a_l \mathbin{\underline{+}} b_l,\ a_u \mathbin{\overline{+}} b_u]$$

のように，丸めモードの変更を変更することで，真の像を包含するような区間を作成します．そのために，丸めモードまで含めた演算規則として，以下のような機械区間演算を定義します:

定義 1.4.1 (機械区間演算)**.** $a_l, a_u, b_l, b_u \in \mathbb{F}$ としたとき，以下として機械区間演算を定義する:

- 加算：$[a_l,\ a_u] + [b_l,\ b_u] = [a_l \mathbin{\underline{+}} b_l,\ a_u \mathbin{\overline{+}} b_u]$
- 減算：$[a_l,\ a_u] - [b_l,\ b_u] = [a_l \mathbin{\underline{-}} b_u,\ a_u \mathbin{\overline{-}} b_l]$
- 乗算：$[a_l,\ a_u] \times [b_l,\ b_u] =$
  $[\min\{a_l \mathbin{\underline{\cdot}} b_l, a_u \mathbin{\underline{\cdot}} b_l, a_l \mathbin{\underline{\cdot}} b_u,\ a_u \mathbin{\underline{\cdot}} b_u\}, \max\{a_l \mathbin{\overline{\cdot}} b_l, a_u \mathbin{\overline{\cdot}} b_l, a_l \mathbin{\overline{\cdot}} b_u, a_u \mathbin{\overline{\cdot}} b_u\}$
- 除算：$[a_l,\ a_u]/[b_l,\ b_u] =$
  $[\min\{a_l \mathbin{\underline{/}} b_l, a_u \mathbin{\underline{/}} b_l, a_l \mathbin{\underline{/}} b_u,\ a_u \mathbin{\underline{/}} b_u\}, \max\{a_l \mathbin{\overline{/}} b_l, a_u \mathbin{\overline{/}} b_l, a_l \mathbin{\overline{/}} b_u, a_u \mathbin{\overline{/}} b_u\}$
  ただし $0 \notin [b_l,\ b_u]$

問題
プログラム

```
#include <stdio.h>
#include <math.h>

int main(void){
  double a = 0.1;
  double b = 1.1;
  double c;

  c = a + b;

  printf("%lf\n", c);
}
```

の a と b 及び，$c = a + b$ の品質保証を行う計算に変えよ．ただし，丸めモードは以下をようなfenv.h を使用すること:

```
#include <fenv.h>

fesetround(FE_UPWARD);        //  これ以降の四則演算を上向き丸めモードへの変更
fesetround(FE_DOWNWARD);      // これ以降の四則演算を下向き丸めモードへの変更
fesetround(FE_TONEAREST);     // これ以降の四則演算を最近点丸めモードへの変更
```

解答例
a と b の保証については前節に取り上げてあることに注意し解答を示す:

```
#include <stdio.h>
#include <fenv.h>
#include <math.h>

int main(void){

  fesetround(FE_DOWNWARD);
    double al = 1.0/10.0;     // 下向き丸めで 1/10 を計算
  fesetround(FE_UPWARD);
    double au = 1.0/10.0;     // 上向き丸めで 1/10 を計算

  fesetround(FE_DOWNWARD);
    double bl = 11.0/10.0;     // 下向き丸めで 11/10 を計算
  fesetround(FE_UPWARD);
    double bu = 11.0/10.0;     // 上向き丸めで 11/10 を計算

  fesetround(FE_TONEAREST);  // 最近点丸めにもどす

  double cl, cu;
  fesetround(FE_DOWNWARD);
    cl = al + bl;
  fesetround(FE_UPWARD);
    cu = au + bu;

  fesetround(FE_TONEAREST);  // 最近点丸めにもどす

  printf("%lf\n", cl);
  printf("%lf\n", cu);
}
```

以下の問題は，ベクトル，行列を扱う際に良く使う演算です．場合によっては，計算途中の結果を区間として保持する必要がなく，一気に計算可能な例があります．どのような演算が一気に計算可能な例になるか覚えておきましょう．

問題
$a, b, c, c_l, c_u \in \mathbb{F}$ とする．そのとき，以下の結果を包含する両端点に浮動小数点数を持つ区間を，丸めモード変更付きの四則演算を用いて作成せよ:

$$(1)\ a \cdot b - c$$
$$(2)\ c - a \cdot b$$
$$(3)\ a \cdot b + [c_l, c_u]$$

解答例
問 (1):

$$\begin{aligned} a \cdot b - c \subset & [a \underline{.} b,\ a \overline{.} b] - c \\ = & [a \underline{.} b,\ a \overline{.} b] - [c, c] \\ \subset & [a \underline{.} b \underline{-} c,\ a \overline{.} b \overline{-} c] \end{aligned}$$

注意点としては，$[a \underline{.} b,\ a \overline{.} b] - c$ は $[a \underline{.} b,\ a \overline{.} b] - [c, c]$ のように c を点区間として計算すればよい．上記の計算により，

```
fesetround(FE_DOWNWARD);
  res_lower = a*b - c;
fesetround(FE_UPWARD);
  res_upper = a*b - c;
```

のように途中結果を区間として保持せずに区間を作成できる．
問 (2):

$$\begin{aligned} c - a \cdot b \subset & c - [a \underline{.} b,\ a \overline{.} b] \\ \subset & [c, c] - [a \underline{.} b,\ a \overline{.} b] \\ \subset & [c \underline{-} a \overline{.} b,\ c \overline{-} a \underline{.} b] \end{aligned}$$

上記の計算により，

```
fesetround(FE_DOWNWARD);
  tmp_lower = a*b;
fesetround(FE_UPWARD);
  tmp_upper = a*b;
  res_upper = c - tmp_lower;
fesetround(FE_DOWNWARD);
  res_lower = c - tmp_upper;
```

のように $a \cdot b$ の結果を区間として保持しなければならない．
　初めて扱った人がしがちなミスとして

誤ったプログラム：

```
fesetround(FE_DOWNWARD);
  res_lower = c - a*b;
fesetround(FE_UPWARD);
  res_upper = c - a*b;
```

としてしまいます．これでは，数学的に正しくない結果を返しているので注意してください．
問 (3):

$$\begin{aligned} a \cdot b + [c_l, c_u] \subset & [a \underline{.} b,\ a \overline{.} b] + [c_l, c_u] \\ \subset & [a \underline{.} b \underline{+} c_l,\ a \overline{.} b \overline{+} c_u] \end{aligned}$$

上記の計算により，

```
fesetround(FE_DOWNWARD);
  res_lower = a*b + cl;
fesetround(FE_UPWARD);
  res_upper = a*b + cu;
```

のように $a \cdot b$ の結果を区間として保持しなければならない．この結果から浮動小数点数の n 次元ベクトル $a \in \mathbb{F}^n$ と $b \in \mathbb{F}^n$ の内積 $a \cdot b$ は区間

$$[a_1 \underline{\cdot} a_2 \underline{+} a_1 \underline{\cdot} a_2 \underline{+} \cdots \underline{+} a_n \underline{\cdot} a_n, \quad a_1 \overline{\cdot} a_2 \overline{+} a_1 \overline{\cdot} a_2 \overline{+} \cdots \overline{+} a_n \overline{\cdot} a_n] \tag{1.2}$$

として得られるため，

```
fesetround(FE_DOWNWARD);
  for (int i = 0; i < n; i++)
    res_lower += a[i]*b[i];
fesetround(FE_UPWARD);
  for (int i = 0; i < n; i++)
    res_upper += a[i]*b[i];
```

のように計算できる．この例のように「一気に計算する技を利用できるのはあくまで数学的に正しい状況のみ」で，基本は「機械区間演算」であることに注意してください．

1.5 浮動小数点数で成分を持つ点行列同士の行列積

前節の (1.2) の結果をまとめると浮動小数点数で成分を持つベクトルの内積は次のようなことが言えます:

定理 1.5.1 (浮動小数点数における内積の包含)**.** 浮動小数点数の n 次元のベクトル $a \in \mathbb{F}^n$ と $b \in \mathbb{F}^n$ は

$$a \cdot b \subset [a \underline{\cdot} b, a \overline{\cdot} b]$$

となる．ただし，$a \underline{\cdot} b$ は内積の演算をすべて下向き丸めで計算し，$a \overline{\cdot} b$ は内積の演算をすべて上向き丸めで計算することを意味し，(1.2) のように計算していることを前提とする．

もちろん，同じように行列積についても以下のことがいえます:

定理 1.5.2 (浮動小数点数における行列積の包含)**.** 浮動小数点数の行列 $A \in \mathbb{F}^{l \times m}$ と $B \in \mathbb{F}^{m \times n}$ は

$$A \cdot B \subset [A \underline{\cdot} B, A \overline{\cdot} B]$$

となる．ただし，$A \underline{\cdot} B$ は行列積の演算をすべて下向き丸めで計算し，$a \overline{\cdot} b$ は行列積の演算をすべて上向き丸めで計算することを意味し，各成分ごとで現れる内積は (1.2) のように計算していることを前提とする．

最後の注意書きは，行列積を工夫するアルゴリズム Strassen などでは引き算が含まれるため，上記の定理の範囲外になってしまいますので注意してください．例えば，$A,\ B \in \mathbb{F}^{n \times n}$ としたときの行列積は以下のように計算できます:

```
#include <fenv.h>
fesetround(FE_UPWARD);              //上向き丸めモードへの変更
C_u=R*A;                            //上向き丸めで行列積の計算
fesetround(FE_DOWNWARD);            //下向き丸めモードへの変更
C_d=R*A;                            //下向き丸めで行列積の計算
```

行列積 $A*B$ に最適化された BLAS(例えば dgemm など) を用いることで非常に高速に点行列同士の積の包含を得ることができます．ただし，例えば 3 つの行列の積 $A*B*C$ などの一般的な場合は区間行列の演算になるため結果の包含は単純には得られないので注意してください．区間同士の行列積については 2 章で取り上げます．

また，注意点としては，$R*A$ を演算する前に丸めモードを変更していますが，行列演算に利用する全ノード・全スレッドの丸めモードが変更されている必要があります．そのため，BLAS などの高速ルーチンを利用する際には全ノード・全スレッドの丸めモードが変更されている保証が必要になります．

1.6 区間演算の落とし穴

区間演算を初めて勉強したとき著者は「すべて区間演算で計算すればいいんじゃない?」って思いました．しかし，残念ながら世の中そんなにうまい話はなく,「区間演算を繰り返すことによる過大評価」という最大の欠点があります．そのために，現在では,「最初に近似解を求めてから，その結果の品質を保証する際に区間演算を使う」という検算手法として区間演算は利用されます．連立一次方程式を例にとると，(1) ガウスの消去法は通常の浮動小数点数の演算で行い近似解を得る，(2) その上で，得られた近似解と真の解の誤差を計算する際に区間演算で数学的に正しい結果で証明，のような流れです．

ここでは，欠点となる例をいくつか紹介したいと思います．まず，$x^2, x \in [-1,1]$ について取り扱いたいと思います．$x^2 = x \cdot x$ だからといって演算を行うと

$$[-1,1] \cdot [-1,1] = [-1,1]$$

となります．もちろん，この結果は数学的には正しいです．しかし，よく考えると $0 \leq x^2, \forall x \in \mathbb{R}$ ですので，負の値が区間に含まれてしまう結果は良い結果とは言えません．この原因は，区間演算の定義に立ち戻ると

$$[-1,1] \cdot [-1,1] = \{x \cdot y \mid x \in [-1,1],\ y \in [-1,1]\}$$

となり，x と y は独立した変数としてみなしているため $x = 1, y = -1$ のような状況が発生し，負の値が生まれます．本来は x^2 の像は

$$[-1,1]^2 = \{x^2 = x \cdot x \mid x \in [-1,1]\}$$

となるため，$x \cdot y$ のように独立には動かず

$$[-1,1]^2 = [0,1]$$

のようになります．あくまで区間演算は 2 つの区間に依存関係がない独立した変数としての結果となるため，依存関係がある演算を行う場合には過大評価になってしまいます．

同じように，$x \in [-1,1]$ に対して，

$$x - x = [-1,1] - [-1,1] = [-2,2] = \{x - y \mid \forall x \in [-1,1],\ \forall y \in [-1,1]\}$$

となってしまうため，過大評価になってしまいます (実際は $\{x - x = 0 \mid \forall x \in [-1,1]\}$)．$x^2$ や $x - x$ のようにわかりきった計算の場合は，専用の演算をプログラムすればよいですが，例えば漸化式

$$\begin{aligned} x_0 \in & [x_l, x_u] \\ x_1 \in & [y_l, y_u] \\ x_n = & f(x_{n-1}, x_{n-2}) \end{aligned}$$

のように依存関係がわかりにくくなってしまう演算は非常に多くあります．解消法としては，平均値形式やアフィン演算などがありますが，本書では取り扱いません．

次の例として回転行列

$$A(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$

と区間ベクトル

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \text{[-1, 1]} \\ \text{[-1, 1]} \end{pmatrix}.$$

について考えてみます．x を図で書いてみると図 1.2 のようになります．

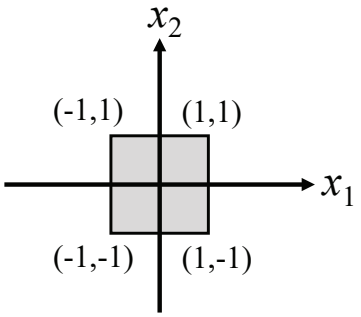

図 1.2: Wrapping effect1

$\pi/4$ だけ x を回転を意味する $A(\pi/4)x$ の像は図 1.3 のようになります．しかしながら，区間演算を用いて計算すると

$$A(\pi/4)x = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \text{[-1, 1]} \\ \text{[-1, 1]} \end{pmatrix} = \begin{pmatrix} [-\sqrt{2}, \sqrt{2}] \\ [-\sqrt{2}, \sqrt{2}] \end{pmatrix},$$

のようになり，図で表すと図 1.4 のオレンジ色の部分になります．図を見てもわかるように，区間演算はそれぞれの軸に対して区間を作成するために，2 次元ベクトル以上の場合は表現できる集合が非常に限定されてしまいます．

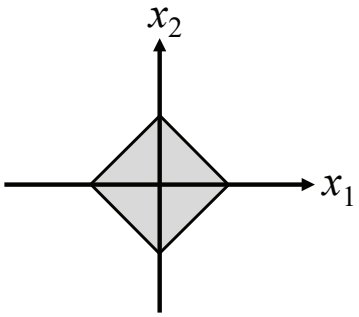

図 1.3: Wrapping effect2

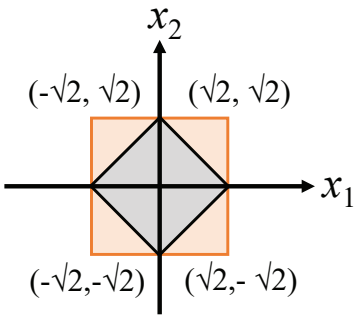

図 1.4: Wrapping effect3

さらに、もう一度，$\pi/4$ だけ回転させる，すなわち $A(\pi/4)\,(A(\pi/4)x)$ とすると真の像は図 1.3 に戻ります．しかしながら，区間演算では

$$A\left(\frac{\pi}{4}\right)\left(A\left(\frac{\pi}{4}\right)x\right) = \begin{pmatrix} [-2,2] \\ [-2,2] \end{pmatrix}$$

から図 1.5 のようになり，非常に過大評価になってしまうことがわかります．

このような影響はラッピングエフェクトと呼ばれます．そのために，区間演算はできる限りあとに行った方が良いです．例えば，計算順序をかえると

$$\left(A\left(\frac{\pi}{4}\right)A\left(\frac{\pi}{4}\right)\right)x = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} [\text{-}1,\ 1] \\ [\text{-}1,\ 1] \end{pmatrix} = \begin{pmatrix} [-1,1] \\ [-1,1] \end{pmatrix}.$$

となることから，精度の良い結果を得ることができます．

1.7　中心半径型の区間と区間演算

今まで扱ってきた区間 $[a_l, a_u] = \{a \in \mathbb{R} \mid a_l \leq a \leq a_u\}$ は，上端と下端から表現される区間であるため，ここでは上端下端型の区間と呼びます．それに対し，中心 $a_m \in \mathbb{R}$ と半径 $a_r \geq 0$ を用いても区間

$$\langle a_m, a_r \rangle := \{a \in \mathbb{R} \mid |a - a_m| \leq a_r\}$$

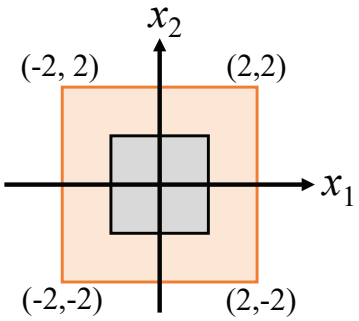

図 1.5: Wrapping effect4

を表現することができ，こちらを中心半径型の区間と呼びます．本節の目標は，中心半径型の区間を利用することで，過大評価にはなるかわりに最大値/最小値の判別が必要ない計算方法を紹介します．この方法を拡張することで，区間同士の行列積に適応することができます．その上，行列積に対し，BLAS などの既存の高速に最適化されたプログラムを利用した高速な品質保証法を実現が可能になります．

まず，上端下端型の区間と中心半径型の区間の関係性は

$$[a_l, a_u] = \left\langle \frac{a_l + a_u}{2},\ \frac{a_u - a_l}{2} \right\rangle$$

であり，

$$\langle a_m, a_r \rangle = [a_m - a_r,\ a_m + a_r]$$

となります．しかしながら，浮動小数点数同士の演算は一般的には浮動小数点数になるとは限らないため,「上端下端型から中心半径型への変換」および「中心半径型から上端下端型への変換」をコンピュータで行う際には誤差が生じます．そのために，双方の変換でも丸めモードの変更を行い元の区間を包含する区間を作成します．

まず,「コンピュータを用いた中心半径型から上端下端型への変換」は単純であり，$a_m, a_r \in \mathbb{F}$ とすると

$$\langle a_m, a_r \rangle \subset [a_m \underline{-} a_r,\ a_m \overline{+} a_r]$$

として変換することができます．それに対し,「コンピュータを用いた上端下端型から中心半径型への変換」は中心の演算に対し，どうすれば良いか考えなければなりません．

定理 1.7.1 (上端下端型から中心半径型への変換)**.** $\mathbb{F}$ を基数 2 の浮動小数点数とし，$a_l, a_u \in \mathbb{F}$ が $a_l \leq a_u$ を満たすとする． そのとき，$a_m, a_r \in \mathbb{F}$ を

$$a_m = (a_l \overline{+} a_u) \overline{/} 2,\ a_r = a_m \overline{-} a_l$$

とすると，

$$[a_l, a_u] \subset \langle a_m, a_r \rangle$$

となる．

証明． $a_m - a_r \leq a_l$ と $a_u \leq a_m + a_r$ を示せばよい．

まず，$a_m - a_r \leq a_l$ から示す．丸めモードの定義から $a_m \overline{-} a_l \geq a_m - a_l$ となることに注意すると

$$a_l + a_r = a_l + (a_m \overline{-} a_l) \geq a_l + (a_m - a_l) = a_m$$

となる．よって，上式から a_r を引くと $a_l \geq a_m - a_r$ になる．

次に，$a_u \leq a_m + a_r$ を示す．まず，$a_l \overline{+} a_u$ の結果は基数 2 の浮動小数点数になる．そのうえで，基数 2 の浮動小数点数に対し，(どのような丸めモードでも)2 を割っても誤差は生じないため，

$$(a_l \overline{+} a_u)\overline{/}2 = (a_l \overline{+} a_u)/2$$

となる．そのうえで，丸めモードの定義から $a_l \overline{+} a_u \geq a_l + a_u$ となることに注意すると

$$\begin{aligned} a_m + a_r =& ((a_l \overline{+} a_u)\overline{/}2) + (a_m \overline{-} a_l) = \left(\frac{a_l \overline{+} a_u}{2}\right) + (a_m \overline{-} a_l) \\ \geq& \frac{a_l + a_u}{2} + a_m - a_l = \frac{a_l + a_u}{2} + \left(((a_l \overline{+} a_u)\overline{/}2\right) - a_l \\ \geq& \frac{a_l + a_u}{2} + \frac{a_l + a_u}{2} - a_l = a_l + a_u - a_l = a_u \end{aligned}$$

□

次に，中心半径型における演算規則を紹介します．足し算と引き算は

$$\begin{aligned} \langle a_m, a_r\rangle \pm \langle b_m, b_r\rangle =& [a_m - a_r,\ a_m + a_r] \pm [b_m - b_r,\ b_m + b_r] \\ =& [a_m \pm b_m - a_r - b_r,\ a_m \pm b_m + a_r + b_r] \\ =& \langle a_m \pm b_m,\ a_r + b_r\rangle \end{aligned}$$

と簡単に得られます (が，わざわざ中心半径型で行うメリットはありません...)．

それに対し，掛け算は工夫が必要であり，中心半径型で計算するメリット (とデメリット) があります．

定理 1.7.2 (中心半径型の区間の積)**.** $a_m, b_m \in \mathbb{R}$ と非負の実数 $a_r, b_r \in \mathbb{R}$ とする． そのとき，中心半径型の区間の積は

$$\langle a_m, a_r\rangle \cdot \langle b_m, b_r\rangle \subset \langle a_m \cdot b_m,\ |a_m| \cdot b_r + a_r \cdot |b_m| + a_r \cdot b_r\rangle$$

のように評価できる．

証明． $\langle a_m, a_r\rangle$ と $\langle b_m, b_r\rangle$ を上端下端型で表すと

$$\begin{aligned} \langle a_m, a_r\rangle =& [a_m - a_r, a_m + a_r] \\ \langle b_m, b_r\rangle =& [b_m - b_r, b_m + b_r] \end{aligned}$$

となる．そのうえ，

$$\langle a_m, a_r\rangle \cdot \langle b_m, b_r\rangle = [a_m - a_r, a_m + a_r] \cdot [b_m - b_r, b_m + b_r]$$

となるため,

$$\begin{aligned}(a_m - a_r)(b_m - b_r) =& a_m b_m - a_m b_r - a_r b_m + a_r b_r \\ (a_m - a_r)(b_m + b_r) =& a_m b_m + a_m b_r - a_r b_m - a_r b_r \\ (a_m + a_r)(b_m - b_r) =& a_m b_m - a_m b_r + a_r b_m - a_r b_r \\ (a_m + a_r)(b_m + b_r) =& a_m b_m + a_m b_r + a_r b_m + a_r b_r\end{aligned}$$

の最小値と最大値がわかればよい．そのうえで,

$$a_m b_m - |a_m| b_r - a_r |b_m| - a_r b_r \leq \begin{matrix} a_m b_m - a_m b_r - a_r b_m + a_r b_r \\ a_m b_m + a_m b_r - a_r b_m - a_r b_r \\ a_m b_m - a_m b_r + a_r b_m - a_r b_r \\ a_m b_m + a_m b_r + a_r b_m + a_r b_r \end{matrix} \leq a_m b_m + |a_m| b_r + a_r |b_m| + a_r b_r$$

となるため,

$$\begin{aligned}\langle a_m, a_r \rangle \cdot \langle b_m, b_r \rangle \subset& [a_m b_m - |a_m| b_r - a_r |b_m| - a_r b_r,\ a_m b_m + |a_m| b_r + a_r |b_m| + a_r b_r] \\ =& \langle a_m b_m,\ |a_m| b_r + a_r |b_m| + a_r b_r \rangle\end{aligned}$$

□

定理 1.7.2 は過大評価というデメリットはあるかわりに，if 文や max, min など分岐を必要せずに計算できることが最大のメリットです．中心半径型の区間の積でもイコールとなる式も知られていますが，if 文や max, min が発生するために，(普通に上端下端型の区間を使えばよいので) あまり利用されません．そのうえで，分岐を必要としない定理 1.7.2 は，区間同士の内積，行列-ベクトル積，行列-行列積が，点同士の内積，行列-ベクトル積，行列-行列積に帰着され，BLAS などの既存の最適化されたルーチンを使用することができます．これにより，行列積の非常に高速な品質保証が可能になります.

第2章　区間ベクトル/行列の演算に対する品質保証法

2.1　記号の準備

本節で定義する記号は本章以降でも利用しますので，定義を忘れた場合に本節に立ち戻って見直してください．まずは，よく利用するベクトルおよび行列を定義したいと思います．$\mathbf{0} \in \mathbb{R}^n$, $O \in \mathbb{R}^{n\times m}$　$I \in \mathbb{R}^{n\times n}$ を

$$\mathbf{0} := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},\ O := \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},\ I := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

のように，それぞれゼロベクトル，ゼロ行列，単位行列とします．

次に，ベクトルおよび行列の半順序 $(\leq)$ を定義したいと思います．2つのベクトル $a = (a_i), b = (b_i) \in \mathbb{R}^n$ あるいは2つの行列 $A = (A_{ij}), B = (B_{ij}) \in \mathbb{R}^{n\times m}$ に対し，

$$a \leq b \text{ はすべての } i \text{ について } a_i \leq b_i \text{ が成立}$$
$$A \leq B \text{ はすべての } i,j \text{ について } A_{ij} \leq B_{ij} \text{ が成立}$$

とします．すなわち，本書ではベクトル・行列の順序を「各成分で比較し，すべての成分で同じ不等号が成立しているとき」を意味します．

続いて，ベクトルおよび行列の絶対値を定義します．ベクトル $a = (a_i) \in \mathbb{R}^n$ および行列 $A = (A_{ij}) \in \mathbb{R}^{n\times m}$ に対する絶対値は

$$|a| := (|a_i|) \in \mathbb{R}^n,\quad |A| := (|A_{ij}|) \in \mathbb{R}^{n\times m}$$

とします．すなわち，各成分に絶対値をつけたベクトル/行列を意味しています．特に，$|A|$ は行列式ではないのでご注意ください．もちろん，$|a| \geq \mathbf{0}$ および $|A| \geq O$ が成立します．

また，ついでにノルムについても定義をしたいと思います．抽象的なノルムについては5.1節の定義5.1.1で記載するとし，ひとまず，記号としてベクトル $x = (x_1, x_2, \cdots x_n)^T \in \mathbb{C}^n$ に対する1ノルム，2ノルム，最大値ノルムをそれぞれ

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|,\quad \|x\|_2 := \sqrt{\sum_{i=1}^{n} |x_i|^2},\quad \|x\|_\infty := \max_{1\leq i\leq n} |x_i|$$

と定義します．次に，正方行列 $A \in \mathbb{C}^{n\times n}$ に固有値を λ_i としたとき，スペクトル半径 $\rho(A)$ を

$$\rho(A) := \max_i |\lambda_i|$$

とします．また，$A=(a_{ij})\in\mathbb{C}^{n\times m}$ に対する行列のノルムを

$$\begin{aligned}\|A\|_1 :=& \max_{1\le j\le m}\sum_{i=1}^{n}|a_{ij}|\\ \|A\|_2 :=& \sqrt{\rho(A^TA)}\\ \|A\|_\infty :=& \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|\end{aligned}$$

と定義します．行列のノルムについても詳細を知りたい場合，5.2.4 項の抽象的な線形作用素ノルムに関する定理 5.2.5 などを参照してください．

また，区間ベクトルおよび区間行列を

$$\begin{aligned}[a,b] :=& \{c\in\mathbb{R}^n \mid a\le c\le b,\ \forall a,b\in\mathbb{R}^n\}\\ [A,B] :=& \{C\in\mathbb{R}^{n\times m} \mid A\le C\le B,\ \forall A,B\in\mathbb{R}^{n\times m}\}\end{aligned}$$

とします．同じように中心半径型の区間ベクトルおよび区間行列も

$$\begin{aligned}\langle c,\ r\rangle :=& \{a\in\mathbb{R}^n \mid |c-a|\le r\in\mathbb{R}^n\}\\ \langle C,\ R\rangle :=& \{A\in\mathbb{R}^{n\times m} \mid |C-A|\le R\in\mathbb{R}^{n\times m}\}\end{aligned}$$

のように定義します．ここで，$|c-a|\le r$ や $|C-A|\le R$ はベクトルや行列の絶対値や不等号であることに注意してください．上端下端型の区間と中心半径型の区間の関係性は，1.7 節と変わらず，

$$\begin{aligned}[a,b] =& \langle (a+b)/2,\ (b-a)/2\rangle, & [A,B] =& \langle (A+B)/2,\ (B-A)/2\rangle\\ \langle c,r\rangle =& [c-r,c+r], & \langle C,R\rangle =& [C-R,C+R]\end{aligned}$$

のようになります．

2.2 区間行列積の計算方法

ここでは，区間行列積の計算方法を紹介します．とはいっても，区間行列積の各成分の結果は，区間ベクトルの内積に帰着されるために，まずは，区間ベクトルの内積にたいする計算方法を紹介します．

まずは，2 つの中心半径型の 2 次元区間ベクトルの内積を考えてみます．すなわち，$a_m=(a_m^i), b_c=(b_m^i)\in\mathbb{R}^2, a_r=(a_r^i), b_r=(b_r^i)\ge\mathbf{0}\in\mathbb{R}^2$ とし，内積

$$\begin{aligned}\langle a_m,a_r\rangle\cdot\langle b_m,b_r\rangle =& \begin{pmatrix}\langle a_m^1,a_r^1\rangle\\ \langle a_m^2,a_r^2\rangle\end{pmatrix}\cdot\begin{pmatrix}\langle b_m^1,b_r^1\rangle\\ \langle b_m^2,b_r^2\rangle\end{pmatrix}\\ =& \langle a_m^1,a_r^1\rangle\cdot\langle b_m^1,b_r^1\rangle+\langle a_m^2,a_r^2\rangle\cdot\langle b_m^2,b_r^2\rangle\end{aligned}$$

を考えます．中心半径型の区間の積は定理 1.7.2 から

$$\begin{aligned}\langle a_m^1,a_r^1\rangle\cdot\langle b_m^1,b_r^1\rangle &\subset \langle a_m^1b_m^1,\ |a_m^1|b_r^1+a_r^1|b_m^1|+a_r^1b_r^1\rangle\\ \langle a_m^2,a_r^2\rangle\cdot\langle b_m^2,b_r^2\rangle &\subset \langle a_m^2b_m^2,\ |a_m^2|b_r^2+a_r^2|b_m^2|+a_r^2b_r^2\rangle\end{aligned}$$

となります．よって，

$$\begin{aligned}
&\langle a_m, a_r\rangle \cdot \langle b_m, b_r\rangle \\
\subset&\langle a_m^1 b_m^1,\ |a_m^1|b_r^1 + a_r^1|b_m^1| + a_r^1 b_r^1\rangle + \langle a_m^2 b_m^2,\ |a_m^2|b_r^2 + a_r^2|b_m^2| + a_r^2 b_r^2\rangle \\
=&\langle a_m^1 b_m^1 + a_m^2 b_m^2,\ (|a_m^1|b_r^1 + |a_m^2|b_r^2) + (a_r^1|b_m^1| + a_r^2|b_m^2|) + (a_r^1 b_r^1 + a_r^2 b_r^2)\rangle \\
=&\langle a_m \cdot b_m,\ |a_m| \cdot b_r + a_r \cdot |b_m| + a_r \cdot b_r\rangle
\end{aligned}$$

となり，定理 1.7.2 の過大評価を許容すれば,「区間ベクトルの内積は，点ベクトルの内積 4 回および点ベクトルの足し算 2 回」に書き直すことができます.

よって次のことの定理がいえます:

定理 2.2.1 (中心半径型の区間ベクトルの内積)**.** $a_m, b_m, \in \mathbb{R}^n$ と非負ベクトル $a_r, b_r \in \mathbb{R}^n$ とする．そのとき，中心半径型の区間の内積は

$$\langle a_m, a_r\rangle \cdot \langle b_m, b_r\rangle \subset\langle a_m \cdot b_m,\ |a_m| \cdot b_r + a_r \cdot |b_m| + a_r \cdot b_r\rangle$$

のように評価できる．

また，同様に区間行列積についても以下が成り立ちます:

定理 2.2.2 (中心半径型の区間行列の行列積)**.** $A_m \in \mathbb{R}^{l\times m}, B_m \in \mathbb{R}^{m\times n}$ と非負行列 $A_r \in \mathbb{R}^{l\times m}, B_r \in \mathbb{R}^{m\times n}$ とする． そのとき，中心半径型の区間行列の行列積は

$$\langle A_m, A_r\rangle \cdot \langle B_m, B_r\rangle \subset\langle A_m \cdot B_m,\ |A_m| \cdot B_r + A_r \cdot |B_m| + A_r \cdot B_r\rangle$$

のように評価できる．

本章の最後に，浮動小数点数で値を持つ上端下端型の区間行列 $[A_l, A_u], [B_l, B_u]$ $(A_l, A_u \in \mathbb{F}^{l\times m}, A_l \leq A_u$ および $B_l, B_u \in \mathbb{F}^{m\times n}, B_l \leq B_u)$ の行列積に関して，品質を保証する方法を紹介します:

定理 2.2.3 (浮動小数点数を両端にもつ上端下端型の区間行列積)**.** 区間行列 $[A_l, A_u], [B_l, B_u]$ $(A_l, A_u \in \mathbb{F}^{l\times m}, A_l \leq A_u$ および $B_l, B_u \in \mathbb{F}^{m\times n}, B_l \leq B_u)$ とする．$A_m, A_r \in \mathbb{F}^{l\times m},\ B_m, B_r \in \mathbb{F}^{m\times n}$ をそれぞれ

$$\begin{aligned}
A_m :=& (A_l \overline{+} A_u)\overline{/}2, & A_r :=& A_m \overline{-} A_l \\
B_m :=& (B_l \overline{+} B_u)\overline{/}2, & B_r :=& B_m \overline{-} B_l
\end{aligned}$$

とし，$C_r \in \mathbb{F}^{l\times n}$ を

$$C_r := |A_m| \overline{\cdot} B_r \overline{+} A_r \overline{\cdot} |B_m| \overline{+} A_r \overline{\cdot} B_r$$

とする．そのとき，

$$[A_l, A_u] \cdot [B_l, B_u] \subset[A_m \underline{\cdot} B_m \underline{-} C_r,\ A_m \overline{\cdot} B_m \overline{+} C_r]$$

となる．

証明．方針としては，1) 定理 1.7.1 を用いて中心半径型に変換，2) 定理 2.2.2 から区間行列積を浮動小数点数を値に持つ点行列の行列積と和に変換， 3) 定理 1.5.2 を用いて浮動小数点数を値に持つ点行列の行列積を計算する．

まず，区間行列 $[A_l, A_u], [B_l, B_u]$ を定理 1.7.1 と $A_m, A_r \in \mathbb{F}^{l\times m},\ B_m, B_r \in \mathbb{F}^{m\times n}$ を用いて中心半径型に変換すると

$$[A_l, A_u] \cdot [B_l, B_u] \subset \langle A_m, A_r\rangle \cdot \langle B_m, B_r\rangle$$

となる．

そのうえで，定理 2.2.2 を上の式に適用すると

$$[A_l, A_u] \cdot [B_l, B_u] \subset \langle A_m \cdot B_m,\ |A_m| \cdot B_r + A_r \cdot |B_m| + A_r \cdot B_r\rangle$$

となる．半径 $|A_m| \cdot B_m + A_r \cdot |B_m| + A_r \cdot B_r$ については，大きくなればなるほど，元の集合を包含する形で集合が大きくなる．そのうえで，定理 1.5.2 を用いて浮動小数点数を値に持つ点行列の行列積を用いると

$$|A_m| \cdot B_m + A_r \cdot |B_m| + A_r \cdot B_r \le C_r$$

となるため，

$$[A_l, A_u] \cdot [B_l, B_u] \subset \langle A_m \cdot B_m,\ C_r\rangle$$

となる．そのうえで，右辺の中心半径型を上端下端型に変換すると

$$\langle A_m \cdot B_m,\ C_r\rangle = [A_m \cdot B_m - C_r,\ A_m \cdot B_m + C_r]$$

となる．さらに，再び定理 1.5.2 を用いて浮動小数点数を値に持つ点行列の行列積を用いると

$$[A_m \cdot B_m - C_r,\ A_m \cdot B_m + C_r] \subset [A_m \dot{-} B_m \underline{-} C_r,\ A_m \dot{+} B_m \overline{+} C_r]$$

よって，

$$[A_l, A_u] \cdot [B_l, B_u] \subset [A_m \dot{-} B_m \underline{-} C_r,\ A_m \dot{+} B_m \overline{+} C_r]$$

となる． □

最後に定理 2.2.3 に対する疑似コードを紹介します:

```
// 区間行列積 [Al, Au] * [Bl, Bu] => [Cl, Cu] を作成
#include <fenv.h>
[Cl, Cu] = interval_matmul(Al, Au, Bl, Bu)
  fesetround(FE_UPWARD);              //上向き丸めモードへの変更
    Am = (Al + Au)/2;
    Ar = Am - Al;
    Bm = (Bl + Bu)/2;
    Br = Bm - Bl;

    Cr = abs(Am)*Br + Ar*abs(Bm) + Ar*Br;
```

```
    Cu = Am*Bm + Cr;

  fesetround(FE_DOWNWARD);            //下向き丸めモードへの変更
    Cl = Am*Bm - Cr;

  fesetround(FE_TONEAREST);  // 最近点丸めにもどす
return Cl, Cu
```

第3章　連立一次方程式の解の品質保証法

本章では，$A \in \mathbb{F}^{n\times n}$, $b \in \mathbb{F}^n$ における連立一次方程式

$$Ax = b \tag{3.1}$$

の近似解 $\hat{x}$ に対して，品質保証を行う方法を紹介します．前章の最初に紹介した 1.1 節とは違い，線形問題である代わりに，n によっては到底，手で解けるとは思えない問題設定です．例えば，$n = 10$ としたら，手で解くのは非常に大変ですが，コンピュータならば $n = 10000$ としても近似解を求めることが可能です．その代わり，あくまで近似解なので，本当の解と近いかどうかはわかりません．そのために，近似解に対し品質保証を実施していきましょう．

3.1　標準的な品質保証法

この節では，連立一次方程式 (3.1) の近似解を品質保証法する，もっとも標準的な方法を紹介したいと思います．次の節以降で紹介する H 行列を用いた品質保証法のほうが，さらに精密になることが知られていますが，最初の一歩ということで，標準的な方法から取り掛かります．

定理 1.1.1 と同じで，第 5 章の最終目標が，以下の定理をものすごく一般化した定理 5.4.2 の証明ですので，ここでは証明をつけずに利用だけしたい思います．

定理 3.1.1. A, $R \in \mathbb{R}^{n\times n}$，$\hat{x}$, $b \in \mathbb{R}^n$，I を単位行列とする．もし

$$\|RA - I\|_\infty < 1 \tag{3.2}$$

を満たすならば，A は逆行列を持ち，$x^* := A^{-1}b \in \mathbb{R}^n$ とすると

$$\|x^* - \hat{x}\|_\infty \leq \frac{\|R(A\hat{x} - b)\|_\infty}{1 - \|RA - I\|_\infty} \tag{3.3}$$

となる．

それでは，定理 3.1.1 をかみ砕いて見ていきましょう．まず，$A \in \mathbb{R}^{n\times n}$ と $b \in \mathbb{R}^n$ は与えられていて，連立一次方程式

$$\text{Find } x \in \mathbb{R}^n \text{ s.t.} Ax = b$$

を満たす解 x^* を探す問題を考えています．$\hat{x}$ は定理内ではなんでも良いベクトルですが，$\hat{x}$ が品質保証をかけたい対象です．すなわち，$Ax = b$ に対し何かしらの方法で求めた近似解 $\hat{x}$ とします．また，R も定理内ではなんでも良い行列ですが，多くの場合，A に対する近似逆行列とします．$\hat{x}$ や R を作成する段階は，丸め誤差を気にする必要はありません．そのために，区間演算を使用する必要はないです．

つぎに，(3.2) をチェックします．ここから，区間演算で計算する必要があります．すなわち，$\|RA - I\|_\infty$ を丸め誤差まで含めて求める必要があります．そのため，多くの場合は計算が簡単な

最大値ノルムを選択することが多いです．その上で，$\|RA-I\|_\infty$ の丸め誤差まで含めた上界が 1 未満であれば，行列 A が正則であることが証明され，解を 1 つ持つことが示されました．

さらに，(3.3) の右辺

$$\frac{\|R(A\hat{x}-b)\|_\infty}{1-\|RA-I\|_\infty} \leq c$$

も全て丸め誤差まで考慮した区間演算で上界 c を求めます．その結果，真の解 x^* と近似解 $\hat{x}$ の差は

$$\|x^*-\hat{x}\|_\infty \leq c$$

となります．最大値ノルムの定義から,「真の解 x^* と近似解 $\hat{x}$ の差の最大値は c よりも小さい」ということを意味します．c が許容範囲であるかどうかは，もちろん，使用したい問題やユーザーによって変わります．ある閾値を決めて，c が閾値よりも大きい場合は,「品質が悪い近似解」というようにし，再度求めなおすようにしましょう．

また，定理 5.4.2 を使うと $\|RA-I\|_\infty$ ではなく $\|I-RA\|_\infty$ であることに気が付くと思います．もちろん，ノルムの定義から $\|RA-I\|_\infty = \|I-RA\|_\infty$ のためどちらでも良いのですが，区間演算を利用する際には，$\|I-RA\|_\infty$ と $\|RA-I\|_\infty$ のどちらのほうが計算しやすいか考えてみると良いと思います．もちろん，区間演算では $\|RA-I\|_\infty$ を計算するほうが楽だからこちらで記載しています.[1] $A\hat{x}-b$ も同様の理由で $b-A\hat{x}$ ではなく，$A\hat{x}-b$ にしております．ただし，$A\hat{x}-b$ はちゃんと区間で結果を得ておかないと，$R(A\hat{x}-b)$ は計算できないので注意してください．

3.2 成分毎評価

定理 3.1.1 では「真の解 x^* と近似解 $\hat{x}$ の差の絶対値最大」を見積もっていましたが，有限次元問題の場合は，簡単に成分毎の評価に切り替えることができます．まず，よく知られている定理であるノイマン級数の定理から準備したいと思います．無限次元 Banach 空間上の線形作用素に関するノイマン級数展開に関するについては定理 5.2.3 を参照してください．

定理 3.2.1. $G \in \mathbb{R}^{n\times n}$ としたとき，以下は同等である

1. $\rho(G) < 1$ (スペクトル半径が 1 未満)
2. $I-G$ が正則で，$(I-G)^{-1} = I+G+G^2+\cdots$ の右辺の級数は収束する

証明． $1 \Rightarrow 2$ の証明

λ を G の固有値とすると $1-\lambda$ は $I-G$ の固有値になる．そのうえ，$\rho(G) = \max_i |\lambda_i| < 1$ より，$1-\lambda_i \neq 0,\ 1 \leq i \leq n$ となり，$I-G$ はゼロ固有値を持たないため，正則である．また，

$$\begin{aligned}
G^{k+1} =& (I+G+\cdots+G^k) - (I+G+\cdots+G^k) + G^{K+1} \\
=& (I+G+\cdots+G^k+G^{K+1}) - (I+G+\cdots+G^k) \\
=& (I+G(I+G+\cdots+G^k)) - (I+G+\cdots+G^k) \\
=& I + (G-I)(I+G+\cdots+G^k) \\
=& I - (I-G)(I+G+\cdots+G^k)
\end{aligned}$$

となるため，左から $(I-G)^{-1}$ をかけると

$$(I-G)^{-1}G^{k+1} = (I-G)^{-1} - (I+G+\cdots+G^k)$$

[1] ヒント: 1.4 節の問題や定理 2.2.3 および，2.1 節の $\|\cdot\|_\infty$ の定義に立ち戻ってみてください．

となる．そのうえで，

$$\|(I-G)^{-1}-(I+G+\cdots+G^k)\| \leq \|(I-G)^{-1}\|\|G^{k+1}\|$$

となる．よく知られている Oldenburger の定理「$\rho(G)<1 \Leftrightarrow G^k \to O,\ (k\to\infty)$」を利用[2]すると

$$\|(I-G)^{-1}-(I+G+\cdots+G^k)\| \to 0\ (k\to\infty)$$

となる．

$2 \Rightarrow 1$ の証明

$\lambda \in \mathbb{C}$ を G の絶対値最大固有値に対応する固有値 (すなわち $\rho(G)=|\lambda|$) とすると

$$Gx=\lambda x$$

となる固有ベクトルが存在する．同様に，

$$G^2x=G(Gx)=\lambda Gx=\lambda^2 x,\ \ G^kx=\lambda^k x$$

となることに注意すると

$$(I+G+G^2+\cdots+G^k)x=(1+\lambda+\lambda^2+\cdots\lambda^k)x$$

となる．等比級数から

$$1+\lambda+\lambda^2+\cdots\lambda^k=\frac{1-\lambda^{N+1}}{1-\lambda}$$

であるため，

$$(I+G+G^2+\cdots+G^k)x=\frac{1-\lambda^{N+1}}{1-\lambda}x$$

となる．そのうえで，$k\to\infty$ としたとき，左辺が収束するために右辺 $(1-\lambda^{N+1})/(1-\lambda)$ も収束ため，等比級数の収束条件より $|\lambda|<1$ となる．ゆえに，$\rho(G)<1$ となる．

□

続いて，スペクトル半径と行列のノルムの関係について準備をします:

定理 3.2.2. $G\in\mathbb{R}^{n\times n}$ とする．行列ノルム $\|\cdot\|$ を

$$\|G\|:=\sup_{x\in\mathbb{R}^n}\frac{\|Gx\|}{\|x\|}$$

とすると，

$$\rho(G)\leq\|G\|$$

となる．(※ 上の行列のノルムは，右辺のベクトルのノルムを $1,2,\infty$ と変更すれば，それぞれ行列の 1 ノルム，2 ノルム，∞ ノルムと一致します．)

[2]すみません，いいわけです．ここは，線形代数の範囲とも言えず，証明 (特に「$\rho(G)<1\Rightarrow G^k\to O,\ (k\to\infty)$」) にはジョルダン標準形を使用します．しかし最初から定義して証明するにも，ここ以外では一切使用しないあまりにもコスパが悪かったために結果のみ使用しました...

証明． G の絶対値最大固有値を λ とし，それに対応する固有ベクトルを $x \in \mathbb{R}^n$ とする．行列のノルムの定義から

$$\|G\| = \sup_{y \in \mathbb{R}^n} \frac{\|Gy\|}{\|y\|} \geq \frac{\|Gx\|}{\|x\|} = \frac{\|\lambda x\|}{\|x\|} = |\lambda| = \rho(G)$$

となるため，題意は示せた．

□

次に，真の解が得られる状況における等式を紹介します:

補題 3.2.1. $A, \in \mathbb{R}^{n \times n}$，$\hat{x}, b \in \mathbb{R}^n$，$I$ を単位行列とする． 行列 $G \in \mathbb{R}^{n \times n}$ を

$$G := I - RA$$

とする．もし RA が逆行列を持つならば，A は逆行列を持ち $x^* := A^{-1}b$ とすると，整数 $N \geq 0$ に対し，

$$x^* - \hat{x} = \sum_{i=0}^{N} G^i R(b - A\hat{x}) + G^{N+1}(x^* - \hat{x})$$

となる．但し，$G^0 = I$ であることに注意する．

証明． $\mathrm{rank}(A)$ を行列の階数とすると，行列積の階数は $\mathrm{rank}(RA) \leq \mathrm{rank}(R)$，$\mathrm{rank}(RA) \leq \mathrm{rank}(A)$ となることが知られている (忘れてしまっている場合，線形代数の教科書で探してみましょう!)．その上，RA が逆行列を持つことから線形代数の基本定理 (あるいは次元定理とも呼ばれます．こちらも忘れている場合は，線形代数の教科書をみなおしましょう) から $\mathrm{rand}(RA) = n$ となるために，$\mathrm{rank}(RA) = \mathrm{rank}(A) = \mathrm{rank}(R) = n$ となり，A も R も全単射となり逆行列を持つ．よって，

$$\begin{aligned}
x^* - \hat{x} =& R(b - A\hat{x}) - R(b - A\hat{x}) + x^* - \hat{x} \\
=& R(b - A\hat{x}) - RA(A^{-1}b - \hat{x}) + x^* - \hat{x} \\
=& R(b - A\hat{x}) - RA(x^* - \hat{x}) + x^* - \hat{x} \\
=& R(b - A\hat{x}) + (I - RA)(x^* - \hat{x})
\end{aligned}$$

から

$$x^* - \hat{x} = R(b - A\hat{x}) + G(x^* - \hat{x}) \tag{3.4}$$

を得る．右辺の $x^* - \hat{x}$ に上の式を代入すると

$$\begin{aligned}
x^* - \hat{x} =& R(b - A\hat{x}) + G\left(R(b - A\hat{x}) + G(x^* - \hat{x})\right) \\
=& (I + G)R(b - A\hat{x}) + G^2(x^* - \hat{x})
\end{aligned}$$

再度，(3.4) を上の式に代入すると

$$x^* - \hat{x} = (I + G + G^2)R(b - A\hat{x}) + G^3(x^* - \hat{x})$$

同じ手続きを繰り返すと題意が得られる．

□

また，ベクトルの絶対値とノルムとの関係性として以下の補題も得られます．

補題 3.2.2. $\boldsymbol{e} \in \mathbb{R}^n$ をすべての要素が 1 のベクトルとする．そのとき，ベクトル $x \in \mathbb{R}^n$ に対し，

$$|x| \leq \|x\|_p \boldsymbol{e},\ \forall p \in \{1, 2, \infty\}$$

となる．

証明． ノルムの定義から $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$ となるため，$\|\cdot\|_\infty$ についてのみ示せばよい．その上，$\|x\|_\infty = \max_{1 \leq i \leq n}(|x_i|)$ であるため，

$$|x| \leq \begin{pmatrix} \|x\|_\infty \\ \|x\|_\infty \\ \vdots \\ \|x\|_\infty \end{pmatrix} \leq \|x\|_\infty \boldsymbol{e}$$

となり，題意は得られた． □

定理 3.1.1, 補題 3.2.1，および，補題 3.2.2 を用いることで，次のような成分毎評価が得られます:

定理 3.2.3. $\boldsymbol{e} \in \mathbb{R}^n$ をすべての要素が 1 のベクトルとする．$A,\ R \in \mathbb{R}^{n \times n}$，$\hat{x},\ b \in \mathbb{R}^n$，$I$ を単位行列とする．行列 $G \in \mathbb{R}^{n \times n}$ を

$$G := I - RA$$

とする．もし

$$\|G\|_\infty < 1 \tag{3.5}$$

を満たすならば，A は逆行列を持ち，$x^* := A^{-1}b \in \mathbb{R}^n$ とすると

$$|x^* - \hat{x}| \leq \left| \sum_{i=0}^{N} G^i R(b - A\hat{x}) \right| + \frac{\|R(A\hat{x} - b)\|_\infty}{1 - \|RA - I\|_\infty} \left| G^{N+1} \right| \boldsymbol{e} \tag{3.6}$$

となる．

証明． (3.5) と定理 3.2.2 より $\rho(G) < 1$ であることからノイマン級数の定理である定理 3.2.1 を用いると RA は逆行列を持つ．そのうえで，補題 3.2.1 から

$$\begin{aligned} |x^* - \hat{x}| &= \left| \sum_{i=0}^{N} G^i R(b - A\hat{x}) + G^{N+1}(x^* - \hat{x}) \right| \\ &\leq \left| \sum_{i=0}^{N} G^i R(b - A\hat{x}) \right| + \left| G^{N+1}(x^* - \hat{x}) \right| \\ &\leq \left| \sum_{i=0}^{N} G^i R(b - A\hat{x}) \right| + \left| G^{N+1} \right| |x^* - \hat{x}| \end{aligned}$$

となる．さらに，補題 3.2.1 および定理 3.1.1 を上式に適用すると

$$\begin{aligned} |x^* - \hat{x}| &\leq \left| \sum_{i=0}^{N} G^i R(b - A\hat{x}) \right| + \left| G^{N+1} \right| \|x^* - \hat{x}\|_\infty \boldsymbol{e} \\ &\leq \left| \sum_{i=0}^{N} G^i R(b - A\hat{x}) \right| + \frac{\|R(A\hat{x} - b)\|_\infty}{1 - \|RA - I\|_\infty} \left| G^{N+1} \right| \boldsymbol{e} \end{aligned}$$

となり，題意は得られた． □

定理 3.2.3 に対して，$i = 0$ のときは特に山本の定理として知られています:

系 3.2.1. $\boldsymbol{e} \in \mathbb{R}^n$ をすべての要素が 1 のベクトルとする．$A,\ R \in \mathbb{R}^{n\times n}$，$\hat{x},\ b \in \mathbb{R}^n$，$I$ を単位行列とする．行列 $G \in \mathbb{R}^{n\times n}$ を

$$G := I - RA$$

とする．もし

$$\|G\|_\infty < 1$$

を満たすならば，A は逆行列を持ち，$x^* := A^{-1}b \in \mathbb{R}^n$ とすると

$$|x^* - \hat{x}| \leq |R(b - A\hat{x})| + \frac{\|R(A\hat{x} - b)\|_\infty}{1 - \|RA - I\|_\infty} |G|\, \boldsymbol{e}$$

となる．

3.3 (発展) さらに評価を極めるには?

定理 3.1.1 や定理 3.2.3 の十分条件は $\|I - RA\|_\infty < 1$ でした．もちろん，R が A の逆行列に近ければ，この十分条件は満足することが期待できます．しかし，A が悪条件と呼ばれる条件数が高い問題では十分条件が通らなくなってきてしまいます．そこで，現状持っている R で品質を保証するためには十分条件 $\|I - RA\|_\infty < 1$ をどこまで良い条件にできるのか?計算コストに見合ったものなのか?など気になってしまいます．

例えば，定理 3.1.1 や定理 3.2.3 の十分条件を $\|I - RA\|_2 < 1$ のように 2 ノルムを用いることで，$\|I - RA\|_2 \leq \|I - RA\|_\infty$ となるために，$\|I - RA\|_2$ が計算量や精度の意味でリーズナブルに評価できれば，良い評価になることが期待できます．しかしながら，$\|I - RA\|_2$ は定義から $\sqrt{\lambda((I - RA)^T(I - RA))}$ のような計算を必要とするため，

- 品質が保証された固有値の結果が必要 (一般的には，連立一次方程式の方が簡単)
- 行列積が発生しており，追加で $O(n^3)$ の計算が必要 (計算コスト増)
- $(I - RA)^T(I - RA)$ は条件数が 2 乗されるため，解きにくくなる

など懸念事項があり，利用されません．

そこで，本節では「計算量や精度の意味でリーズナブルに評価可能な範囲における現在最強の十分条件」を紹介します．まず，着目する点は，まだ，ノルムを使用していない状態である補題 3.2.1 です．この補題をみてみると RA が正則であれば，A も正則で解を持ち，イコールで評価できることがわかります．そこで，RA の正則性について，何とか検証できないか?って考えてみると定理 3.2.1 があります．$G = I - RA$ とすると，定理 3.2.1 から,「$\rho(G) < 1 \Leftrightarrow I - G(= RA)$ が正則で $\sum G^i$ は収束する」といえます．すなわち，$\rho(G) < 1$ を検証できればうれしいです．

しかしながら，残念なことに，まだ，$\rho(G) < 1$ の「計算量や精度の意味でリーズナブルな評価法」はまだ見つかっていません．そこで，少しだけ，次の定理を用いて過大評価にします:

定理 3.3.1. 任意の行列 $A \in \mathbb{R}^{n\times n}$ について

$$\rho(A) \leq \rho(|A|)$$

となる．

すなわち，$\rho(G) \leq \rho(|G|) < 1$ として考えて，$\rho(|G|) < 1$ を検証する方法を考えます．$\rho(|G|)$ はどうなのか?っと疑問に思うので，関係性を示すと

$$\rho(G) \leq \rho(|G|) \leq \||G|\|_\infty = \|G\|_\infty$$

となるため，定理 3.1.1 や定理 3.2.3 の十分条件 $\|G\|_\infty < 1$ を検証するよりかはよさそうだ!とわかると思います．

つぎに，対角行列 $D \in \mathbb{R}^{n\times n}$(対角成分以外はすべてゼロ) と非対角行列 $E \in \mathbb{R}^{n\times n}$(対角成分はすべてゼロ) を次のように定義します:

$$RA = D + E$$

そのとき，R^{new} を

$$R^{new} := D^{-1}R$$

として，定義すると，$R^{new}A$ は対角成分がすべて 1 の行列になります．もともと，定理 3.1.1 や定理 3.2.3 の R はなんでもよいものだったため，R^{new} を R とみなして，検証を考えます．(ちなみに，この作業をしなくても良いのですが，対角成分についてほんの少しだけ評価が良くなります．)

そのうえ，次の定理 (Fiedler-Ptaák の定理の系) がいえます:

定理 3.3.2. $A,\ R \in \mathbb{R}^{n\times n}$，$\hat{x},\ b \in \mathbb{R}^n$，$I$ を単位行列とする．対角行列 $D \in \mathbb{R}^{n\times n}$(対角成分以外はすべてゼロ) と非対角行列 $E \in \mathbb{R}^{n\times n}$(対角成分はすべてゼロ) を:

$$RA = D + E$$

とする．行列 $G \in \mathbb{R}^{n\times n}$ を

$$G := I - D^{-1}RA$$

とする．行列 $M \in \mathbb{R}^{n\times n}$ を

$$M := I - |G|$$

とする．そのとき，以下は同等である:

1. $\rho(|G|) < 1$
2. M が正則で，$M^{-1} \geq O$
3. $Mv > \mathbf{0}$ となる正のベクトル $v > \mathbf{0}$ が存在

すなわち,「$Mx > \mathbf{0}$ となる正のベクトル $x > \mathbf{0}$ を 1 つ作成すれば良いです．そのうえ,「M が正則で，$M^{-1} \geq O$」と「任意の $v \in \mathbb{R}^n$ に対して $Mv \geq \mathbf{0} \Rightarrow v \geq \mathbf{0}$ 」です．そのため，例えば，

$$Mv = \boldsymbol{e}$$

となる方程式を ($O(n^2)$ の範疇で) 近似的に解くことで得られる近似解 $\hat{v}$ を用意し，次を丸め誤差まで含めた区間演算で検証すれば良いです:

$$M\hat{v} > \mathbf{0} \quad \text{かつ} \quad \hat{v} > \mathbf{0}$$

他にも，べき乗法を利用する方法もあります.

上記を満たす $\hat{v}$ が 1 つ見つかれば，$\rho(|G|) < 1$ であるために，$\rho(G) < 1$ から，補題 3.2.1 が使用できます．実際に，補題 3.2.1 の $i = 0$ とし，D^{-1} も加え，$G := I - D^{-1}RA$ として評価すると

$$\begin{aligned}|x^* - \hat{x}| =& |D^{-1}R(b - A\hat{x}) + G(x^* - \hat{x})| \\ \leq& |D^{-1}R(b - A\hat{x})| + |G(x^* - \hat{x})| \\ \leq& |D^{-1}R(b - A\hat{x})| + |G||x^* - \hat{x}|\end{aligned}$$

となり,

$$\begin{aligned}&(I - |G|)|x^* - \hat{x}| \leq |D^{-1}R(b - A\hat{x})| \\ \Leftrightarrow& M|x^* - \hat{x}| \leq |D^{-1}R(b - A\hat{x})|\end{aligned}$$

となります．そのうえ,「$Mv > \mathbf{0}$ となる正のベクトル $v > \mathbf{0}$ が存在」していることから,「M が正則で，$M^{-1} \geq O$」であるために，左から M^{-1} をかけると

$$|x^* - \hat{x}| \leq M^{-1}|D^{-1}R(b - A\hat{x})|$$

となります．そのために，あとは M^{-1} が計算量や精度の意味でリーズナブルに評価できれば，目標が達成するために，M^{-1} の評価について紹介します.

まず,

$$M\hat{v} > \mathbf{0} \ \text{ かつ } \ \hat{v} > \mathbf{0}$$

を満たす $\hat{v}$ に対し,

$$u := M\hat{v} > \mathbf{0}$$

とベクトル u を定義します．さらに，ベクトル $w \in \mathbb{R}^n$ を

$$w_k := \max_{1 \leq i \leq n} \frac{|G|_{ik}}{u_i} \ \text{ for } \ 1 \leq k \leq n$$

とします (ただし，$|G|_{ik}$ は行列 G の絶対をとったときの (i, k) 成分を意味します．)．そのとき,

$$I - M = |G| \leq uw^T$$

となるため，左から $M^{-1}(\geq O)$ をかけると

$$M^{-1} - I \leq M^{-1}uw^T = vw^T$$

となるため,

$$M^{-1} \leq I + vw^T$$

となる．そのために,

$$|x^* - \hat{x}| \leq (I + vw^T)|D^{-1}R(b - A\hat{x})|$$

となります．さらに，補題 3.2.1 を加えると次のような定理になります.

定理 3.3.3. $A,\ R \in \mathbb{R}^{n\times n}$，$\hat{x},\ b \in \mathbb{R}^n$，$I$ を単位行列とする．対角行列 $D \in \mathbb{R}^{n\times n}$(対角成分以外はすべてゼロ) と非対角行列 $E \in \mathbb{R}^{n\times n}$(対角成分はすべてゼロ) を:

$$RA = D + E$$

とする (すなわち，RA の対角成分が D，非対角成分が E です．)．行列 $G \in \mathbb{R}^{n\times n}$ を

$$G := I - D^{-1}RA$$

とする．行列 $M \in \mathbb{R}^{n\times n}$ を

$$M := I - |G|$$

とする．もし，

$$M\hat{v} > \mathbf{0} \ \text{ かつ } \ \hat{v} > \mathbf{0}$$

となるベクトル $\hat{v}$ が存在するならば，A は正則で，ベクトル $u, w \in \mathbb{R}^n$ を

$$\begin{aligned} u :=& M\hat{v} \\ w_k :=& \max_{1\le i\le n} \frac{|G|_{ik}}{u_i} \ \text{ for } \ 1 \le k \le n \end{aligned}$$

とすると真の解 $x^* = A^{-1}b$ と $\hat{x}$ の差は

$$|x^* - \hat{x}| \le \left|\sum_{i=0}^{N} G^i D^{-1}R(b - A\hat{x})\right| + \left|G^{N+1}\right| (I + \hat{v}w^T)|D^{-1}R(b - A\hat{x})|$$

となる．

第4章　固有値問題の固有値に対する品質保証法

4.1　全固有値に対する一般的な品質保証法

本章では $A, B \in \mathbb{C}^{n\times n}$ としたとき，固有値問題

$$\text{Find } (\lambda, x) \in \mathbb{C}\times\mathbb{C}^n \text{ s.t. } Ax = \lambda Bx,\ x \neq \mathbf{0}$$

を考えます．定理 1.1.1 や定理 3.1.1 などでは，第 5 章の最終目標である定理 5.4.2 を利用してきました．線型方程式や非線形方程式と同様に固有値問題にも定理 5.4.2 を使うことも可能ですが，重複固有値の取り扱いや全固有値と全固有ベクトルを 1 つの方程式としてみなす巨大な方程式を考えなければならないため，直接利用することはあまりおすすめしません．そのかわりに数値解析でもよく利用されるゲルシュゴリンの定理を利用します:

定理 4.1.1 (ゲルシュゴリンの定理)**.** n 次元行列 $A = (a_{ij})$ に対し

$$r_i := \sum_{j=1, j\neq i}^{n} |a_{ij}|,\ U_i := \{z \in \mathbb{C} \mid\ |z - a_{ii}| \leq r_i\}$$

とすると固有値問題 $Ax = \lambda x$ のすべての固有値は

$$\bigcup_{i=1}^{n} U_i$$

に含まれる．そのうえ，特に，単連結領域 C が m 個の U_i からなる場合，重複度も含めて m 個の固有値が C 内に存在する．

定理 4.1.1 では $B = I$ としたときに，A の対角成分を中心に，非対角成分の絶対値の和を半径として真の固有値を含む集合を作成しています．しかし，$Ax = \lambda Bx$ に対しては，定理 4.1.1 は直接的には利用できません．そこで，$Ax = \lambda Bx$ を固有値を変えずに $\tilde{X},\ Y \in \mathbb{C}^{n\times n}$ および対角行列 $\tilde{\Lambda} \in \mathbb{C}^{n\times n}$ を式変形をします (固有ベクトルは変わるので注意してください)．それぞれの行列の意味合いは

- $\tilde{X}$: 近似固有ベクトルを並べた行列
- $Y \approx (B\tilde{X})^{-1}$
- $\tilde{\Lambda}$: 対角行列，対角成分に近似固有値を並べた行列

ですが，どんな行列でもかまいません．

補題 4.1.1. $A, B \in \mathbb{C}^{n\times n}$ を与えられている行列とする．$\tilde{X},\ Y \in \mathbb{C}^{n\times n}$ および対角行列 $\tilde{\Lambda} \in \mathbb{C}^{n\times n}$ を与えられている行列とする．もし $YB\tilde{X}$ が正則ならば，固有値問題

$$\text{Find } (\lambda, x) \in \mathbb{C}\times\mathbb{C}^n \text{ s.t. } Ax = \lambda Bx$$

と固有値問題

$$\text{Find } (\mu, y) \in \mathbb{C} \times \mathbb{C}^n \text{ s.t. } \left(\tilde{\Lambda} + (YB\tilde{X})^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right) y = \mu y$$

のそれぞれの固有値は一致する，すなわち，$\lambda_i = \mu_i$ となる．

証明． YBX が正則であることと，行列積によって行列の階数が上がることはないために，$Y, B, \tilde{X} \in \mathbb{R}^{n\times n}$ は各々正則である．そのうえ，

$$\begin{aligned}
&Ax = \lambda Bx \\
\Leftrightarrow &A\tilde{X}\tilde{X}^{-1}x = \lambda B\tilde{X}\tilde{X}^{-1}x \\
\Leftrightarrow &A\tilde{X}y = \lambda B\tilde{X}y, \quad y := \tilde{X}^{-1}x \\
\Leftrightarrow &YA\tilde{X}y = \lambda YB\tilde{X}y \\
\Leftrightarrow &YB\tilde{X}\tilde{\Lambda}y - YB\tilde{X}\tilde{\Lambda}y + YA\tilde{X}y = \lambda YB\tilde{X}y \\
\Leftrightarrow &YB\tilde{X}\tilde{\Lambda}y + Y\left(A\tilde{X} - B\tilde{X}\tilde{\Lambda}\right)y = \lambda YB\tilde{X}y \\
\Leftrightarrow &\tilde{\Lambda}y + \left(YB\tilde{X}\right)^{-1}Y\left(A\tilde{X} - B\tilde{X}\tilde{\Lambda}\right)y = \lambda y \\
\Leftrightarrow &\left(\tilde{\Lambda} + \left(YB\tilde{X}\right)^{-1}Y\left(A\tilde{X} - B\tilde{X}\tilde{\Lambda}\right)\right)y = \lambda y
\end{aligned}$$

□

ゲルシュゴリンの定理 (定理 4.1.1) と補題 4.1.1 を用いると $Ax = \lambda Bx$ に対する全固有値に対する品質保証法を考えることができます:

定理 4.1.2. $A, B \in \mathbb{C}^{n\times n}$ を与えられている行列とする．$\tilde{X},\ Y \in \mathbb{C}^{n\times n}$ 与えられている行列とする．対角行列 $\tilde{\Lambda} \in \mathbb{C}^{n\times n}$ を対角成分

$$\Lambda_{ii} := \tilde{\lambda}_i$$

を持つ対角行列とする．$Q \in \mathbb{R}^{n\times n}$ を

$$Q := (YB\tilde{X})^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})$$

とする．もし $YB\tilde{X}$ が正則ならば,

$$r := |Q|\boldsymbol{e},\ U_i := \{z \in \mathbb{C} \mid\ |z - \tilde{\lambda}_i| \le r_i\}$$

とすると固有値問題 $Ax = \lambda Bx$ のすべての固有値は

$$\bigcup_{i=1}^{n} U_i$$

に含まれる．そのうえ，特に，単連結領域 C が m 個の U_i からなる場合，重複度も含めて m 個の固有値が C 内に存在する．

証明． 補題 4.1.1 から，固有値問題

$$\text{Find } (\mu, y) \in \mathbb{C} \times \mathbb{C}^n \text{ s.t. } \left(\tilde{\Lambda} + Q\right) y = \mu y$$

を考えればよい．そのうえ，上の固有値問題に対し，ゲルシュゴリンの定理を適用すると，

$$q_i := \sum_{j=1, j\neq i}^{n} |Q_{ij}|, \quad \tilde{U}_i := \left\{ z \in \mathbb{C} \mid \left| z - (\tilde{\lambda}_i + Q_{ii}) \right| \leq q_i \right\}$$

としたとき，

$$\bigcup_{i=1}^{n} \tilde{U}_i$$

内に固有値問題 $Ax = \lambda Bx$ のすべての固有値が含まれており，そのうえ，特に，単連結領域 C が m 個の $\tilde{U}_i$ からなる場合，重複度も含めて m 個の固有値が C 内に存在する．

また，

$$\tilde{U}_i = \left\{ z \in \mathbb{C} \mid \left| z - (\tilde{\lambda}_i + Q_{ii}) \right| \leq q_i \right\} \subset \left\{ z \in \mathbb{C} \mid \left| z - \tilde{\lambda}_i \right| \leq |Q_{ii}| + q_i \right\}$$

となり，

$$|Q_{ii}| + q_i = \sum_{j=1}^{n} |Q_{ij}| = (|Q|\boldsymbol{e})_i = r_i$$

となる．よって，

$$\tilde{U}_i \subset U_i$$

となる．

□

この定理 4.1.2 に対して，ノイマン級数の定理 (定理 3.2.1 や定理 5.2.3) を適用すると，次の宮島の定理がいえます:

定理 4.1.3. $A, B \in \mathbb{C}^{n\times n}$ を与えられている行列とする．$\tilde{X},\ Y \in \mathbb{C}^{n\times n}$ 与えられている行列とする．対角行列 $\tilde{\Lambda} \in \mathbb{C}^{n\times n}$ を対角成分

$$\Lambda_{ii} := \tilde{\lambda}_i$$

を持つ対角行列とする．$Q \in \mathbb{R}^{n\times n}$ を

$$Q := (YB\tilde{X})^{-1} Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})$$

とし，行列 $G \in \mathbb{R}^{n\times n}$ を

$$G := I - YB\tilde{X}$$

とする．もし

$$\|G\|_\infty < 1$$

ならば，$YB\tilde{X}$ は正則で，

$$r := |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} + \frac{\|Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})\|_\infty}{1 - \|G\|_\infty} |G|\boldsymbol{e},\ U_i := \{ z \in \mathbb{C} \mid \ |z - \tilde{\lambda}_i| \leq r_i \}$$

とすると固有値問題 $Ax = \lambda Bx$ のすべての固有値は

$$\bigcup_{i=1}^{n} U_i$$

に含まれる．そのうえ，特に，単連結領域 C が m 個の U_i からなる場合，重複度も含めて m 個の固有値が C 内に存在する．

証明． $\|G\|_\infty < 1$ であるから，ノイマン級数の定理より $I - G = YB\tilde{X}$ は正則である．そのうえ，

$$\|(I-G)^{-1}\|_\infty \leq \frac{1}{1-\|G\|_\infty}$$

となる．よって，定理 4.1.2 の $|Q|\boldsymbol{e}$ のみを検討すればよい．

$$\begin{aligned}
|Q|\boldsymbol{e} =& |(YB\tilde{X})^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\
=& |(I-G)^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\
=& |(I+G+G^2+\cdots)Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\
=& |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) + G(I+G+G^2+\cdots)Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\
=& |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) + G(I-G)^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\
\leq& |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} + |G||(I-G)^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\
\leq& |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} + |G| \left\| \left| (I-G)^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} \right\|_\infty \boldsymbol{e} \\
\leq& |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} + |G| \left\|(I-G)^{-1}\right\|_\infty \left\| \left| Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} \right\|_\infty \boldsymbol{e} \\
\leq& |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} + |G| \frac{1}{1-\|G\|_\infty} \left\| Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right\|_\infty \boldsymbol{e} \\
=& |Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} + \frac{\|Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})\|_\infty}{1-\|G\|_\infty}|G|\boldsymbol{e}
\end{aligned}$$

となり，題意は得られた． □

定理 4.1.3 は形式的には，連立一次方程式の解の品質保証法であった山本の定理 (系 3.2.1) の固有値バージョンといえます．著者は，固有値問題でも，このような「うまい式変形」を行うことで，近似固有値とそのエラーみたいな形式に書けることを初めて知った時には感動を覚えました．

4.2 (発展) さらに評価を極めるには?

3.3 節でも取り上げたように，今度は定理 4.1.3 の十分条件 $\|G\|_\infty < 1$ をどこまで緩められるか，定理 4.1.2 まで戻って考えてみたいと思います．まず，対角行列 $D \in \mathbb{R}^{n\times n}$ と非対角行列 $E \in \mathbb{R}^{n\times n}$ をそれぞれ

$$YB\tilde{X} = D + E$$

のように $YB\tilde{X}$ を対角成分と非対角成分に分離する行列とします．さらに Y を

$$Y_{\text{new}} := D^{-1}Y$$

としても，定理 4.1.2 は成立します．そのうえ，

$$G := I - D^{-1}YB\tilde{X}$$

とすると $D^{-1}YB\tilde{X}$ の対角成分はすべて 1 となるため，G の対角成分は 0 となります．

定理 4.1.3 では十分条件 $\|G\|_\infty < 1$ を利用して，ノイマン級数の定理から $YB\tilde{X}$ の正則性を検証していました．この条件を緩めるために，$\rho(G) < 1$ と置き換えたいところですが，3.3 節と同様に，まだ，リーズナブルな評価が見つかっていません．そのために，十分条件を変更し，

$$\rho(G) \leq \rho(|G|) < 1$$

を検証することとします．もちろん，$\rho(|G|) \leq \||G|\|_\infty = \|G\|_\infty$ であるため，$\|G\|_\infty < 1$ よりかは良いことが期待されます．そのうえで，連立一次方程式の場合と同様に次の定理 (Fiedler-Ptaák の定理の系) がいえます:

定理 4.2.1. $Y, B, \tilde{X} \in \mathbb{R}^{n\times n}$ を与えられているとし，I を単位行列とする．対角行列 $D \in \mathbb{R}^{n\times n}$(対角成分以外はすべてゼロ) と非対角行列 $E \in \mathbb{R}^{n\times n}$(対角成分はすべてゼロ) を:

$$YB\tilde{X} = D + E$$

とする．行列 $G \in \mathbb{R}^{n\times n}$ を

$$G := I - D^{-1}YB\tilde{X}$$

とする．行列 $M \in \mathbb{R}^{n\times n}$ を

$$M := I - |G|$$

とする．そのとき，以下は同等である:

1. $\rho(|G|) < 1$
2. M が正則で，$M^{-1} \geq O$
3. $Mv > \mathbf{0}$ となる正のベクトル $v > \mathbf{0}$ が存在

この定理を利用して，次のような定理を導出することができます:

定理 4.2.2. $A, B \in \mathbb{C}^{n\times n}$ を与えられている行列とする．$\tilde{X}$, $Y \in \mathbb{C}^{n\times n}$ 与えられている行列とする．対角行列 $\tilde{\Lambda} \in \mathbb{C}^{n\times n}$ を対角成分

$$\Lambda_{ii} := \tilde{\lambda}_i$$

を持つ対角行列とする．$Q \in \mathbb{R}^{n\times n}$ を

$$Q := (YB\tilde{X})^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})$$

とする．対角行列 $D \in \mathbb{R}^{n\times n}$(対角成分以外はすべてゼロ) と非対角行列 $E \in \mathbb{R}^{n\times n}$(対角成分はすべてゼロ) を:

$$YB\tilde{X} = D + E$$

とする．行列 $G \in \mathbb{R}^{n\times n}$ を

$$G := I - D^{-1}YB\tilde{X}$$

とする．行列 $M \in \mathbb{R}^{n\times n}$ を

$$M := I - |G|$$

とする．もし

$$M\hat{v} > \mathbf{0} \ \text{ かつ } \ \hat{v} > \mathbf{0}$$

となるベクトル $\hat{v}$ が存在するならば，$D^{-1}YB\tilde{X}$ は正則で，ベクトル $u, w \in \mathbb{R}^n$ を

$$\begin{aligned} u :=& M\hat{v} \\ w_k :=& \max_{1\le i\le n} \frac{|G|_{ik}}{u_i} \ \text{ for } \ 1 \le k \le n \end{aligned}$$

とすると，

$$\begin{aligned} r :=& \left| \sum_{i=0}^{N} G^i D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} + |G|^{N+1}(I + \hat{v}w^T)D^{-1}|Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e}, \\ U_i :=& \{z \in \mathbb{C} \mid \ |z - \tilde{\lambda}_i| \le r_i\} \end{aligned}$$

とすると固有値問題 $Ax = \lambda Bx$ のすべての固有値は

$$\bigcup_{i=1}^{n} U_i$$

に含まれる．そのうえ，特に，単連結領域 C が m 個の U_i からなる場合，重複度も含めて m 個の固有値が C 内に存在する．

証明． 十分条件 $M\hat{v} > \mathbf{0}$ かつ $\hat{v} > \mathbf{0}$ となる $\hat{v}$ が存在することから，定理 4.2.1 より $\rho(|G|) < 1$ となる．そのうえ，$\rho(G) \le \rho(|G|) < 1$ から，ノイマン級数の定理より $D^{-1}YB\tilde{X}$ は正則である．さらに，3.3 節の議論より

$$O \le M^{-1} \le I + \hat{v}w^T$$

となる．そのうえで，定理 4.1.2 の $|Q|\boldsymbol{e}$ のみを検討すると

$$\begin{aligned} |Q|\boldsymbol{e} =& |(D^{-1}YB\tilde{X})^{-1}D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\ =& |(I-G)^{-1}D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\ =& |(I + G + G^2 + \cdots)D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda})|\boldsymbol{e} \\ =& \left| \left(\sum_{i=0}^{N} G^i \right) D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) + G^{N+1}(I + G + G^2 + \cdots)D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} \\ \le& \left| \sum_{i=0}^{N} G^i D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} + \left|G^{N+1}\right| \left|(I + G + G^2 + \cdots)\right| \left| D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} \\ \le& \left| \sum_{i=0}^{N} G^i D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} + \left|G^{N+1}\right| (I + |G| + |G|^2 + \cdots) \left| D^{-1}Y(A\tilde{X} - B\tilde{X}\tilde{\Lambda}) \right| \boldsymbol{e} \end{aligned}$$

となる．$\rho(|G|) < 1$ からノイマン級数の定理より $I + |G| + |G|^2 + \cdots = (I - |G|)^{-1} = M^{-1}$ となるため

$$\begin{aligned}
|Q|\boldsymbol{e} &\leq \left|\sum_{i=0}^{N} G^i D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} + \left|G^{N+1}\right| M^{-1} \left|D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} \\
&\leq \left|\sum_{i=0}^{N} G^i D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} + \left|G^{N+1}\right| (I + \hat{v} w^T) \left|D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e}
\end{aligned}$$

□

上記の定理をそのまま適用すると，$\sum_{i=0}^{N} G^i$ で $O(n^3)$ の計算がたくさんかかってしまいます．そのために，$\rho(|G|) < 1$ であることを利用して，

$$\begin{aligned}
|Q|\boldsymbol{e} &\leq \left|\sum_{i=0}^{N} G^i D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} + \left|G^{N+1}\right| (I + \hat{v} w^T) \left|D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} \\
&\leq \sum_{i=0}^{N} |G^i| \left|D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} + \left|G^{N+1}\right| (I + \hat{v} w^T) \left|D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} \\
&\leq \sum_{i=0}^{N+1} |G|^i \left|D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e} + |G|^{N+1} \hat{v} w^T \left|D^{-1} Y (A\tilde{X} - B\tilde{X}\tilde{\Lambda})\right| \boldsymbol{e}
\end{aligned}$$

として計算することをお勧めします．

第5章　方程式の品質保証のための基本定理

5.1　Banach 空間

Banach 空間を知らないという場合，実数全体の集合 $\mathbb{R}$ や複素数全体の集合 $\mathbb{C}$ の直積空間 $\mathbb{R}^N$ や $\mathbb{C}^N$，はたまた $\mathbb{R}$ あるいは $\mathbb{C}$ を表す記号 $\mathbb{K}$ と読み替えても，ほとんどの章は差し支えありません．では，なぜ Banach 空間で紹介するのか?という疑問が生まれると思います．偏微分方程式の解は一般的に無限次元となり，$\mathbb{R}^N$ では表せません．このような問題を扱う際には無限次元まで扱える Banach 空間が必要になります．$\mathbb{R}^N$ も Banach 空間であるため，連立一次方程式の解の品質保証を行いたい場合は $\mathbb{R}^N$ で十分です．

5.1.1　Banach 空間の定義と性質

みなさんは有理数 $\mathbb{Q}$ と実数 $\mathbb{R}$ の違いを明確に答えられますか? 誤解を恐れずにいうと「有理数に抜け目のないように穴を埋めたものが実数」といえます．すなわち，実数の上を歩いたとしても落とし穴にはまることがない集合，完備な集合といえます．実数の完備性は色々な出発点と定理があるが，その中に「Cauchy 列ならば収束列」であることを示す出発点あるいは定理があったことを思い出しておくと，すんなりとこの節を読むことができると思います．

この節では，実数に限らず一般的な線形空間でも，抜け目がない集合を定義することが目的です．この抜け目のない線形空間は Banach 空間と呼ばれます．非常に抽象的な定義で難しいと感じるかもしれませんが，実数の完備性を思い出すと，実は一般化しただけの自然にも思える定義に感じれると思います．

では，線形空間がないと話が進みませんので，線形空間の公理から始めましょう．

公理 5.1.1 (線形空間の公理)**.** 空でない集合 X が係数体 $\mathbb{K}$ 上の線形空間であるとは，任意の $u, v \in X$ とスカラー $\alpha \in \mathbb{K}$ に対して，加法 $u+v \in X$ とスカラー乗法 $\alpha u \in X$ が定義されていて，任意の $u, v, w \in X$ とスカラー $\alpha, \beta \in \mathbb{K}$ に対して次の (i)-(viii) が成り立つことである．
(i) $(u+v)+w = u+(v+w)$
(ii) $u+v = v+u$
(iii) $u+0 = u$ となる $0 \in X$ が一意に存在
(iv) $u+(-u) = 0$ となる $-u \in X$ が一意に存在
(v) $\alpha(u+v) = \alpha u + \alpha v$
(vi) $(\alpha+\beta)u = \alpha u + \beta u$
(vii) $(\alpha\beta)u = \alpha(\beta u)$
(viii) $1u = u,\ 1 \in \mathbb{K}$

再び，実数の完備性を思い出してみましょう．実数の場合は，真っ先に $\varepsilon - N$ 論法を使って数列の収束を勉強したと思います．この際に，自然と絶対値記号を使っていたと思います．しかし，

一般的な線形空間には絶対値記号はありません．例えば，$\mathbb{C}^2$ の元

$$\begin{pmatrix} 2+2i \\ -1+3i \end{pmatrix}$$

に対する「絶対値に相当するもの」は何でしょうか? ここで重要なのは，あくまで一般的な線形空間 X の点列に対して $\varepsilon - N$ 論法を使って収束を定義したいことに着目すると「絶対値に相当するもの」を $\mathbb{C}^2$ の元につけた結果が $\mathbb{C}^2$ では $\varepsilon - N$ 論法の ε がよくわからない状態になってしまいますよね．そのために，どんな線形空間 X の元でも「絶対値に相当するもの」をつけた結果は実数 $\mathbb{R}$ になるように定義します．もちろん,「絶対値に相当するもの」であるため，絶対値の性質も含めて定義する必要もあります．これらの点に留意しながら次の「絶対値に相当するもの」の定義，すなわちノルムの定義を見てみましょう．

定義 5.1.1 (ノルムとノルム空間の定義)**.** X を係数体 $\mathbb{K}$ 上の線形空間とする．X で定義された関数 $\|\cdot\| : X \to \mathbb{R}$ が X のノルムであるとは
(i) $\|u\| \geq 0,\ u \in X$
(ii) $\|u\| = 0 \Leftrightarrow u = 0$
(iii) $\|\alpha u\| = |\alpha|\|u\|,\ (\alpha \in \mathbb{K},\ u \in X)$
(iv) $\|u+v\| \leq \|u\| + \|v\|$
が成立することである．さらに X に一つのノルムが指定されているとき，X はノルム空間という．

これで，線形空間にノルムという絶対値に相当する道具が追加されたノルム空間が定義できました．まだ，ノルム空間が X しか出てきていないので $\|\cdot\|$ でも問題ありませんが，ノルム空間が多数出てきた場合，$\|\cdot\|_X$ のように右下にどのノルム空間のノルムなのか明記します．続いてノルムを使って，点列の収束を定義してみましょう．

定義 5.1.2 (ノルム空間の収束と極限)**.** X をノルム空間とする．X の点列 $(u_n) \subset X$ は

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N},\ \forall n \geq N \text{ に対して } \|u_n - u\| < \varepsilon$$

のとき，点 $u \in X$ に収束するといい,

$$\|u_n - u\| \to 0,\ (n \to \infty)$$

と表す．このとき，u を u_n の極限といい

$$u_n \to u,\ (n \to \infty)$$

と表す.

定理 5.1.1 (極限の一意性)**.** X の点列 $(u_n) \subset X$ が収束するならば，その極限は一意である．

証明 . 極限を v と w の 2 つ持つと仮定して矛盾を示す．つまり，v と w が極限であることから,

$$\forall \varepsilon > 0,\ \exists N_1 \in \mathbb{N},\ \forall n \geq N_1 \text{ に対して } \|u_n - v\| < \varepsilon$$

と

$$\forall \varepsilon > 0,\ \exists N_1 \in \mathbb{N},\ \forall n \geq N_2 \text{ に対して } \|u_n - w\| < \varepsilon$$

が成立する．そのうえで，ϵ は任意であることから，

$$\epsilon = \frac{\|v-w\|}{2}$$

としたとき，上の論理式を満たす N_1 と N_2 がそれぞれ存在する．すなわち，

$$\exists N_1 \in \mathbb{N}\ \forall n \geq N_1 \text{ に対して } \|u_n - v\| < \frac{\|v-w\|}{2}$$

と

$$\exists N_1 \in \mathbb{N}\ \forall n \geq N_2 \text{ に対して } \|u_n - w\| < \frac{\|v-w\|}{2}$$

となる．そのうえで，$N_{\max} = \max(N_1, N_2)$ とすると，

$$\forall n \geq N_{\max} \text{ に対して } \|u_n - v\| < \frac{\|v-w\|}{2} \text{ および } \|u_n - w\| < \frac{\|v-w\|}{2}$$

となる．よって，任意の $n \geq N_{\max}$ に対して，

$$\|v-w\| \leq \|v-u_n\| + \|u_n - w\| < \frac{\|v-w\|}{2} + \frac{\|v-w\|}{2} = \|v-w\|$$

となり，$\|v-w\| < \|v-w\|$ から，矛盾する．

□

同様にノルム空間上の Cauchy 列もノルムを用いて次のように定義できます．

定義 5.1.3 (Cauchy 列)**.** X をノルム空間とする．そのとき X の点列 (u_n) が Cauchy 列であるとは

$$u_n - u_m \to 0,\ (n, m \to \infty)$$

が成立することである．即ち

$$\|u_n - u_m\| \to 0,\ (n, m \to \infty)$$

が成立することである．

ちなみに $\varepsilon - N$ 論法を用いてちゃんと書くと，X の点列 (u_n) が

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N},\ \forall n, m \geq N \text{ に対して } \|u_n - u_m\| < \varepsilon$$

を満たすときに点列 (u_n) が Cauchy 列であるといいます．

点列の収束と Cauchy 列は一見すると似ているように見えますが，明確に意図の違いがあります．点列の収束の定義は，X の点列 (u_n) の極限 u が X に属していることが既にわかっています．しかし，抽象的なノルム空間 X の点列 (u_n) の極限が X に属しているか，それとも属していないかなんてわかりません．そのために，極限を使わずに収束してそうな点列として Cauchy 列が定義されたと思っておいてください．もちろん，ノルム空間 X の収束する点列 (u_n) ならば，収束しそうな点列である Cauchy 列になることは簡単にいえます．実際に，収束する点列 $(u_n) \subset X$ は極限 $u \in X$ を持ちます．その上，

$$\begin{aligned} &\|u_n - u_m\| = \|u_n - u + u - u_m\| \\ \leq\ &\|u_n - u\| + \|u - u_m\| \to 0,\ (n, m \to \infty). \end{aligned}$$

から，収束する点列 $(u_n) \subset X$ ならば，Cauchy 列になります．

Cauchy 列が収束しそうな点列であることを踏まえたうえで，次の完備に関する定義を見てみましょう．

定義 5.1.4 (完備)**.** X をノルム空間とする．X が完備であるとは，任意の Cauchy 列 (u_n) が X の中で極限を持つことである．すなわち，任意の Cauchy 列 $(u_n) \subset X$ が

$$\|u_n - u\| \to 0,\ (n \to \infty)$$

となる極限 u を X 内に持つことである．

ノルム空間 X が完備であるとき，どんな Cauchy 列 (u_n) でも，ちゃんと極限 u を X 内に持つことを保証してくれています．この節の目的は「抜け目のない線形空間」を定義することでしたね．ノルム空間が完備であれば抜け目のない線形空間の出来上がりです．すなわち，Banach 空間の定義は次のようになります．

定義 5.1.5 (Banach 空間)**.** ノルム空間 X が Banach 空間であるとは，X が完備であることである．

この節の最後に，後々必要になる Cauchy 列の性質を示しておきます．まず，Cauchy 列の性質を示すために必要になるノルムの性質として逆三角不等式を示します:

定理 5.1.2 (逆三角不等式)**.** X をノルム空間とする．任意の $u, v \in X$ について次の不等式を満たす:

$$|\|u\| - \|v\|| \leq \|u - v\|$$

証明 . 任意の $u, v \in X$ について

$$\begin{aligned} \|u\| &= \|u - v + v\| \leq \|u - v\| + \|v\| \\ \|v\| &= \|v - u + u\| \leq \|v - u\| + \|u\| = \|u - v\| + \|u\| \end{aligned}$$

となる．よって

$$\begin{aligned} \|u\| - \|v\| &\leq \|u - v\| \\ \|v\| - \|u\| &\leq \|u - v\| \end{aligned}$$

となるため,

$$|\|u\| - \|v\|| \leq \|u - v\|$$

を持つ.

□

続いて，Cauchy 列とは別の点列の性質を定義します:

定義 5.1.6 (有界列)**.** X をノルム空間とする．そのとき X の点列 (u_n) が有界列とは任意の $n \in \mathbb{N}$ に対して

$$\|u_n\| \leq M$$

となる定数 $M > 0$ が存在することである．

有界列のポイントは点列 (u_n) に対して，1 つの M が存在することです．すなわち，有界列なら，どんな u_n でも共通の M を用いて $\|u_n\| \leq M$ という不等式を満たします．

定理 5.1.3 (Cauchy 列ならば有界列)**.** X をノルム空間とする．そのとき X の点列 (u_n) が Cauchy 列ならば有界列でもある．

証明 . X の点列 (u_n) が Cauchy 列であるため，$\varepsilon - N$ 論法を用いた表記で

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N},\ \forall n, m \geq N \text{ に対して } \|u_n - u_m\| < \varepsilon$$

を満たす．$\varepsilon = 1$ としても，それに対応した N が存在し，任意の $n \geq N$ に対して

$$\|u_n - u_N\| < 1$$

を満たす．(ここで m を N に固定したことに注意．)

任意の $n \geq N$ に対して $\|u_n\|$ が $\|u_N\|$ で評価できることを示す．逆三角不等式である定理 5.1.2 を用いると

$$|\|u_n\| - \|u_N\|| \leq \|u_n - u_N\| < 1$$

となる．絶対値の性質より $|\|u_n\| - \|u_N\|| < 1$ は

$$\|u_N\| - 1 < \|u_n\| < \|u_N\| + 1$$

となる．よって

$$M = \max\{\|u_1\|,\ \|u_2\|,\ \cdots,\ \|u_{N-1}\|,\ \|u_N\| + 1\}$$

とすると，任意の $n \in \mathbb{N}$ について

$$\|u_n\| \leq M$$

が成り立つため，点列 (u_n) は有界列である．

□

5.1.2 ノルム空間の位相構造と Banach 空間の閉部分空間

まず，線形代数の復習である線形部分空間を定義しましょう．

定義 5.1.7 (線形部分空間)**.** 線形空間 X の空ではない集合 M が任意の元 $u, v \in M$ と任意の係数体 $\alpha \in \mathbb{K}$ に対して

$$u + v \in M$$
$$\alpha u \in M$$

を満たすとき，M は線形空間 X の線形部分空間と呼ぶ．

定義 5.1.7 の線形部分空間は線形代数で習っている通り，M も X の演算のもとに線形空間となります．では，Banach 空間の線形部分空間は，必ず Banach 空間になるでしょうか? 答えは残念ながら No です．Banach 空間の線形部分空間は Banach 空間にはならず，完備性がないノルム空間になってしまう場合もあります．では，Banach 空間の線形部分空間が Banach 空間になるための条件は何でしょうか? 本節では，この条件を紹介することが目標です．その際に定義するノルム空間の位相構造も，距離空間における集合と位相の復習となり，本節以外にも出てくるので覚えておきましょう．まず，ノルム空間における開球，開集合，閉集合を定義します．

定義 5.1.8 (ノルム空間の開球)**.** X をノルム空間とする．$x \in X$ とし，$r > 0$ を正実数とする．そのとき，集合

$$B_X(x, r) := \{y \in X \mid \|x - y\|_X < r\}$$

を中心 x，半径 r の開球という．X が明らかな場合は $B_X(x, r)$ を省略して $B(x, r)$ と表記する．

定義 5.1.9 (ノルム空間の開集合)**.** X をノルム空間し，M を X の部分集合とする．任意の $x \in M$ に対して，$U(x, r) \subset M$ となる $r > 0$ が存在する場合，M が開集合であるという．

定義 5.1.10 (ノルム空間の閉集合)**.** X をノルム空間とし，M を X の部分集合とする．M が閉集合であるとは，M の任意の点列 (u_n) の極限 $u \in X$ が M にも属することである．すなわち，点列 $(u_n) \subset M$ について

$$u_n \to u,\ (n \to \infty) \Rightarrow u \in M$$

であるとき，M は閉集合であるという．

集合と位相で習った講義で習った一般的な位相空間 X の部分集合 M の閉集合の定義は補集合

$$X^c := \{x \in X \mid x \notin M\}$$

としたとき，X^c が開集合になることでしたね．ノルム空間の場合は,「X^c が開集合」と「定義 5.1.10」は同値になります．そのため，本書では，良く使う定義 5.1.10 を閉集合の定義としました．もちろん,「X^c が開集合」を閉集合の定義として，定理として定義 5.1.10 と同値であることを導く方針もあります．

それでは本節の目標である「Banach 空間の線形部分空間が Banach 空間になるための条件」を考えていきます．先に答えを言ってしまうと,「Banach 空間の線形部分空間が Banach 空間になるための条件」は次に定義する「閉部分空間」と同値になります．では閉部分空間について定義しましょう．

定義 5.1.11 (閉部分空間)**.** X をノルム空間とし，M を X の線形部分空間が閉集合であるとき，M を閉部分空間であるという．

端的に言えば,「線形部分空間」かつ「閉集合」のとき,「閉部分空間」と呼ばれます．さて，いよいよ,「Banach 空間の線形部分空間が Banach 空間になるための条件」を示しましょう．

定理 5.1.4 (線形部分空間と Banach 空間の関係)**.** X を Banach 空間とし，M を X の線形部分空間とする．そのとき,

$$M \text{ が } X \text{ の閉部分空間} \ \Leftrightarrow\ M \text{ が } X \text{ のノルムで Banach 空間}$$

証明 .「M が X の閉部分空間 $\Rightarrow$ M が X のノルムで Banach 空間 の証明」

点列 (u_n) を M の任意の Cauchy 列とする．M は Banach 空間 X の部分集合であることから，Banach 空間 X の完備性 (定義 5.1.4 と定義 5.1.5) より，M の任意の Cauchy 列 (u_n) の極限 u が X 内に存在する．さらに，M が X の閉部分空間であることから，閉集合の定義 5.1.10 より任意の点列の極限は M に属することから，Cauchy 列 (u_n) の極限 u は M にも属する．よって，ノルム空間 M の任意の Cauchy 列 (u_n) の極限 u が M に属するため，M は Banach 空間となる．

「M が X の閉部分空間 $\Leftarrow$ M が X のノルムで Banach 空間 の証明」

M の任意の点列を (u_n) とすると，$u_n \to u, (n \to \infty)$ となる極限 $u \in X$ が M にも属することを示せれば，定義 5.1.9 より M が閉集合であることが示せる．点列 (u_n) は X の Cauchy 列でもあり，X が Banach 空間であることから完備性より X 内に極限 $v \in X$ を持つ．また，同様に，点列 (u_n) は M の Cauchy 列でもあること，および定理の仮定から M が Banach 空間であることから完備性より M 内に極限 $u \in M$ を持つ．そのうえで，$M \subset X$ であることと，極限の一意性 (定理 5.1.1) より $u = v \in M \subset X$ となり，M の任意の点列を (u_n) の極限 u は M に属する．

□

5.2 作用素の基礎

本章では作用素というものを導入します．作用素とは写像の一般化にあたる概念です．そのため，行列も作用素に含まれますので，ピンと来ない場合は行列として考えてみるといいかもしれません．ただし，写像の一般化であるため，写像と作用素の違いに注意しながら読むことをお勧めします．本章の目標の一つは精度保証付き数値計算の証明で良く利用される Neumann 級数に関する定理を紹介することです．

5.2.1 作用素とは?写像との違いに注意しながら...

まず，作用素を定義しましょう:

定義 5.2.1 (作用素)**.** ある線形空間 X から別の線形空間 Y への作用素 A とは,

$$\mathcal{D}(A) := \{u \in X \mid Au \in Y\}$$

としたとき，$\mathcal{D}(A)$ のどんな元に対しても，それぞれ集合 Y のただ一つの元を指定する規則のことである．また，$\mathcal{D}(A)$ は A の定義域と呼ばれ

$$\mathcal{R}(A) := \{Au \in Y \mid u \in \mathcal{D}(A)\}$$

を値域と呼ぶ.

作用素は一見すると写像と同じものに感じますが，作用素は写像の一般化であることに注意が必要です．実際に，$\mathcal{D}(A) \subsetneq X$ の場合，上記の作用素 A は X から Y への写像ではありません．また，$\mathcal{D}(A) = X$ の場合にのみ，作用素 A は X から Y への写像となります．そのため，写像と作用素を同じものと思って取り扱うと引っかかってしまうことが多々あるので注意が必要です．写像と作用素の違う例の一つが逆作用素です．写像の場合は単射かつ全射ならば逆写像を持ちますが，作用素の場合は定義が変わってしまいます.

まず，作用素における単射と全射の定義を見てみましょう:

定義 5.2.2 (単射)**.** 線形空間 X から線形空間 Y への作用素 A が

$$u_1 \neq u_2,\ \forall u_1, u_2 \in \mathcal{D}(A) \Rightarrow A(u_1) \neq A(u_2)$$

を満たすときに作用素 A は単射であるという.

定義 5.2.3 (全射)**.** 線形空間 X から線形空間 Y への作用素 A が

$$Y = \mathcal{R}(A)$$

を満たすときに作用素 A は全射であるという.

単射と全射の定義は作用素の特性である X と定義域 $\mathcal{D}(A)$ が違うこと以外は写像と同じです．それに対し，逆作用素の定義は次のようになります:

定義 5.2.4 (逆作用素)**.** 線形空間 X から線形空間 Y への作用素 A とし，その定義域を $\mathcal{D}(A) \subset X$，値域を $\mathcal{R}(A) \subset Y$ とする．そのとき,

$$A^{-1}(A(u)) = u,\ u \in \mathcal{D}(A)$$
$$A(A^{-1}(v)) = v,\ v \in \mathcal{R}(A)$$

かつ

$$\mathcal{D}(A^{-1}) = \mathcal{R}(A)$$
$$\mathcal{R}(A^{-1}) = \mathcal{D}(A)$$

となる Y から X への作用素 A^{-1} を A の逆作用素と呼ぶ.

この逆作用素の定義のもとで，次の定理をみると逆作用素と逆写像の違いが明確になると思います:

定理 5.2.1 (単射と逆作用素の関係)**.** 線形空間 X から線形空間 Y への作用素 A とすると,

$$A \text{が逆作用素を持つ} \Leftrightarrow A \text{が単射である}$$

証明 . 「A が逆作用素を持つ $\Rightarrow$ A が単射である」の証明
単射の定義 5.2.2 の対偶「任意の $u_1, u_2 \in \mathcal{D}(A)$ に対し $A(u_1) = A(u_2) \Rightarrow u_1 = u_2$」を満たすことを確かめる．$A$ の逆作用素を A^{-1} とすると，任意の $u_1, u_2 \in \mathcal{D}(A)$ に対し

$$\begin{aligned} & A(u_1) = A(u_2) \\ \Rightarrow \quad & A^{-1}(A(u_1)) = A^{-1}(A(u_2)) \\ \Rightarrow \quad & u_1 = u_2 \end{aligned}$$

となる．最後の変形は逆作用素 A^{-1} の定義に由来する.

「A が単射である $\Rightarrow$ A が逆作用素 A^{-1} を持つ」の証明
A の値域の定義 $\mathcal{R}(A) = \{A(u) \in Y \mid u \in \mathcal{D}(A)\}$ より，任意の $v \in \mathcal{R}(A)$ に対し

$$A(u) = v$$

となる $u \in \mathcal{D}(A)$ が存在する．その上，A が単射であるため，単射の定義 5.2.2 の対偶「任意の $u_1, u_2 \in \mathcal{D}(A)$ に対し $A(u_1) = A(u_2) \Rightarrow u_1 = u_2$」より，$u \in \mathcal{D}(A)$ はどんな $v \in \mathcal{R}(A)$ に対してもただ一つの元である．そのため，作用素の定義 5.2.1 より，上記の $v \in \mathcal{R}(A)$ に対してただ一つの元 $u \in \mathcal{D}(A)$ を指定する規則として

$$B(v) = u$$

となる定義域 $\mathcal{D}(B) = \mathcal{R}(A)$ と値域 $\mathcal{R}(B) = \mathcal{D}(A)$ となる Y から X への作用素 B が定義できる．その上，$B(v) = u$ の v に $v = A(u)$ を代入すると

$$B(A(u)) = u$$

となる．同様に，$A(u) = v$ の u に $u = B(v)$ を代入すると

$$A(B(v)) = v$$

となる．よって，定義域 $\mathcal{D}(B) = \mathcal{R}(A)$ と値域 $\mathcal{R}(B) = \mathcal{D}(A)$ となる Y から X への作用素 B は A の逆作用素であるため，A は逆作用素を持つ.

□

写像の場合，逆写像を持つ必要十分条件は単射かつ全射でした．それに対し，逆作用素を持つ必要十分条件には作用素 A が全射であることを必要としません．なぜなら，逆作用素 A^{-1} は作用素であるため定義域 $\mathcal{D}(A^{-1})$ と Y が一致しなくてもよいからです．これは，写像と作用素が明確に違うことを表しています．そのために，1 から作用素を扱うには，基礎から丁寧に定義しなければなりません．

では，なぜ，苦労してまで写像を捨てて，X と $\mathcal{D}(A)$ が別になる作用素を利用するのか疑問に思うと思います．もちろん，苦労する分，うまみが出てくるから作用素を導入します．X と $\mathcal{D}(A)$ を別にする理由は固有値やその一般化であるスペクトルにあります．実際に，線形代数で習った行列の固有値問題は行列 A は正方行列であったと思います．正方行列ということは行列 A は $X = \mathcal{D}(A) = \mathbb{R}^N$ から $Y = \mathbb{R}^N$ への作用素になります．そのため，線形代数で習った固有値問題は $X = \mathcal{D}(A) = Y$ となる状況です．しかし，単純な 2 階微分作用素 d^2/dx^2 を考えてみると，$\mathcal{D}(A)$ には 2 階以上微分できる関数である必要があるのに対し，Y の元は微分ができる必要はありません．では，このような 2 階微分作用素 d^2/dx^2 に対して固有値問題を考えたい場合，どうすればよいでしょうか? 写像のままで進めようとすると $X = \mathcal{D}(A) \neq Y$ で写像 A の固有値を定義する必要があります．これがまかり通ると，正方行列以外でも固有値が定義できていることになってしまい，線形代数の固有値問題とは別ものになってしまいます．このような状況で作用素のうまみがでてきます．$X = Y$ を連続関数全体にし，$\mathcal{D}(A)$ のみ 2 階微分可能な関数空間に設定すれば，今までの線形代数の固有値問題を壊さずに定義していくことができます．行列の場合は，たまたま $X = \mathcal{D}(A)$ になったという解釈です．

では，次に作用素 A と B が等しい (すなわち $A = B$) とはどういうことか，ちゃんと定義をしましょう．

定義 5.2.5 (作用素の等号)**.** 線形空間 X から線形空間 Y への作用素 A と B が等しいとは

$$\mathcal{D}(A) = \mathcal{D}(B)$$

かつ

$$Au = Bu,\ \forall u \in \mathcal{D}(A) = \mathcal{D}(B)$$

が成立することであり，

$$A = B$$

と表記する．

定義をみてもわかるとおり，定義域も一致していることが重要です．

この節の最後に作用素の連続性について定義します．作用素の連続性については，本書籍ではあまり利用しないので簡単に留めておく程度で十分です．

定義 5.2.6 (作用素の連続)**.** ノルム空間 X からノルム空間 Y への作用素 A が $u \in \mathcal{D}(A)$ で連続であるとは

$$u_n \to u,\ (n \to \infty)$$

となる任意の $u_n \in \mathcal{D}(A) \subset X$ に対して

$$Au_n \to Au,\ (n \to \infty)$$

を満たすときである．さらに，A が任意の $u \in \mathcal{D}(A)$ において連続であるとき，A は連続であるという．

上記の作用素 A の $u \in \mathcal{D}(A)$ における連続と同値になる $\varepsilon - \delta$ 論法を用いた定義は次のようになります:

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \|u_n - u\|_X < \delta \text{ となる } \forall u_n \in \mathcal{D}(A) \text{ に対して } \|Au_n - Au\|_Y < \varepsilon$$

「作用素 A は $u \in \mathcal{D}(A)$ で連続」と「作用素 A は連続」は連続である点が，$u \in \mathcal{D}(A)$ のただ 1 点における連続と，定義域 $\mathcal{D}(A)$ の任意の元で連続という大きな違いがあるので注意して下さい．

連続な作用素も作用素の一つのクラスになります．作用素はあまりにも広すぎる定義であるため，クラス分けして,「このクラスの作用素ならば，こんな性質を持ちます」といった感じで解析していきます．

5.2.2 線形作用素

本節では最も標準的な作用素のクラスである線形作用素を紹介します．線形作用素では作用素同士の加法とスカラー乗法なども定義できます．では，線形作用素の定義を見てみましょう:

定義 5.2.7 (線形作用素)**.** 線形空間 X から線形空間 Y への作用素 A が，任意の $u, v \in \mathcal{D}(A) \subset X$ と $\alpha \in \mathbb{K}$ に対し，

$$\begin{aligned} &\mathcal{D}(A) \text{ が } X \text{ の線形部分空間} \\ &A(u+v) = Au + Av \\ &A(\alpha u) = \alpha Au \end{aligned}$$

を満たすとき，A を線形作用素と呼ぶ．

線形作用素の定義は線形写像の定義を作用素に置き換えたものです．もちろん $\mathcal{D}(A) \subset X$ であることに注意してください．同じように線形作用素に対する加法，スカラー乗法，合成作用素を定義していきましょう．これらを定義する際は，特に定義域に注意しましょう．

定義 5.2.8 (線形作用素の加法)**.** 線形空間 X から線形空間 Y への線形作用素 A と B の和を

$$(A+B)u := Au + Bu,\ u \in \mathcal{D}(A) \cap \mathcal{D}(B)$$

と定義する．このとき，X から Y への作用素 $A+B$ の定義域は

$$\mathcal{D}(A+B) = \mathcal{D}(A) \cap \mathcal{D}(B)$$

とする．

定義 5.2.9 (線形作用素のスカラー乗法)**.** 線形空間 X から線形空間 Y への線形作用素 A の $\alpha \in \mathbb{K}$ によるスカラー倍を

$$(\alpha A)u := \alpha(Au),\ u \in \mathcal{D}(A)$$

と定義する．このとき，X から Y への作用素 αA の定義域は

$$\mathcal{D}(\alpha A) := \mathcal{D}(A)$$

とする．

定義 5.2.10 (合成作用素)**.** X, Y, Z を線形空間とする．A を Y から Z への線形作用素とし，B を X から Y への線形作用素とする．そのとき，A と B の合成作用素 AB は

$$(AB)u := A(Bu),\ u \in \{v \in \mathcal{D}(B) \mid Bv \in \mathcal{D}(A)\}$$

と定義する．このとき，X から Z への合成作用素 AB の定義域は

$$\mathcal{D}(AB) := \{v \in \mathcal{D}(B) \mid Bv \in \mathcal{D}(A)\}$$

とする．

線形作用素同士の合成作用素は線形作用素になるとは限らないので注意してください．なぜなら，$\mathcal{D}(B)$ が X の線形部分空間でも，$\mathcal{D}(AB)$ が X の線形部分空間になるとは限らないからです．$\mathcal{D}(AB)$ が X の線形部分空間であれば，合成作用素 AB は線形作用素になります．

線形作用素の場合，単射性をチェックするための定理がいくつかあります．線形作用素が単射であることを確かめる際には，単射の定義をチェックするよりも楽な場合が多いので紹介します:

定理 5.2.2 (線形作用素に対する単射性 (1))**.** 線形空間 X から線形空間 Y への線形作用素 A において以下は同値である:
i) 線形作用素 A が単射である
ii) $Au = 0,\ u \in \mathcal{D}(A)\ \Rightarrow\ u = 0$

証明 . 単射の定義の対偶は

$$Au_1 = Au_2,\ \forall u_1, u_2 \in \mathcal{D}(A) \Rightarrow u_1 = u_2$$

となる．その上，A は線形作用素であるため

$$Au_1 = Au_2\ \Leftrightarrow\ A(u_1 - u_2) = 0$$

となる．$u_1 - u_2 \in \mathcal{D}(A)$ を $u \in \mathcal{D}(A)$ とおきなおせば，(i) $\Rightarrow$ (ii) は証明された．また，証明を逆に追うことで，(ii) $\Rightarrow$ (i) も示せる．

□

定理 5.2.3 (線形作用素に対する単射性 (2))**.** ノルム空間 X からノルム空間 Y への線形作用素 A とする．不等式

$$\|u\|_X \le K\|Au\|_Y,\ u \in \mathcal{D}(A)$$

を満たす定数 $K > 0$ が存在するならば，線形作用素 A は単射である．

証明 . A が線形作用素であるため，定理 5.2.2 ii) を使って証明する．ノルムの定義より

$$Au = 0,\ \forall u \in \mathcal{D}(A) \Leftrightarrow \|Au\|_Y = 0$$

となる．さらに，$Au = 0$ ならば,

$$\|u\|_X \le K\|Au\|_Y = 0,\ u \in \mathcal{D}(A)$$

より $\|u\|_X = 0$ となる．よって，再びノルムの定義より

$$\|u\|_X = 0,\ \forall u \in \mathcal{D}(A) \Leftrightarrow u = 0$$

より，$Au = 0$ ならば $u = 0$ となる．

□

定理 5.2.3 と逆作用素の定義より，線形作用素 A に対して不等式

$$\|u\|_X \leq K\|Au\|_Y,\ u \in \mathcal{D}(A)$$

を満たす定数 $K > 0$ が存在するならば，線形作用素 A は Y から X への逆作用素 A^{-1} を持ちます．但し，A^{-1} の定義域 $\mathcal{D}(A^{-1}) = \mathcal{R}(A)$ は Y と一致するかどうかはわかりません．もし，A が全射でもあるならば，もちろん A^{-1} の定義域 $\mathcal{D}(A^{-1}) = \mathcal{R}(A)$ と Y は一致します．

5.2.3　有界な線形作用素

前節では作用素に「線形」という条件を追加しました．本節では，作用素に「線形」かつ「有界」という条件を追加した場合の作用素のクラスの性質を紹介します．では，まず，有界な線形作用素を定義しましょう:

定義 5.2.11 (有界な線形作用素)**.** *ノルム空間 X から Y への線形作用素 A に対し，*

$$\|Au\|_Y \leq K\|u\|_X,\ u \in \mathcal{D}(A)$$

を満たす正の定数 K が存在するとき，線形作用素 A を有界な作用素と呼ぶ.

有界な線形作用素の定義は非常に悩ましく，$X = \mathcal{D}(A)$ の場合にのみ限り，有界な線形作用素と呼ぶ文献も数多くあります．もちろん $X = \mathcal{D}(A)$ に限らない定義の文献もあります．しかし，本書籍では $X \neq \mathcal{D}(A)$ でも次の定理が成り立つことから，有界な線形作用素の定義には $X = \mathcal{D}(A)$ は入れないので注意しでください.

定理 5.2.4 (有界な線形作用素と連続な線形作用素)**.** *ノルム空間 X からノルム空間 Y への線形作用素 A に対し，*

$$A \text{ が有界 } \Leftrightarrow \ A \text{ が連続}$$

証明 . 「A が有界 $\Rightarrow$ A が連続」の証明
連続性の定義より，$u_n \to u$ となる任意の $u_n \in \mathcal{D}(A)$ に対して $Au_n \to Au$ となることを確かめる．$u_n \to u$ となる任意の $u_n \in \mathcal{D}(A)$ から $\|u_n - u\|_X \to 0,\ (n \to \infty)$ を持つ．その上，A は有界であることから

$$\|Au_n - Au\|_Y \leq M\|u_n - u\|_X \to 0,\ (n \to \infty)$$

となる．よって，$u_n \to u,\ (n \to \infty)$ ならば，$Au_n \to Au$ であるため，A は連続である．

「A が連続 $\Rightarrow$ A が有界」の証明
背理法によって証明する．すなわち，A が連続ならば，任意の $M_2 > 0$ に対して

$$\|Au\|_Y > M_2\|u\|_X$$

を満たす $u \in \mathcal{D}(A)$ が存在すると仮定して矛盾をみつける．この仮定より自然数 n に対して，

$$\|Au_n\|_Y > n\|u_n\|_X$$

を満たす $u_n \in \mathcal{D}(A)$ が存在する．このとき，$\|u_n\|_X \neq 0$ であることに注意する．ノルム空間 X はノルム空間の定義より線形空間であるため，ゼロ元 $0 \in X$ を持つ．その上，線形作用素の定義よ

り $\mathcal{D}(A)$ は X の部分空間であるため，ゼロ元 $0 \in \mathcal{D}(A) \subset X$ を持つ．その上，A が連続であるため，A は $0 \in \mathcal{D}(A)$ でも連続である．$\varepsilon - \delta$ 論法による A の $0 \in \mathcal{D}(A) \subset X$ における連続の定義を記述すると

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \|u_n\|_X < \delta \text{ となる } \forall u_n \in X \text{ に対して } \|Au_n\|_Y < \varepsilon$$

となる．その上，ε を $n\|u_n\|_X$ とすると，$\delta_n > 0$ が存在し，$\|u_n\|_X < \delta$ となる任意の $u_n \in \mathcal{D}(A)$ に対して，

$$\|Au_n\|_Y < n\|u_n\|_X$$

となる．有界ではないという仮定と組み合わせると

$$n\|u_n\|_X < \|Au_n\|_Y < n\|u_n\|_X$$

となるため矛盾する．

□

5.2.4 定義域が X の全体となる有界な線形作用素全体の集合 $\mathcal{B}(X,Y)$

今まで作用素に「連続」,「線形」,「有界かつ線形」などのクラス分けをしました．この節では，さらに「有界かつ線形かつ $X = \mathcal{D}(A)$」となる作用素を考えます．しかし，今までと大きく違う点は,「有界かつ線形かつ $X = \mathcal{D}(A)$」となる作用素となるものすべて集めた集合を考えるので注意してください．では，まず，本節で主役となる集合の定義をしましょう．

定義 5.2.12 (定義域が X の全体となる有界な線形作用素全体の集合 $\mathcal{B}(X,Y)$)**.** Banach 空間 X から Banach 空間 Y の有界な線形作用素 A の定義域が X，すなわち

$$X = \mathcal{D}(A)$$

となるもの全体の集合を

$$\mathcal{B}(X,Y)$$

と書く．また，$X = Y$ の場合は省略して $\mathcal{B}(X)$ と記述する．

集合 $\mathcal{B}(X)$ の元 A は，$X = \mathcal{D}(A)$ となる有界な線形作用素であるため写像でもあります．しかし，写像と呼ぶことはなく，あくまで作用素と呼びます．その理由は，$X = \mathcal{D}(A)$ となる有界な線形作用素 A が全射ではないが単射の場合は，逆作用素 A^{-1} を持ちますが，写像 A と呼んでしまうと逆写像 A^{-1} は持ちません．このように，$X = \mathcal{D}(A)$ となる有界な線形作用素と写像は一見すると同じに見えますが，やはり $X = \mathcal{D}(A)$ となる有界な線形作用素も写像のより一般的な定義になっていることに注意が必要です．

本節の主役である集合 $\mathcal{B}(X,Y)$ に着目する理由は，$\mathcal{B}(X,Y)$ に適切なノルムを入れると Banach 空間になるからです．ここでは，まず，$\mathcal{B}(X,Y)$ に適切なノルムを入れ，ノルム空間になることを示した上で，Banach 空間になることを証明します．5.1 章の山場であるため，最初は証明を追うことができないかもしれません．証明を飛ばして結果だけ見る場合でも，証明後に掲載している Neumann 級数に関する定理は数値計算の品質保証で度々でてくるため，必ず把握しておいて下さい．

定理 5.2.5 ($\mathcal{B}(X,Y)$ は Banach 空間)**.** X をノルム空間とし，Y を Banach 空間とする．定義域が X 全体となる X から Y への有界な線形作用素全体の集合 $\mathcal{B}(X,Y)$ のノルムを

$$\|A\|_{\mathcal{B}(X,Y)} := \sup_{u\in X\setminus\{0\}} \frac{\|Au\|_Y}{\|u\|_X},\ A\in\mathcal{B}(X,Y) \tag{5.1}$$

とすると $\mathcal{B}(X,Y)$ は Banach 空間となる．

証明 . 定義 5.2.8 の作用素の加法と定義 5.2.9 の作用素のスカラー乗法をもとに線形空間の公理 5.1.1 が満たされていることは簡単に導かれる．但し，$\mathcal{B}(X,Y)$ のゼロ元は任意の $u\in X$ を $0\in Y$ へ写す作用素であることに注意が必要である．

「ノルム空間」

$\|A\|_{\mathcal{B}(X,Y)}$ がノルムの定義 5.1.1 を満たすことを示せばよい．ノルム空間 X と Banach 空間 Y であるため $\|\cdot\|_X\geq 0$ と $\|\cdot\|_Y\geq 0$ であることから

$$\frac{\|Au\|_Y}{\|u\|_X}\geq 0$$

となるため，$\|A\|_{\mathcal{B}(X,Y)}\geq 0$ となり，ノルムの定義 5.1.1 (i) はいえる．

次に，$A=0$ ならば $\|Au\|_Y=0$ であるため，

$$\|A\|_{\mathcal{B}(X,Y)} = \sup_{u\in X\setminus\{0\}} \frac{\|Au\|_Y}{\|u\|_X} = \sup_{u\in X\setminus\{0\}} \frac{0}{\|u\|_X} = 0$$

である．さらに，任意の $u\in X\setminus\{0\}$ について

$$\frac{\|Au\|_Y}{\|u\|_X}=0 \Leftrightarrow \|Au\|_Y=0 \Leftrightarrow Au=0$$

任意の $u\in X\setminus\{0\}$ を $0\in Y$ へ写す作用素は $\mathcal{B}(X,Y)$ が線形空間より一意に存在し，$A=0$ である．よって，ノルムの定義 5.1.1 (ii) も示された．

続いて，$\alpha\in\mathbb{K}$ としたとき，Y は Banach 空間であるため $\|\cdot\|_Y$ はノルムの定義を満たすため，

$$\|\alpha A\|_{\mathcal{B}(X,Y)} = \sup_{u\in X\setminus\{0\}} \frac{\|\alpha Au\|_Y}{\|u\|_X} = |\alpha| \sup_{u\in X\setminus\{0\}} \frac{\|Au\|_Y}{\|u\|_X} = |\alpha|\|A\|_{\mathcal{B}(X,Y)}$$

となるため，ノルムの定義 5.1.1 (iii) も示された．

最後に任意の $A,B\in\mathcal{B}(X,Y)$ について

$$\begin{aligned}
\|A+B\|_{\mathcal{B}(X,Y)} &= \sup_{u\in X\setminus\{0\}} \frac{\|(A+B)u\|_Y}{\|u\|_X} = \sup_{u\in X\setminus\{0\}} \frac{\|Au+Bu\|_Y}{\|u\|_X} \\
&\leq \sup_{u\in X\setminus\{0\}} \frac{\|Au\|_Y+\|Bu\|_Y}{\|u\|_X} \\
&\leq \sup_{u\in X\setminus\{0\}} \frac{\|Au\|_Y}{\|u\|_X} + \sup_{u\in X\setminus\{0\}} \frac{\|Bu\|_Y}{\|u\|_X} \\
&= \|A\|_{\mathcal{B}(X,Y)} + \|B\|_{\mathcal{B}(X,Y)}
\end{aligned}$$

となり，ノルムの定義 5.1.1 (iv) も示されたため，$\mathcal{B}(X,Y)$ はノルム空間である．

「Banach 空間」
Banach 空間であることを証明するには $\mathcal{B}(X,Y)$ の任意の Cauchy 列 $(A_n) \subset \mathcal{B}(X,Y)$ が極限 T を $\mathcal{B}(X,Y)$ 内に持つことを示せばよい．証明の流れは，まず，極限の候補 $\tilde{A}$ の定義できるか確認する．続いて極限の候補 $\tilde{A}$ が $\mathcal{B}(X,Y)$ に属していることを示す．最後に，極限の候補 $\tilde{A}$ が Cauchy 列 (A_n) の極限であることを示す.

まず，極限の候補 $\tilde{A}$ が定義に取り掛かる．任意の Cauchy 列 $(A_n) \subset \mathcal{B}(X,Y)$ は Cauchy 列の定義 5.1.3 より

$$\|A_n - A_m\|_{\mathcal{B}(X,Y)} \to 0,\ (n,m \to \infty)$$

となる．任意の $u \in X \setminus \{0\}$ に対して，点列 $(A_n u) \subset Y$ は

$$\begin{aligned} \|A_n u - A_m u\|_Y &= \frac{\|(A_n - A_m)u\|_Y}{\|u\|_X}\|u\|_X \\ &\leq \sup_{\phi \in X\setminus\{0\}} \frac{\|(A_n - A_m)\phi\|_Y}{\|\phi\|_X}\|u\|_X \\ &= \|A_n - A_m\|_{\mathcal{B}(X,Y)}\|u\|_X \to 0,\ (n,m \to \infty) \end{aligned}$$

を持つため，点列 $(A_n u) \subset Y$ は Cauchy 列になる．その上，Y は Banach 空間であるため，Y の任意の Cauchy 列は収束し，Y 内に極限 $\tilde{A}u$ となるような X から Y への作用素 $\tilde{A}$ が存在する．ここで，任意の $u \in X$ に対して極限 $\tilde{A}u$ が定義されることから，$\tilde{A}$ の定義域は $\mathcal{D}(\tilde{A}) = X$ である．これで，$\mathcal{B}(X,Y)$ の任意の Cauchy 列 (A_n) の極限の候補 $\tilde{A}$ が定義できた.

続いて，定義した極限の候補 $\tilde{A}$ が $\mathcal{B}(X,Y)$ に属しているか確認する．$\tilde{A}$ が有界な線形作用素であり，かつ $\mathcal{D}(\tilde{A}) = X$ であることを示せばよい．$\mathcal{B}(X,Y)$ の任意の Cauchy 列 Cauchy 列 (A_n) の元 A_n は線形作用素であるため，線形作用素の定義 5.2.7 より任意の $\alpha, \beta \in \mathbb{K}$ と $u, v \in X$ について

$$A_n(\alpha u + \beta v) = \alpha A_n u + \beta A_n v$$

を持つ．よって $n \to \infty$ とすると

$$\tilde{A}(\alpha u + \beta v) = \alpha \tilde{A}u + \beta \tilde{A}v$$

となり，極限の候補 $\tilde{A}$ は線形作用素である．次に極限の候補 $\tilde{A}$ が有界作用素であることを示す．点列 (A_n) は Cauchy 列であるため定理 5.1.3 より有界列でもある．すなわち，どんな $n \in \mathbb{N}$ に対しても

$$\|A_n\|_{\mathcal{B}(X,Y)} \leq M$$

となる $n \in \mathbb{N}$ に依存しない定数 M が存在する．この $n \in \mathbb{N}$ に依存しない定数 M は，任意の $u \in X$ について

$$\|A_n u\|_Y \leq M\|u\|_X$$

も満たす．$A_n u \to \tilde{A}u,\ (n \to \infty)$ であるため，上の不等式に対して $n \to \infty$ とすると M が n に依存しないため

$$\|\tilde{A}u\|_Y \leq M\|u\|_X$$

を得る．よって，点列 (A_n) の極限の候補 $\tilde{A}$ は $\mathcal{B}(X,Y)$ に属する.

最後に，点列 (A_n) の極限が $\tilde{A}$ であることを示す．任意の $u \in X$ に対して，点列 $(A_n u) \subset Y$ は Y 内に極限 $\tilde{A}u$ を持つこと，すなわち

$$A_n u \to \tilde{A}u,\ (n \to \infty)$$

を持つことから

$$\|A_n u - A_m u\|_Y \to \|A_n u - \tilde{A}u\|_Y,\ (m \to \infty)$$

となる．その上，$\mathcal{B}(X,Y)$ のノルムの定義と $\tilde{A} \in \mathcal{B}(X,Y)$ から

$$\|A_n - A_m\|_{\mathcal{B}(X,Y)} \to \|A_n - \tilde{A}\|_{\mathcal{B}(X,Y)},\ (m \to \infty)$$

を得る．点列 (A_n) が Cauchy 列であるため

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N},\ \forall n, m \geq N \text{ に対して } \|A_n - A_m\|_{\mathcal{B}(X,Y)} < \varepsilon$$

を満たす．その上，$m \to \infty$ とすると

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N},\ \forall n \geq N \text{ に対して } \|A_n - \tilde{A}\|_{\mathcal{B}(X,Y)} < \varepsilon$$

となり，$\tilde{A} \in \mathcal{B}(X,Y)$ は Cauchy 列 (A_n) の極限である．よって，任意の Cauchy 列は $\mathcal{B}(X,Y)$ 内に極限を持つため，ノルム空間 $\mathcal{B}(X,Y)$ は Banach 空間である．

□

これで，5.1 章の山場は越えました．続いて，Banach 空間 $\mathcal{B}(X,Y)$ のノルムに関するよく使われる性質を 2 つほど紹介しておきます．紹介する 2 つの性質はほぼ当たり前なので，証明は冗長に感じるかもしれません．

定理 5.2.6 ($\mathcal{B}(X,Y)$ のノルムの性質 (1))**.** X をノルム空間とし，Y を Banach 空間とする．そのとき，任意の $u \in X$ と任意の $A \in \mathcal{B}(X,Y)$ について以下の不等式が成り立つ:

$$\|Au\|_Y \leq \|A\|_{\mathcal{B}(X,Y)} \|u\|_X$$

証明． $u = 0$ の場合は明らかに成り立つため，$u \in X \setminus \{0\}$ について考える．$u \in X \setminus \{0\}$ について

$$\|Au\|_Y = \frac{\|Au\|_Y}{\|u\|_X}\|u\|_X \leq \sup_{\phi \in X \setminus \{0\}} \frac{\|A\phi\|_Y}{\|\phi\|_X}\|u\|_X = \|A\|_{\mathcal{B}(X,Y)}\|u\|_X$$

となるため，題意は示された．

□

定理 5.2.7 ($\mathcal{B}(X,Y)$ のノルムの性質 (2))**.** X をノルム空間とし，Y と Z を Banach 空間とする．そのとき，任意の $B \in \mathcal{B}(X,Y)$ と $A \in \mathcal{B}(Y,Z)$ の合成作用素 AB は $\mathcal{B}(X,Z)$ に属する．その上，

$$\|AB\|_{\mathcal{B}(X,Z)} \leq \|A\|_{\mathcal{B}(Y,Z)} \|B\|_{\mathcal{B}(X,Y)}$$

証明． 合成作用素の定義 5.2.10 から

$$\mathcal{D}(AB) = \{v \in \mathcal{D}(B) = X \mid Bv \in \mathcal{D}(A) = Y\}$$

となるが，$B \in \mathcal{B}(X,Y)$ であるため，任意の $v \in X$ に対して Bv は Y に属する．よって，

$$\mathcal{D}(AB) = \mathcal{D}(B) = X$$

となる．その上，A も B も線形作用素であることから，任意の $u, v \in X$ と任意の $\alpha, \beta \in \mathbb{K}$ に対して

$$AB(\alpha u + \beta v) = A(B\alpha u + B\beta v) = A(\alpha Bu + \beta Bv) = A\alpha Bu + A\beta Bv = \alpha ABu + \beta ABv$$

となるため，合成作用素 AB は定義域が X 全体となる線形作用素である．また，$A \in \mathcal{B}(Y,Z)$, $B \in \mathcal{B}(X,Y)$ であるため，任意の $u \in X$ について，定理 5.2.6 から

$$\|ABu\|_Z \leq \|A\|_{\mathcal{B}(Y,Z)}\|Bu\|_Y \leq \|A\|_{\mathcal{B}(Y,Z)}\|B\|_{\mathcal{B}(X,Y)}\|u\|_X$$

となり，定義域が X 全体となる線形作用素 AB は有界な線形作用素でもある．よって AB は $\mathcal{B}(X,Z)$ に属する．その上，

$$\begin{aligned}
\|AB\|_{\mathcal{B}(X,Z)} &= \sup_{u \in X\setminus\{0\}} \frac{\|ABu\|_Z}{\|u\|_X} \\
&\leq \sup_{u \in X\setminus\{0\}} \frac{\|A\|_{\mathcal{B}(Y,Z)}\|B\|_{\mathcal{B}(X,Y)}\|u\|_X}{\|u\|_X} \\
&= \|A\|_{\mathcal{B}(Y,Z)}\|B\|_{\mathcal{B}(X,Y)}
\end{aligned}$$

□

それでは，5.1 章の最後に Neumann 級数に関する定理を紹介に入ります．Neumann 級数に関する定理を紹介にあたり，まず，恒等作用素を定義します:

定義 5.2.13 (X 上の恒等作用素)**.** X を Banach 空間とする．任意の $u \in X$ に対して

$$Iu = u$$

となる $I \in \mathcal{B}(X)$ を X 上の恒等作用素と呼ぶ.

もちろん，X 上の恒等作用素 $I \in \mathcal{B}(X)$ は，任意の $A \in \mathcal{B}(X)$ に対して

$$AI = IA = A$$

となります．また，もちろん，

$$\mathcal{R}(I) = X$$

です.

では，Neumann 級数に関する定理を紹介します．Neumann 級数に関する定理は数値計算の品質保証で重要になる定理ですので覚えておいて下さい.

定理 5.2.8 (Neumann 級数)**.** X を Banach 空間とする．$B \in \mathcal{B}(X)$ とし，$I \in \mathcal{B}(X)$ を X 上の恒等作用素とする．もし

$$\|I - B\|_{\mathcal{B}(X)} < 1$$

ならば，B は逆作用素をもち $B^{-1} \in \mathcal{B}(X)$ となる．そのうえ，

$$B^{-1} = I + (I-B) + (I-B)^2 + \cdots = \sum_{i=0}^{\infty} (I-B)^i$$

で，かつ

$$\|B^{-1}\|_{\mathcal{B}(X)} \leq \frac{1}{1 - \|I-B\|_{\mathcal{B}(X)}}$$

証明．

$$S_n = I + (I-B) + (I-B)^2 + \cdots + (I-B)^n$$

とすると，B と I はともに $\mathcal{B}(X)$ に属するため，加法 $I-B$ や合成作用素 $(I-B)(I-B)$ なども $\mathcal{B}(X)$ に属する．よって S_n も $\mathcal{B}(X)$ に属する．

続いて，点列 $(S_n) \subset \mathcal{B}(X)$ が極限 S を $\mathcal{B}(X)$ 内に持つか確認する．定理 5.2.7 より

$$\|(I-B)^i\|_{\mathcal{B}(X)} \leq \|I-B\|^i_{\mathcal{B}(X)},\ i = 0, 1, \cdots$$

となるため，$n > m > 0$ となる整数に対して

$$\|S_n - S_m\|_{\mathcal{B}(X)} = \left\| \sum_{i=m+1}^{n} (I-B)^i \right\|_{\mathcal{B}(X)} \leq \sum_{i=m+1}^{n} \|I-B\|^i_{\mathcal{B}(X)}$$

となる．定理の仮定より $\|I-B\|_{\mathcal{B}(X)} < 1$ であるため，

$$\sum_{i=m+1}^{n} \|I-B\|^i_{\mathcal{B}(X)} \to 0,\ (n,\ m \to \infty)$$

となる．よって，

$$\|S_n - S_m\|_{\mathcal{B}(X)} \to 0,\ (n,\ m \to \infty)$$

となるため，点列 (S_n) は Cauchy 列である．その上，$\mathcal{B}(X)$ は Banach 空間であるため，任意の Cauchy 列は極限を $\mathcal{B}(X)$ に持つため，点列 (S_n) は

$$\|S_n - S\|_{\mathcal{B}(X)} \to 0,\ (n,\ m \to \infty)$$

となる極限 $S \in \mathcal{B}(X)$ を持つ．

次に S が B^{-1} になることを示す．合成作用素の定義 5.2.10 に従って合成作用素 BS_n を考える．X は Banach 空間であり，$B, S_n \in \mathcal{B}(X)$ であるため，定理 5.2.7 合成作用素 BS_n は $\mathcal{B}(X)$ に属する．その上，点列 $(BS_n) \subset \mathcal{B}(X)$ は

$$\|BS_n - BS\|_{\mathcal{B}(X)} \leq \|B\|_{\mathcal{B}(X)} \|S_n - S\|_{\mathcal{B}(X)} \to 0,\ (n \to \infty)$$

となるため，極限 BS を $\mathcal{B}(X)$ 内にもつ．一方で，

$$\begin{aligned} BS_n &= (I - (I-B))\, S_n = S_n - (I-B)S_n \\ &= \sum_{i=0}^{n} (I-B)^i - \sum_{i=1}^{n+1} (I-B)^i \\ &= I - (I-B)^{n+1} \end{aligned}$$

となり，定理の仮定より $\|I-B\|_{\mathcal{B}(X)}<1$ を持つため

$$\|BS_n - I\|_{\mathcal{B}(X)} = \|(I-B)^{n+1}\|_{\mathcal{B}(X)} \leq \|I-B\|^{n+1}_{\mathcal{B}(X)} \to 0,\ (n\to\infty)$$

となるため，点列 (BS_n) は極限 I も $\mathcal{B}(X)$ 内に持つ．よって，極限の一意性より

$$BS = I$$

を得る．$\mathcal{R}(I)=X$ であるため，$\mathcal{R}(BS)=X$ である．その上，$X=\mathcal{R}(BS)\subset\mathcal{R}(B)$ と $\mathcal{R}(B)\subset X$ となるため，$\mathcal{R}(B)=X$ となる．さらに，$S\in\mathcal{B}(X)$ であることから，

$$\mathcal{D}(S)=\mathcal{R}(B)=X$$

となる．

同様の議論を $S_nB\in\mathcal{B}(X)$ について行うと

$$SB = I$$

と

$$\mathcal{D}(B)=\mathcal{R}(S)=X$$

が得られる．そのため，B は逆作用素を持ち，逆作用素 $B^{-1}=S\in\mathcal{B}(X)$ である．

また，

$$S_n = I+(I-B)+(I-B)^2+\cdots+(I-B)^n \to B^{-1},\ (n\to\infty)$$

より

$$B^{-1} = I+(I-B)+(I-B)^2+\cdots = \sum_{i=0}^{\infty}(I-B)^i$$

となる．

最後に

$$\left\|B^{-1}\right\|_{\mathcal{B}(X)} = \left\|\sum_{i=0}^{\infty}(I-B)^i\right\|_{\mathcal{B}(X)} \leq \sum_{i=0}^{\infty}\|I-B\|^i$$

となり，初項 1，公比 $\|I-B\|_{\mathcal{B}(X)}<1$ の総和より

$$\left\|B^{-1}\right\|_{\mathcal{B}(X)} \leq \frac{1}{1-\|I-B\|_{\mathcal{B}(X)}}$$

□

Neumann の定理 5.2.8 は B をそのまま使う場合もありますが，$B=RA$ のように二つの線形作用素の合成作用素として利用することも良くあります．そのとき，RA が全単射であった場合における R と A の単射と全射に関する議論を行います．そのため，次の定理も Neumann の定理 5.2.8 とセットで覚えておきましょう．

定理 5.2.9. X と Y を Banach 空間とする．$A\in\mathcal{B}(X,Y)$, $R\in\mathcal{B}(Y,X)$ とする．もし RA が全単射ならば，A は単射であり，R は全射である．

証明．「A が単射」の証明
定理 5.2.2 (線形作用素に対する単射性 (1)) の ii) を用いて証明する．$u \in X$ とし，RA が単射であることに注意すると

$$Au = 0 \Rightarrow RAu = 0 \Rightarrow u = 0$$

よって，A は単射である．
「R が全射」の証明
RA が全射であるため任意の $g \in X$ に対して，

$$RAu = g$$

となる $u \in X$ が存在する．その上，$v = Au$ とすると任意の $g \in X$ に対して

$$Rv = g$$

となる $v \in Y$ が存在するため，R は全射である．

□

Neumann の定理 5.2.8 と定理 5.2.9，及び，線形代数の次元定理を利用することで定理 3.1.1 を証明することができます．実際に $\|I - RA\| < 1$ より Neumann の定理 5.2.8 から $RA \in \mathbb{R}^n$ は全単射になります．そのうえで，定理 5.2.9 から $R \in \mathbb{R}^n$ は全射となり，$A \in \mathbb{R}^n$ は単射になります．さらに，線形代数の次元定理を利用することで，R と A がそれぞれ全単射となることもいえます．よって，

$$\begin{aligned} Ax^* = b \Leftrightarrow \quad & A(x^* - \hat{x}) = b - A\hat{x} \\ \Leftrightarrow \quad & RA(x^* - \hat{x}) = R(b - A\hat{x}) \\ \Leftrightarrow \quad & x^* - \hat{x} = (RA)^{-1}R(b - A\hat{x}) \end{aligned}$$

となるため

$$\|x^* - \hat{x}\|_\infty = \|(RA)^{-1}R(b - A\hat{x})\|_\infty \leq \|(RA)^{-1}\|_\infty \|R(b - A\hat{x})\|_\infty \leq \frac{\|R(b - A\hat{x})\|_\infty}{1 - \|I - RA\|_\infty}$$

が成立します．

5.3 非線形解析の基礎

本書では線形，非線形，有限次元，無限次元を問わず方程式のコンピュータで得られた解の品質を保証するための定理を紹介することが目的です．そのためには，無限次元空間上の非線形作用素を解析するための道具が必要になります．有限次元の言葉を借りて説明すると，全微分，Riemann 積分，区間上のベクトル値関数に対する微分積分学の基本定理 (Fundamental theorem of calculus) が必要になります．Banach 空間上においては Fréchet 微分，Bochner 積分，区間上の Banach 空間値関数に対する微分積分学の基本定理が必要になります．

5.3.1 Fréchet 微分

本節では Banach 空間上の作用素に対する微分として Fréchet 微分を紹介します．ベクトル空間 $\mathbb{R}^n$ から $\mathbb{R}^m$ への写像に対する微分は方向微分と全微分がありましたね．Fréchet 微分は全微分の拡張にあたります．本書では取り扱いませんが，方向微分の拡張は Gâteaux 微分と呼ばれます．

定義 5.3.1 (Fréchet 微分)**.** X, Y を Banach 空間とし，開部分集合 $U \subset X$ とする．定義域を $\mathcal{D}(f) = U$ とする U から Y への作用素 f は U 上で連続とする．ある点 $v \in U$ に対し，$v + h \in U$ となる任意の $h \in X$ について

$$\frac{\|f(v+h) - f(v) - f'[v]h\|_Y}{\|h\|_X} \to 0, \ (h \to 0)$$

を満たす線形作用素 $f'[v] \in \mathcal{B}(X, Y)$ が存在するとき，作用素 f は点 v において Fréchet 微分可能といい，$f'[v] \in \mathcal{B}(X, Y)$ を f の点 v における Fréchet 微分と呼ぶ．

Fréchet 微分の定義 5.3.1 はランダウの記号 o を用いると

$$f(v+h) - f(v) = f'[v]h + o(h)$$

となる $o(h)$ が

$$\frac{\|o(h)\|_Y}{\|h\|_X} \to 0, \ (h \to 0)$$

を満たすときです．

定理 5.3.1 (Fréchet 微分の線形性)**.** X, Y を Banach 空間とし，開部分集合 $U \subset X$ とする．定義域を $\mathcal{D}(f_1) = \mathcal{D}(f_2) = U$ とする U から Y への作用素 f_1 と f_2 が，各々 $v \in U$ で Fréchet 微分可能とし，各々の Fréchet 微分を $f_1'[v]$ と $f_2'[v]$ と表記する．$\mathcal{D}(g) = U$ とする U から Y への作用素 g を

$$g(u) = \alpha f_1(u) + \beta f_2(u), \ u \in U, \ \alpha, \beta \in \mathbb{K}$$

とすると，作用素 g も $v \in U$ にて Fréchet 微分可能である．その上，g の $v \in U$ における Fréchet 微分を $g'[v]$ と表記すると

$$g'[v] = \alpha f_1'[v] + \beta f_2'[v]$$

が成立する．

証明．作用素 f_1 と f_2 が v において Fréchet 微分可能であるため，ランダウの記号を用いた表記

$$\frac{\|o_1(h)\|_Y}{\|h\|_X} \to 0,\ (h \to 0)$$

となる

$$f_1(v+h) - f_1(v) = f_1'[v]h + o_1(h)$$

と，

$$\frac{\|o_2(h)\|_Y}{\|h\|_X} \to 0,\ (h \to 0)$$

となる

$$f_2(v+h) - f_2(v) = f_2'[v]h + o_2(h)$$

を持つ．

$$\begin{aligned} g(v+h) &= \alpha f_1(v+h) + \beta f_2(v+h) \\ &= \alpha\left(f_1(v) + f_1'[v]h + o_{f_1}(h)\right) + \beta\left(f_2(v) + f_2'[v]h + o_{f_2}(h)\right) \\ &= (\alpha f_1(v) + \beta f_2(v)) + \left(\alpha f_1'[v] + \beta f_2'[v]\right)h + (\alpha o_{f_1}(h) + \beta o_{f_2}(h)) \\ &= g(v) + \left(\alpha f_1'[v] + \beta f_2'[v]\right)h + (\alpha o_{f_1}(h) + \beta o_{f_2}(h)) \end{aligned}$$

となる．その上，

$$\frac{\|\alpha o_{f_1}(h) + \beta o_{f_2}(h)\|_Y}{\|h\|_X} \le \frac{|\alpha|\|o_{f_1}(h)\|_Y}{\|h\|_X} + \frac{|\beta|\|o_{f_2}(h)\|_Y}{\|h\|_X} \to 0,\ (h \to 0)$$

となるため，作用素 g は $v \in U$ において Fréchet 微分可能で，g の Fréchet 微分 $g'[v]$ は

$$g'[v] = \alpha f_1'[v] + \beta f_2'[v]$$

となる．

□

定理 5.3.2 (Fréchet 微分の連鎖律)**.** X, Y, Z を Banach 空間とする．開部分集合を $U_1 \subset X$ とし，定義域を $\mathcal{D}(f_1) = U_1$ とする U_1 から Y への作用素 f_1 が，$v \in U_1$ で Fréchet 微分可能とする．また，と $U_2 \subset \mathcal{R}(f_1) \subset Y$ とし，定義域を $\mathcal{D}(f_2) = U_2$ とする U_2 から Z への作用素 f_2 が，$f_1(v) \in U_2 \subset \mathcal{R}(f_1)$ で Fréchet 微分可能とする．各々の Fréchet 微分を $f_1'[v]$ と $f_2'[f_1(v)]$ と表記する．$\mathcal{D}(g) = U_1$ とする U_1 から Z への作用素 g を

$$g(u) = f_2(f_1(u)),\ u \in U_1$$

とすると，作用素 g も $v \in U_1$ にて Fréchet 微分可能である．その上，g の $v \in U_1$ における Fréchet 微分を $g'[v]$ と表記すると

$$g'[v] = f_2'[f_1(v)]f_1'[v]$$

が成立する．

証明．作用素 f_1 が U_1 で Fréchet 微分可能であるため，ランダウの記号を用いた表記

$$\frac{\|o_1(h_1)\|_Y}{\|h_1\|_X} \to 0,\ (h_1 \to 0)$$

となる

$$f_1(v+h_1) - f_1(v) = f_1'[v]h_1 + o_1(h_1)$$

となる．同様に作用素 f_2 が U_2 で Fréchet 微分可能であるため

$$\frac{\|o_2(h_2)\|_Z}{\|h_2\|_Y} \to 0,\ (h_2 \to 0)$$

となる

$$f_2(f_1(v)+h_2) - f_2(f_1(v)) = f_2'[f_1(v)]h_2 + o_2(h_2)$$

をもつ．

$$f_2(f_1(v+h_1)) = f_2(f_1(v) + f_1'[v]h_1 + o_1(h_1))$$

となる．$h_2 = f_1'[v]h_1 + o_1(h_1)$ とおくと

$$\begin{aligned}
f_2(f_1(v+h_1)) &= f_2(f_1(v) + f_1'[v]h_1 + o_1(h_1)) \\
&= f_2(f_1(v) + h_2) \\
&= f_2(f_1(v)) + f_2'[f_1(v)]h_2 + o_2(h_2) \\
&= f_2(f_1(v)) + f_2'[f_1(v)](f_1'[v]h_1 + o_1(h_1)) + o_2(h_2) \\
&= f_2(f_1(v)) + f_2'[f_1(v)]f_1'[v]h_1 + \big(f_2'[f_1(v)]o_1(h_1) + o_2(h_2)\big)
\end{aligned}$$

となる．その上，Fréchet 微分の定義より $f_2'[f_1(v)] \in \mathcal{B}(Y,Z)$ であるため，$\|f_2'[f_1(v)]\|_{\mathcal{B}(Y,Z)} \leq M$ となる定数 $M > 0$ が存在する．また，

$$\begin{aligned}
\frac{\|f_2'[f_1(v)]o_1(h_1) + o_2(h_2)\|_Z}{\|h_1\|_X} &\leq \frac{\|f_2'[f_1(v)]o_1(h_1)\|_Z}{\|h_1\|_X} + \frac{\|o_2(h_2)\|_Z}{\|h_1\|_X} \\
&\leq M\frac{\|o_1(h_1)\|_Y}{\|h_1\|_X} + \frac{\|o_2(h_2)\|_Z}{\|h_2\|_Y}\frac{\|h_2\|_Y}{\|h_1\|_X}
\end{aligned}$$

となり，

$$\begin{aligned}
\frac{\|h_2\|_Y}{\|h_1\|_X} &= \frac{\|f_1'[v]h_1 + o_1(h_1)\|_Y}{\|h_1\|_X} \\
&\leq M + \frac{\|o_1(h_1)\|_Y}{\|h_1\|_X}
\end{aligned}$$

から

$$\begin{aligned}
&\frac{\|f_2'[f_1(v)]o_1(h_1) + o_2(h_2)\|_Z}{\|h_1\|_X} \\
\leq\ & M\frac{\|o_1(h_1)\|_Y}{\|h_1\|_X} + \frac{\|o_2(h_2)\|_Z}{\|h_2\|_Y}\left(M + \frac{\|o_1(h_1)\|_Y}{\|h_1\|_X}\right) \to 0,\ h_1, h_2 \to 0
\end{aligned}$$

となるため，作用素 g は $v \in U_1$ において Fréchet 微分可能で，g の Fréchet 微分 $g'[v]$ は

$$g'[v] = f_2'[f_1(v)]f_1'[v]$$

となる．

□

5.3.2 Bochner 積分

本節では Bochner 積分を紹介します．本書では，定理には証明をつけるスタンスですが，本節のみ証明を省きます．理由としては Lebesgue 積分の Banach 空間に値をとる関数への拡張であり，測度論から Lebesgue 積分論，Banach 空間に値をとる関数の可測性を経て Bochner 積分が定義され，定理が証明されるので，本書のゆったりペースでは 1 冊かけて説明しなくてはならなくなってしまいます．測度論及び Lebesgue 積分論を知っているという前提で，証明や詳細を知りたい場合は，宮寺功「関数解析」の第 6 章を参照してください．

本節では Bochner 積分のエッセンスのみ提示します．f を実数の閉区間 $[a,b]$ から $\mathbb{R}$ の写像としたとき，f の $[a,b]$ 上の Riemann 積分は

$$\int_a^b f(x)dx$$

のように微積分で習っていると思います．Bochner 積分では，

1. 区間 $[a,b]$ 上ではなく一般的な測度空間 S 上の積分を考える
2. f を Banach 空間 X への作用素にする

ことです．すなわち，f を測度空間 S から Banach 空間 X への作用素としたときの，f の S 上の積分を定義することが Bochner 積分の趣旨です．

しかし，本書では，具体的には一般的な測度空間 S を扱う必要がなく，実数の閉区間 $[a,b]$ から Banach 空間 X への作用素 f の積分のみで十分です．その上，結果的に言うと，いわゆる，積分の線形性や交換則などの Riemann 積分で習った性質もちゃんと受け継がれています．また，区間 $[a,b]$ からベクトル $\mathbb{R}^n$ への写像を区間上のベクトル値関数と呼ぶ風習にならって，区間 $[a,b]$ から Banach 空間 X への作用素を区間上の Banach 空間値関数と呼びます．

集合 S のいくつかの部分集合からなる集合族を Σ を完全加法集合体とし，μ を Σ 上で定義される測度とします．測度と完全加法集合体について細かい定義は省略しますが，測度とは面積や体積の一般化された概念であり，完全加法集合体とは測度を定義するには十分な性質を備えた集合の集まりです．集合，集合族，測度の三つの組 (S,Σ,μ) を測度空間と呼びます．

では，まず，単純関数を定義します．単純関数のイメージは区間 $[a,b]$ から $\mathbb{R}$ への関数が，区間を $[a,c_1),[c_1,c_2),\cdots[c_n,b]$ のように分割したときに，各々の分割した区間上で定数になる場合の一般化になります．

定義 5.3.2 (単純関数)**.** (S,Σ,μ) を測度空間とし，可測集合 $A_i\in\Sigma,\ i=1,2,\cdots,n$ を

$$A_i\cap A_j,\ i\neq j\quad \text{が空集合}$$

かつ

$$S=\bigcup_{i=1}^{\infty}A_i$$

とする．X を Banach 空間とし，f を定義域が S となる S から X への作用素が各 A_i 上で Banach 空間 X としての定数をとるとき，f を単純関数と呼ぶ．

続いて，単純関数の Bochner 積分を定義します．

定義 5.3.3 (単純関数の Bochner 積分)**.** (S,Σ,μ) を測度空間とし，X を Banach 空間とする．f を S から X への単純関数とする．すなわち，

$$f(s)=f_n,\ (s\in A_n)$$

となる Banach 空間 X の定数 $f_n \in X$ と

$$A_i \cap A_j,\ i \neq j \quad \text{が空集合}$$

かつ

$$S = \bigcup_{i=1}^{\infty} A_i$$

となる可測集合 $A_i \in \Sigma,\ i = 1, 2, \cdots, n$ が存在する．そのとき，

$$\|f(s)\|_X \text{が } S \text{ 上で Lebesgue 可積分}$$

のとき，$f(s)$ は S 上で Bochner 可積分であるといい，$f(s)$ の S 上の Bochner 積分を

$$\int_S f(s)d\mu := \sum_{i=1}^{\infty} f_n \mu(A_i) \in X$$

と定義する．

$\mu(A_i)$ は集合 A_i の測度です．厳密ではない言葉で表現すると，集合 A_i の面積に相当するものですので，$\mu(A_i)$ の定義には $\mu(A_i) \geq 0$ が含まれます．そのため，単純関数の Bochner 積分の定義より

$$\left\| \int_S f(s)d\mu \right\|_X \leq \int_S \|f(s)\|_X d\mu$$

を満たします．(上式の右辺になれば Lebesgue 積分で十分ですね．)

つづいては一般的な関数の Bochner 積分の定義を定義します:

定義 5.3.4 (Bochner 積分)**.** (S, Σ, μ) を測度空間とし，X を Banach 空間とする．定義域 $\mathcal{D}(f) = S$ となる S から X への作用素を f とする．$s \in S$ に対してほとんど至るところで

$$f_n(s) \to f(s),\ (n \to \infty)$$

かつ

$$\int_S \|f(s) - f_n(s)\|_X d\mu \to 0,\ (n \to \infty)$$

となるような S 上で Bochner 可積分な単純関数の列 $\{f_n(s)\}$ が存在するとき，$f(s)$ は S 上で Bochner 可積分であるという．その上，$f(s)$ の S 上の Bochner 積分を $\int_S f_n(s)d$ の極限，すなわち

$$\int_S f_n(s)d\mu \to \int_S f(s)d\mu,\ (n \to \infty)$$

として定義する．

これで Bochner 積分が定義できました．続いて，Bochner 積分の線形性や有界作用素との関係など本書で使う性質を証明を付けずに紹介します．

定理 5.3.3 (Bochner 積分のノルム評価)**.** (S,Σ,μ) を測度空間とし，X を Banach 空間とする．f を S から X への S 上の Bochner 可積分な関数とする．そのとき

$$\left\|\int_S f(s)d\mu\right\|_X \leq \int_S \|f(s)\|_X d\mu$$

となる．

定理 5.3.4 (Bochner 積分の線形性)**.** (S,Σ,μ) を測度空間とし，X を Banach 空間とする．f と g を S から X への S 上の Bochner 可積分な関数とする．$\alpha,\beta\in\mathbb{K}$ としたとき，

$$\int_S (\alpha f(s)+\beta g(s))\,d\mu = \alpha\int_S f(s)d\mu + \beta\int_S g(s)d\mu$$

となる．

定理 5.3.5 (Bochner 積分の線形性)**.** (S,Σ,μ) を測度空間とし，X を Banach 空間とする．f と g を S から X への S 上の Bochner 可積分な関数とする．$\alpha,\beta\in\mathbb{K}$ としたとき，

$$\int_S (\alpha f(s)+\beta g(s))\,d\mu = \alpha\int_S f(s)d\mu + \beta\int_S g(s)d\mu$$

となる．

定理 5.3.6 (有界線形作用素と Bochner 積分の関係)**.** (S,Σ,μ) を測度空間とし，X と Y を Banach 空間とする．f を S から X への S 上の Bochner 可積分な関数とする．$B\in\mathcal{B}(X,Y)$ とすると $\alpha,\beta\in\mathbb{K}$ としたとき，

$$\int_S Bf(s)d\mu = B\int_S f(s)d\mu$$

となる．

続いて，測度空間 (S,Σ,μ) の S を閉区間 $[a,b]\subset\mathbb{R}$ のように限定した場合の性質を紹介します:

定義 5.3.5 (閉区間上の Bochner 積分)**.** X を Banach 空間とし，f を閉区間 $[a,b]$ から X への $[a,b]$ 上の Bochner 可積分な関数とする．そのとき，$[a,b]$ 上の Bochner 積分を

$$\int_a^b f(s)ds := \int_{[a,b]} f(s)d\mu$$

と記述する．

定理 5.3.7 (閉区間上の連続作用素と Bochner 積分)**.** X を Banach 空間とし，f を閉区間 $[a,b]$ から X への $[a,b]$ 上の Bochner 可積分な関数とする．f が任意の $t\in[a,b]$ において連続ならば $[a,b]$ 上で Bochner 可積分である．

5.3.3 閉区間上の Banach 空間値関数に対する微分積分学の基本定理

本章の最後に Fréchet 微分と閉区間上の Bochner 積分の関係性を示す微分積分学の基本定理を導出します.

定理 5.3.8 (Fréchet 微分と閉区間上における Bochner 積分に対する微分積分学の基本定理 (i))**.** X を Banach 空間とする．f を閉区間 $[a,b]$ から X への $[a,b]$ 上で連続な関数とする．$t \in (a,b)$ とし，関数 g を

$$g(t) := \int_a^t f(s)ds$$

とすると，g は $t \in (a,b)$ において Fréchet 微分可能である．その上，g の t の Fréchet 微分を $g'[t]$ とすると

$$g'[t] = f(t)$$

となる．

証明． f は任意の $t \in [a,b]$ において連続であるため，

$$\forall \varepsilon > 0\ \exists \delta > 0\ |s-t| < \delta \text{ となる } \forall s \in [a,b] \text{ に対して } \|f(s) - f(t)\|_X < \varepsilon$$

が成立する．その上，g の t の Fréchet 微分可能か定義にあてはめると

$$\begin{aligned}
\frac{\|g(t+h) - g(t) - f(t)h\|_X}{|h|} &= \frac{\left\| \int_a^{t+h} f(s)ds - \int_a^t f(s)ds - f(t)h \right\|_X}{|h|} \\
&= \frac{\left\| \int_t^{t+h} f(s)ds - f(t)h \right\|_X}{|h|} \\
&= \frac{\left\| \int_t^{t+h} f(s)ds - (f(t)(t+h) - f(t)t) \right\|_X}{|h|} \\
&= \frac{\left\| \int_t^{t+h} f(s)ds - \int_t^{t+h} f(t)ds \right\|_X}{|h|} \\
&= \frac{\left\| \int_t^{t+h} (f(s) - f(t))\, ds \right\|_X}{|h|} \\
&\leq \frac{\int_t^{t+h} \|f(s) - f(t)\|_X\, ds}{|h|} \\
&< \frac{\varepsilon \int_t^{t+h} ds}{|h|} = \varepsilon
\end{aligned}$$

となる．その上，$\varepsilon > 0$ は任意であるため，

$$\frac{\|g(t+h) - g(t) - f(t)h\|_X}{|h|} \to 0,\ (h \to 0)$$

となるため，g は Fréchet 微分可能で g の t における Fréchet 微分は $f(t)$ となる．

□

定理 5.3.9 (Fréchet 微分と閉区間上における Bochner 積分に対する微分積分学の基本定理 (ii))**.** X を Banach 空間とする．f を閉区間 $[a,b]$ から X への $[a,b]$ 上で連続な関数とし，その上，任意

の $s \in (a,b)$ において Fréchet 微分可能とする．f の $s \in (a,b)$ における Fréchet 微分を $f'[s]$ とし，$f' : [a,b] \to X$ で連続[1] であるとすると，

$$\int_a^b f'[s]ds = f(b) - f(a)$$

となる．

証明． f' が $[a,b]$ から X で連続であるため定理 5.3.7 から $f'[s],\ s \in [a,b]$ で Bochner 可積分ある．その上，g を

$$g(t) = \int_a^t f'[s]ds$$

とすると定理 5.3.8 から，g は $t \in (a,b)$ において Fréchet 微分可能で，

$$g'[t] = f'[t]\ \forall t \in (a,b)$$

となる．

$$F(t) = g(t) - f(t)$$

とおくと，g も f も $t \in (a,b)$ において Fréchet 微分可能であることから，

$$F'[t] = g'[t] - f'[t] = 0,\ \forall t \in (a,b)$$

となる．また，$F'[t]$ がゼロであることから，Fréchet 微分の定義により $F(t)$ は t に依存しない作用素になる．よって

$$F(t) = g(t) - f(t) = c$$

となる $t \in [a,b]$ によって変化しない $c \in X$ が存在する．その上，上の式の t に a を代入すると

$$g(a) - f(a) = 0 - f(a) = c$$

となるため，$c = -f(a)$ である．よって，$g(t) - f(t) = c$ に注意すると

$$\int_a^t f'[s]ds = g(t) = f(t) + c = f(t) - f(a)$$

が成り立つ．ゆえに $t = b$ とすると

$$\int_a^b f'[s]ds = f(a) - f(b)$$

となる．

□

[1] $f'[v] \in \mathcal{B}(X,Y)$ が $h \in Y$ で連続とは

$$\forall \varepsilon > 0,\ \exists > 0,\ \|h_n - h\|_X < \delta \text{ となる } \forall h_n \in X \text{ に対して } \|f'[v]h_n - f'[v]h\|_Y < \varepsilon$$

が成立することです．その一方で，$f' : [a,b] \to \mathcal{B}(X,Y)$ が $h \in [a,b]$ で連続とは

$$\forall \varepsilon > 0,\ \exists > 0,\ |h_n - h| < \delta \text{ となる } \forall h_n \in \mathbb{R} \text{ に対して } \|f'[h_n] - f'[h]\|_{\mathcal{B}(X,Y)} < \varepsilon$$

が成立することです．

定理 5.3.10 (Fréchet 微分と閉区間上の Bochner 積分に対する微分積分学の基本定理 (iii))**.** X, Y を Banach 空間とし，U を X の開部分集合とする．作用素 $F : U \to Y$ とし，任意の $v \in U$ において Fréchet 微分可能であるとする．$u, v \in U$ とし，

$$w(t) = (1-t)u + tv,\ t \in [0,1]$$

が U に属すると仮定する．$F' : U \to \mathcal{B}(X,Y)$ が U 上で連続であると仮定する．そのとき，

$$F(v) - F(u) = \int_0^1 F'[(1-t)u + tv](v-u)dt$$

となる．

証明． $w : [0,1] \to U$ は Fréchet 微分可能で

$$w'[t] = v - u$$

となる．そのうえで，$w' : [0,1] \to X$ は $\|w'[s] - w'[t]\|_X = 0$ となるため，$[0,1]$ 上で連続である．さらに，作用素 $F : U \to Y$ が任意の $v \in U$ において Fréchet 微分可能であるため，

$$f(t) := F(w(t))$$

となる f は Fréchet 微分の連鎖律定理 5.3.2 より

$$f'[t] = F'[w(t)]w'[t] = F'[w(t)](v-u)$$

となるため，任意の $[0,1]$ 上で Fréchet 微分可能である．また，仮定から $F' : U \to \mathcal{B}(X,Y)$ が U 上で連続していることより，$f' : [0,1] \to \mathcal{B}(X,Y)$ は $[0,1]$ 上で連続となる．よって，f に対して定理 5.3.9 を利用でき，

$$f(1) - f(0) = \int_0^1 f'[t]dt$$

となる．ゆえに，

$$F(v) - F(u) = f(1) - f(0) = \int_0^1 f'[t]dt = \int_0^1 F'[w(t)](v-u)dt$$

となり，題意は示せた．

□

5.4 方程式の解の品質保証の準備

本章では，数値計算を用いて得られる方程式の解を品質保証するための基本的な定理を紹介します．品質保証するための基本定理はBanach空間の抽象的な関数方程式における解の存在定理が根幹にあり，不動点定理やNewton-Kantorovichの定理を品質保証のために使いやすくした定理だと思ってください．今まで導出してきたNeumann級数の定理5.2.8，Fréchet微分 (定義5.3.1)，Fréchet微分と閉区間上のBochner積分に対する微分積分学の基本定理 (iii)(定理5.3.10) のほかに，Banachの不動点定理を利用します．そのため，まず，Banachの不動点定理の紹介から始めます．

5.4.1 Banachの不動点定理

定義 5.4.1 (不動点)**.** X を係数体が $\mathbb{K}$ のBanach空間とする．M は空でない閉集合で $M \subset X$ を満たすとする．A を M から M への写像とする．$x \in M$ がAの不動点であるとは，x が

$$x = Ax$$

を満たすことである．

定義 5.4.2 (縮小写像)**.** X を係数体が $\mathbb{K}$ のBanach空間とする．M は空でない閉集合で $M \subset X$ を満たすとする．写像 $A : M \to M$ が k 次の縮小写像であるとは，$0 \leq k < 1$ を満たす定数 k が存在し，$\forall x, y \in M$ について

$$\|Ax - Ay\| \leq k\|x - y\|$$

を満たすことである．

定理 5.4.1 (Banachの不動点定理)**.** X を係数体が $\mathbb{K}$ のBanach空間とする．M は空でない閉集合で $M \subset X$ を満たすとする．A は M から M への k 次の縮小写像とする．そのとき，問題

$$\text{Find } u \in M \text{ s.t. } u = Au \tag{5.2}$$

は真の解 u^* を M 内にただ一つ持つ．即ち，写像 A は M 上にただ一つ不動点 u^* を持つ．

系 5.4.1. X,M,A,k は定理5.4.1で定義されている集合，写像及び定数とする．u_0 を閉集合 M の元として与えられていると仮定する．そのとき反復法

$$u_{n+1} = Au_n,\ n = 0, 1, \cdots$$

によって得られる点列 (u_n) は (5.2) を満たす真の唯一解 u^* に収束する．さらに，$\forall n \in \mathbb{N}$ について不等式

$$\|u_n - u^*\| \leq k^n(1-k)^{-1}\|u_1 - u_0\| \tag{5.3}$$
$$\|u_{n+1} - u^*\| \leq k(1-k)^{-1}\|u_{n+1} - u_n\| \tag{5.4}$$
$$\|u_{n+1} - u^*\| \leq k\|u_n - u^*\| \tag{5.5}$$

が成立する．

注意 5.4.1. (5.3)，(5.4)，(5.5) はそれぞれ事前誤差評価，事後誤差評価，収束率と呼ばれる．

証明 (Banach の不動点定理). u_0 を閉集合 M の元として与えられていると仮定する．点列 (u_n) は反復法

$$u_{n+1} = Au_n,\ n = 0, 1, \cdots \tag{5.6}$$

によって得られる．そのとき，証明のプロセスは次のように考える:
I) (u_n) が Cauchy 列になること，さらに Banach 空間の完備性を使うことで，$u_n \to u,\ n \to \infty$ となる u が X 内に存在することを示す.
II) u が (5.2) を満たす真の解 u^* と一致することを示す (解の存在性).
III) 真の解 u^* が M 内で一意であることを示す.

I)
(5.6) より

$$\|u_n - u_{n+1}\| = \|Au_{n-1} - Au_n\|$$

となる．仮定より A は k 次の縮小写像であるため,

$$\|Au_{n-1} - Au_n\| \le k\|u_{n-1} - u_n\|$$

となる定数 k が存在する．同様に $\|u_{n-1} - u_n\|$ に (5.6) と A の縮小写像の性質を使うと最終的に

$$\|u_n - u_{n+1}\| \le k^n\|u_0 - u_1\| \tag{5.7}$$

を得る.
次に三角不等式 (定義 5.1.1 (iv)) より，$n = 0, 1, 2, \cdots,\ m > n$ について

$$\begin{aligned} \|u_n - u_m\| &= \|(u_n - u_{n+1}) + (u_{n+1} - u_{n+2}) + \cdots + (u_{m-1} - u_m)\| \\ &\le \|u_n - u_{n+1}\| + \|u_{n+1} - u_{n+2}\| + \cdots + \|u_{m-1} - u_m\| \end{aligned} \tag{5.8}$$

となる．上の式に (5.7) を適用すると

$$\begin{aligned} \|u_n - u_m\| &\le \|u_n - u_{n+1}\| + \|u_{n+1} - u_{n+2}\| + \cdots + \|u_{m-1} - u_m\| \\ &\le k^n\|u_0 - u_1\| + k^{n+1}\|u_0 - u_1\| + \cdots + k^{m-1}\|u_0 - u_1\| \\ &= k^n(1 + k + \cdots + k^{m-n-1})\|u_0 - u_1\| \end{aligned} \tag{5.9}$$

となる．ここで，k は $0 \le k < 1$ であるため，$1 + k + \cdots + k^{m-n-1} \le 1 + k + \cdots + k^{m-1}$ となる．さらに，等比級数から

$$1 + k + \cdots + k^{m-1} = \frac{1 - k^m}{1 - k} \tag{5.10}$$

であるため，(5.9) は

$$\|u_n - u_m\| \le \frac{k^n(1 - k^m)}{1 - k}\|u_0 - u_1\| \tag{5.11}$$

となる．よって，k は $0 \le k < 1$ から $k^n \to 0,\ n \to \infty$ と $k^m \to 0,\ m \to \infty$ となる．即ち,

$$\|u_n - u_m\| \to 0,\ (n, m \to \infty)$$

となるため，点列 (u_n) は Cauchy 列である．さらに X は Banach 空間であるため，X は完備である (定義 5.1.5)．即ち，任意の Cauchy 列が X の中で極限を持つ (定義 5.1.4)．よって，点列 (u_n) は

$$u_n \to u,\ n \to \infty$$

となる $u \in X$ が存在する．

II)

u_0 を M の元とする．仮定より A は M から M の写像であるため，$A(M) \subset M$ となる．即ち，$u_1 = Au_0$ が成立する $u_1 \in M$ が存在する．同様に，$\forall n \in \mathbb{N}$ について $u_n \in M$ が存在する．さらに，M は閉集合であるため，I) で存在を示した u は M に属する (定義 5.1.10)．よって Au も M に属する．その上で仮定より A は k 次の縮小写像であるため，

$$\|Au_n - Au\| \leq k\|u_n - u\|$$

を得る．I) より点列 (u_n) は極限を持つため，$\|u_n - u\| \to 0,\ n \to \infty$ となる (定義 5.1.2)．即ち

$$\|Au_n - Au\| \to 0,\ n \to \infty$$

となるため，Au は Au_n の極限である (定義 5.1.2)．よって，(5.6) について $n \to \infty$ とすると

$$u = Au$$

が成立する．よって，u は (5.2) を満たす真の解 u^* となる．

III)

$u^*, v^* \in M$ をそれぞれ $u^* = Au^*$ と $v^* = Av^*$ を満たすとする．そのとき，A は k 次の縮小写像であるため，

$$\|u^* - v^*\| = \|Au^* - Av^*\| \leq k\|u^* - v^*\|$$

を得る．ここで $0 \leq k < 1$ であるため，不等式を満たすものは $u^* = v^*$ の場合のみである．即ち，(5.2) を満たす真の解は一意である．

□

証明 (Banach の不動点定理の系)**.** 点列 (u_n) が収束すること，極限が (5.2) を満たす真の解 u^* であることは定理 5.4.1 の証明の I) と II) と同一の証明で示せている．

(5.3) の証明

(5.11) について $m \to \infty$ とすると $k^m \to 0$ より

$$\|u_n - u\| \quad \leq \quad \frac{k^n}{1-k}\|u_1 - u_0\|$$

(5.4) の証明

(5.8) より

$$\begin{aligned}
\|u_{n+1} - u_m\| &\leq \|u_{n+1} - u_{n+2}\| + \|u_{n+2} - u_{n+3}\| + \cdots + \|u_{m-1} - u_m\| \\
&\leq k\|u_n - u_{n+1}\| + k^2\|u_n - u_{n+1}\| + \cdots + k^{m-n-1}\|u_n - u_{n+1}\| \\
&= k(1 + k + \cdots + k^{m-n-1})\|u_n - u_{n+1}\|
\end{aligned}$$

を得る．ここで $1+k+\cdots+k^{m-n-1} \leq 1+k+\cdots+k^{m-1}$ より，

$$\|u_{n+1}-u_m\| \quad \leq \quad k(1+k+\cdots+k^{m-1})\|u_n-u_{n+1}\|$$

となる．よって (5.10) より

$$\|u_{n+1}-u_m\| \quad \leq \quad k\frac{1-k^m}{1-k}\|u_n-u_{n+1}\|$$

となる．よって $m \to \infty$ とすると

$$\|u_{n+1}-u\| \quad \leq \quad \frac{k}{1-k}\|u_{n+1}-u_n\|$$

(5.5) の証明

$$\|u_{n+1}-u\| = \|Au_n - Au\| \leq k\|u_n-u\|$$

□

5.4.2　方程式の解の品質保証のための基本定理

本節では方程式の解の品質保証を行うための基本的な定理を紹介する．X, Y を Banach 空間とし，非線形方程式

$$\text{Find } u \in X \text{ s.t. } F(u)=0$$

を求める問題を考えます．この問題は有限次元や無限次元を問わず，連立一次方程式や非線形方程式，偏微分方程式などを取り扱うことができます．この問題に対して近似解の品質保証を行う定理として，以下の Newton-Kantorovich の定理の亜種を紹介します:

定理 5.4.2 (基本定理)**.** X と Y を Banach 空間とする．$F : X \to Y$ を与えられた作用素とし，$\hat{u} \in X$ を与えられているとする ($\hat{u}$ はコンピュータで求めた近似解をイメージで!)．$R \in \mathcal{B}(Y,X)$ とする．F は $\hat{u}$ で Fréchet 微分可能とし $F'[\hat{u}]$ と表記する．$F'[\hat{u}]$ は全射であるとする．η と δ を不等式

$$\|RF(\hat{u})\|_X \leq \eta$$

と

$$\|I-RF'[\hat{u}]\|_{\mathcal{B}(X)} \leq \delta < 1$$

を満たす定数とする．

$\bar{B}(0, 2\eta/(1-\delta)) = \{v \in X \mid \|v\|_X \leq 2\eta/(1-\delta)\}$ とする．F を $\bar{B}(\hat{u}, 2\eta/(1-\delta))$ 上で Fréchet 微分可能であるとし，$F' : \bar{B}(\hat{u}, 2\eta/(1-\delta)) \to \mathcal{B}(X,Y)$ が $\bar{B}(\hat{u}, 2\eta/(1-\delta))$ 上で連続であるとする．

K を不等式

$$\|R\left(F'[\hat{u}]-F'[\hat{u}+v]\right)\|_{\mathcal{B}(X)} \leq K,\ v \in \bar{B}\left(0, \frac{2\eta}{1-\delta}\right)$$

を満たす定数とする．もし $2K+\delta \leq 1$ ならば，

$$\|u^*-\hat{u}\|_X \leq \frac{\eta}{1-(K+\delta)} =: \rho$$

に対し，真の解 u^* は $\bar{B}(\hat{u}, \rho)$ 内に存在する．その上，$\bar{B}(\hat{u}, 2\eta/(1-\delta))$ 内で一意である．

証明．まず，R と $F'[\hat{u}]$ が全単射であることを示す．$F'[\hat{u}]$ は F の $\hat{u}$ における Fréchet 微分であることから，$F'[\hat{u}]$ は $\mathcal{B}(X,Y)$ に属する．さらに，Neumann 級数の定理 5.2.8 と仮定 $\|I-RF'[\hat{u}]\|_{\mathcal{B}(X)} \leq \delta < 1$ より，$RF'[\hat{u}]$ は全単射である．その上，定理 5.2.9 より，$F'[\hat{u}]$ は単射であり，R は全射である．また，定理の仮定より $F'[\hat{u}]$ は全単射となるため，逆作用素 $F'[\hat{u}]^{-1}$ が存在し，$\mathcal{B}(Y,X)$ に属する．また，R の単射性については，定理 5.2.2 と逆作用素 $F'[\hat{u}]^{-1}$ を用いて

$$\text{Find } \phi \in Y \text{ s.t. } R\phi = 0 \Rightarrow RF'[\hat{u}]F'[\hat{u}]^{-1}\phi = 0 \Rightarrow \phi = 0$$

となるため，R も全単射である．

続いて，Banach の不動点定理 5.4.1 を用いて解の存在を示す．まず，作用素方程式 $F(u)=0$ を不動点方程式に変形する．$w := u^* - \hat{u}$ とする．R が単射であるため，

$$\begin{aligned} & F(u^*) = 0 \\ \Leftrightarrow & w = w - RF(\hat{u}+w) \\ \Leftrightarrow & w = -RF(\hat{u}) + w - R\left(F(\hat{u}+w) - F(\hat{u})\right). \end{aligned}$$

$\mathcal{T} : X \to X$ を

$$\mathcal{T}(w) := -RF(\hat{u}) + w - R\left(F(\hat{u}+w) - F(\hat{u})\right)$$

となる非線形作用素とし，不動点方程式 $w = \mathcal{T}(w)$ の解の存在を Banach の不動点定理 5.4.1 を用いて示す．

Banach の不動点定理 5.4.1 では M を決めて，$\mathcal{T}$ が M から M への縮小写像になることを確認しなければならない．特に，ポイントの一つは $\mathcal{T}$ の定義域を M としたときに，値域が M に含まれることを確かめなければならない．すなわち，$\mathcal{T}(M) \subset M$ となるように M を選ぶことが重要である．この定理では，$M = \bar{B}(0,\rho),\ \rho = \eta/(1-(K+\delta))$ と選ぶ．まず，M として選んだ閉球 $\bar{B}(0,\rho)$ と定理で出てくる，もう一つの閉球 $\bar{B}(0, 2\eta/(1-\delta))$ の関係性を確認する． 定理の仮定より $1-\delta > 0$ と $1-(\delta+2K) \geq 0$ を持つため，

$$\begin{aligned} \rho - \frac{2\eta}{1-\delta} &= \frac{\eta(1-\delta)}{(1-(\delta+K))(1-\delta)} - \frac{2\eta(1-(\delta+K))}{(1-(\delta+K))(1-\delta)} \\ &= \frac{\eta\left((1-\delta) - 2(1-\delta) + 2K\right)}{(1-(\delta+K))(1-\delta)} = \frac{-\eta\left(1-\delta-2K\right)}{(1-(\delta+K))(1-\delta)} \leq 0 \end{aligned}$$

となる．よって，$\rho \leq 2\eta/(1-\delta)$ から $\bar{B}(0,\rho) \subset \bar{B}(0, 2\eta/(1-\delta))$ となる．

次に $\mathcal{T}(\bar{B}(0,\rho)) \subset \bar{B}(0,\rho)$ を示す．仮定から F' が $\bar{B}(\hat{u}, 2\eta/(1-\delta))$ 上で連続であることから，任意の $w \in \bar{B}(0,\rho)$ に対し Fréchet 微分と Bochner 積分に対する微分積分学の基本定理 5.3.10 を用

いることで

$$\begin{aligned}\|\mathcal{T}(w)\|_X \leq& \|RF(\hat{u})\|_X + \|w - R\left(F(\hat{u}+w) - F(\hat{u})\right)\|_X \\ \leq& \eta + \left\| w - R\int_0^1 F'[(1-t)\hat{u} + t(\hat{u}+w)]w dt \right\|_X \\ \leq& \eta + \int_0^1 \left\| w - RF'[\hat{u}+tw]w \right\|_X dt \\ \leq& \eta + \int_0^1 \left\| I - RF'[\hat{u}+tw] \right\|_{\mathcal{B}(X)} \|w\|_X dt \\ \leq& \eta + \int_0^1 \left(\left\| R(F'[\hat{u}] - F'[\hat{u}+tw]) \right\|_{\mathcal{B}(X)} + \left\| I - RF'[\hat{u}] \right\|_{\mathcal{B}(X)} \right) \|w\|_X dt \\ \leq& \eta + \int_0^1 \left(\left\| R(F'[\hat{u}] - F'[\hat{u}+tw]) \right\|_{\mathcal{B}(X)} + \delta \right) \|w\|_X dt\end{aligned}$$

を得る．さらに $t \in [0,1]$ に対し $tw \in \bar{B}(0,\rho) \subset \bar{B}(0, 2\eta/(1-\delta))$ となるため，定理の仮定 $\|R(F'[\hat{u}] - F'[\hat{u}+tw])\|_{\mathcal{B}(X)} \leq K$ から

$$\begin{aligned}\|\mathcal{T}(w)\|_X \leq& \eta + (K+\delta)\|w\|_X \\ \leq& \eta + (K+\delta)\rho \\ =& \eta + \frac{\eta(K+\delta)}{1-(\delta+K)} \\ =& \frac{\eta - \eta(\delta+K)}{1-(\delta+K)} + \frac{\eta(K+\delta)}{1-(\delta+K)} = \rho\end{aligned}$$

となる．よって，任意の $w \in \bar{B}(0,\rho)$ に対して $\|\mathcal{T}(w)\|_X \leq \rho$ となることから，$\mathcal{T}(\bar{B}(0,\rho)) \subset \bar{B}(0,\rho)$ となる．

次に $\mathcal{T} : \bar{B}(0,\rho) \to \bar{B}(0,\rho)$ が縮小写像になることを確認する．すなわち

$$\|\mathcal{T}(w_1) - \mathcal{T}(w_2)\|_X \leq k\|w_1 - w_2\|_X,\ \forall w_1, w_2 \in \bar{B}(0,\rho)$$

となる定数 k が 1 未満になることを確認しなければならない．この定理では，一意性の範囲を広げるために，$\bar{B}(0,\rho) \subset \bar{B}(0, 2\eta/(1-\delta))$ であることを用いて

$$\|\mathcal{T}(w_1) - \mathcal{T}(w_2)\|_X \leq k\|w_1 - w_2\|_X,\ \forall w_1, w_2 \in \bar{B}\left(0, \frac{2\eta}{1-\delta}\right)$$

となる定数 k が 1 未満になることを確かめる．その上，任意の $w_1, w_2 \in \bar{B}(0, 2\eta/(1-\delta))$ に対し，

$$\begin{aligned}\|\mathcal{T}(w_1) - \mathcal{T}(w_2)\|_X =& \|w_1 - w_2 - R\left(F(\hat{u}+w_1) - F(\hat{u}+w_2)\right)\|_X \\ =& \left\| (w_1 - w_2) - R\int_0^1 F'[\hat{u} + (1-t)w_2 + tw_1](w_1 - w_2) dt \right\|_X \\ \leq& \int_0^1 \left\| I - RF'[\hat{u} + (1-t)w_2 + tw_1] \right\|_{\mathcal{B}(X)} dt \|w_1 - w_2\|_X \\ \leq& \int_0^1 \left(\left\| R\left(F'[\hat{u}] - F'[\hat{u} + (1-t)w_2 + tw_1]\right) \right\|_{\mathcal{B}(X)} + \delta \right) dt \|w_1 - w_2\|_X\end{aligned}$$

となる．任意の $w_1, w_2 \in \bar{B}(0, 2\eta/(1-\delta))$ と $0 \leq t \leq 1$ に対し，$\|(1-t)w_2 + tw_1\|_X \leq (1-t)\|w_2\|_X + t\|w_1\|_X \leq 2\eta/(1-\delta)$ となるため，

$$\|\mathcal{T}(w_1) - \mathcal{T}(w_2)\|_X \leq (K+\delta)\|w_1 - w_2\|_X$$

となる．その上，仮定 $2K+\delta \leq 1$ かつ $\delta < 1$ より $K+\delta < 1$ も満たすため，$\mathcal{T}$ は $\bar{B}(0,\rho)$ から $\bar{B}(0,\rho)$ への縮小写像となる．よって，Banach の不動点定理 5.4.1 より不動点方程式 $w = \mathcal{T}(w)$ を満たす解が $\bar{B}(0,\rho)$ 内に一意に存在する．$w = u^* - \hat{u}$ であったことを思い出すと，作用素方程式 $F(u) = 0$ を満たす解が $\bar{B}(\hat{u},\rho)$ 内に一意に存在する．

最後に $\bar{B}(\hat{u}, 2\eta/(1-\delta))$ 内で一意になることを示す．$\bar{B}(0, 2\eta/(1-\delta))$ 内に不動点方程式 $w = \mathcal{T}(w)$ の解が 2 つあったとする．すなわち 2 つの解 $w_1^*, w_2^* \in \bar{B}(\hat{u}, 2\eta/(1-\delta))$ とし，$w_1^* = \mathcal{T}(w_1^*)$ と $w_2^* = \mathcal{T}(w_2^*)$ を満たすとする．そのとき，

$$\|w_1^* - w_2^*\|_X = \|\mathcal{T}(w_1^*) - \mathcal{T}(w_2^*)\|_X \leq (K+\delta)\,\|w_1^* - w_2^*\|_X$$

となる．ここで，$2K+\delta < 1$ であるため，不等式を満たすものは $w_1^* = w_2^*$ の場合のみである．すなわち，不動点方程式 $w = \mathcal{T}(w)$ を満たす解は $\bar{B}(0, 2\eta/(1-\delta))$ 内に一意である．よって，作用素方程式 $F(u) = 0$ を満たす解が $\bar{B}(\hat{u}, 2\eta/(1-\delta))$ 内に一意に存在する．

□

定理 5.4.2 では,「$F'[\hat{u}]$ は全射」であることと,「$\|I - RF'[\hat{u}]\|_{\mathcal{B}(X)} \leq \delta < 1$」を仮定することで，$F'[\hat{u}]$ が全単射であることを証明しています．有限次元の問題であれば，R は $R \approx F'[\hat{u}]^{-1}$ とすれば，簡単に作れるので，$\|I - RF'[\hat{u}]\|_{\mathcal{B}(X)} \leq \delta < 1$ は，さほど大きな問題にはなりません．しかしながら，無限次元問題の場合，R を作ることすら困難になり始めます．そのうえで，$\|I - RF'[\hat{u}]\|_{\mathcal{B}(X)} \leq \delta < 1$ を満たす保証もありません．そのために,「$F'[\hat{u}]$ は全射」であることと,「$\|I - RF'[\hat{u}]\|_{\mathcal{B}(X)} \leq \delta < 1$」とは別の方法であらかじめ，$F'[\hat{u}]$ が全単射であることを証明した上で，定理 5.4.2 を使うことも多々あります．もちろん，その場合も，定理 5.4.2 の系として次のように簡単に得られます:

系 5.4.2. X と Y を Banach 空間とする．$F : X \to Y$ を与えられた作用素とし，$\hat{u} \in X$ を与えられているとする ($\hat{u}$ はコンピュータで求めた近似解をイメージで!)．F は $\hat{u}$ で Fréchet 微分可能とし $F'[\hat{u}]$ と表記する．$F'[\hat{u}]$ は全単射であるとする．η を不等式

$$\|F'[\hat{u}]^{-1}F(\hat{u})\|_X \leq \eta$$

を満たす定数とする．

$\bar{B}(0,2\eta) = \{v \in X \mid \|v\|_X \leq 2\eta\}$ とする．F を $\bar{B}(\hat{u},2\eta)$ 上で Fréchet 微分可能であるとし，$F' : \bar{B}(\hat{u},2\eta) \to \mathcal{B}(X,Y)$ が $\bar{B}(\hat{u},2\eta)$ 上で連続であるとする．K を不等式

$$\|F'[\hat{u}]^{-1}\left(F'[\hat{u}] - F'[\hat{u}+v]\right)\|_{\mathcal{B}(X)} \leq K,\ v \in \bar{B}\,(0,2\eta)$$

を満たす定数とする．もし $2K \leq 1$ ならば,

$$\|u^* - \hat{u}\|_X \leq \frac{\eta}{1-K} =: \rho$$

に対し，真の解 u^* は $\bar{B}(\hat{u},\rho)$ 内に存在する．その上，$\bar{B}(\hat{u},2\eta)$ 内で一意である．

第6章　射影を用いた無限次元線形問題の解法

ここでは，Banach 空間 X とし，有界な線形作用素 $L \in \mathcal{B}(X)$ と $g \in X$ に対して問題

$$\text{Find } \phi \in X \text{ s.t. } L\phi = g \tag{6.1}$$

の解を求める方法を考えてみましょう．ここで求める方法は，コンピュータではなく，解析的に求める方法を指します．しかし，この問題は，系 5.4.2 の $\left\|F'[\hat{u}]^{-1}F(\hat{u})\right\|_X$ や $\left\|F'[\hat{u}]^{-1}\left(F'[\hat{u}]-F'[\hat{u}+v]\right)\right\|_{\mathcal{B}(X)}$ を求める際に利用することを想定しています．例えば，$\phi = F'[\hat{u}]^{-1}F(\hat{u})$ とすれば，

$$\text{Find } \phi \in X \text{ s.t. } F'[\hat{u}]\phi = F(\hat{u})$$

となるので，無限次元線形問題の解 ϕ を求める問題に帰着することができます．その上，解 ϕ のノルム $\|\phi\|_X$ を計算すれば系 5.4.2 の利用に一歩近づくために，無限次元非線形問題の解の品質保証法の準備になります!

6.1　直和と射影 ～計算できる部分とできない部分の分離～

コンピュータは残念ながら有限次元の計算しか実行できません．そのため，無限次元の問題も有限次元に近似して解く必要があります．ここでは，無限次元空間 X を有限次元部分と無限次元部分にわけるための道具として，一般的な Banach 空間上の射影を紹介します．

まず，Banach 空間上の直和として，代数的直和を定義します:

定義 6.1.1 (代数的直和)**.** Banach 空間 X とする．また，X_1,X_2 を X の線形部分空間とする (X_1 と X_2 は「閉」である必要がないので注意!)．ただし，X_1 と X_2 のノルムは，X のノルムと同一とする．そのとき，X が X_1 と X_2 の代数的直和であるとは

- $X = X_1 + X_2 := \{x_1 + x_2 \mid \forall x_1 \in X_1,\ \forall x_2 \in X_2\}$
- $X_1 \cap X_2 = \{0\}$ (すなわち，$x = x_1 + x_2, x_1 \in X_1, x_2 \in X_2$ が一意に表せること)

が成立することをいう.

代数的直和は分解した線形部分空間 X_1, X_2 がともに閉である必要がありません．そのため，定理 5.1.4 から X_1 や X_2 が Banach 空間ではない可能性もでてきてしまいます．X_1 や X_2 が Banach 空間ではなくなってしまったら，折角分解しても使いづらくなってしまいますね．そこで，X_1, X_2 が Banach 空間になるための必要十分条件を考えていきましょう.

それでは，まず，射影を定義します:

定義 6.1.2 (射影)**.** X をノルム空間とする．定義域を X とした X 上の線形作用素 P が

$$P^2 = P$$

となるとき，射影作用素，あるいは，単に射影と呼ぶ.

射影の定義には，Banach 空間であることも，有界性も連続性も直交性も仮定していないことに注意して下さい．では，代数的直和と射影の関係を見ていきましょう．

定理 6.1.1. Banach 空間 X がその線形部分空間 X_1,X_2 の代数的直和であるとする．そのとき，$x \in X$ について

$$x = x_1 + x_2,\ x_1 \in X_1, x_2 \in X_2$$

とし，$\mathcal{D}(P) = X$ となる X 上の線形作用素 P を

$$Px = x_1,\ \ (I - P)x = x_2$$

とすると，線形作用素 P と $I - P$ は射影となる．

証明． まず，$Px_2 = 0$ を証明する．$Px_2 \neq 0$ となる $x_2 \in X_2$ が存在すると仮定し，矛盾を示す．定義より，

$$x = x_1 + x_2 = x_1 + Px_2 + (I - P)x_2$$

に対して，P の値域 $\mathcal{R}(P) = X_1$ であることと，X_1 が線形空間であることに注意すると，$x_1 + Px_2 \in X_1$ は $x_1 + Px_2 \neq x_1$ となる．また，$\mathcal{R}(I - P) = X_2$ であるため $(I - P)x_2 \in X_2$ である．しかし，これは代数的直和の定義 $X_1 \cap X_2 = \{0\}$ より x の分解の一意性に矛盾する．よって $Px_2 = 0$ となる．

そのうえで，

$$Px_2 = P(I - P)x = Px - P^2x = 0$$

より

$$Px = P^2x$$

となるため，P は射影となる．

また，

$$(I - P)(I - P)x = (I - 2P + P^2)x = (I - 2P + P)x = (I - P)x$$

となるため，$I - P$ も射影となる．

□

定理 6.1.1 の証明中に $Px_2 = 0,\ \forall x_2 \in X_2$ を示しました．これ以外にも

$$Px_1 = x_1,\ \forall x_1 \in X_1, \qquad (I - P)x_2 = x_2,\ \forall x_2 \in X_2, \qquad (I - P)x_1 = 0,\ \forall x_1 \in X_1$$

といった性質を代数的直和の分解の一意性に対する矛盾をつく背理法で証明することができます．実際に，$Px_1 = x_1$ を証明するには，$Px_1 \neq x_1$ を仮定し，$x = x_1 + x_2 = Px_1 + (I - P)x_1 + x_2$ となるため分解の一意性に対して矛盾することを示せば良いです．

では，いよいよ，X が X_1 と X_2 の代数的直和である際に，X_1, X_2 が Banach 空間になるための必要十分条件を見てみましょう．

定理 6.1.2. Banach 空間 X が X_1 と X_2 の代数的直和であるとする．$x \in X$ について

$$x = x_1 + x_2,\ x_1 \in X_1, x_2 \in X_2$$

とし，$\mathcal{D}(P) = X$ となる X 上の線形作用素 P を

$$Px = x_1,\ \ (I - P)x = x_2$$

とすると以下が成立する:

$$P \text{が連続} \Leftrightarrow X_1, X_2 \text{が共に Banach 空間}$$

(「作用素が連続である」を忘れてしまった場合は定義 5.2.6 を思い出そう!)

証明．「P が連続 $\Rightarrow X_1, X_2$が共に Banach 空間 の証明」

まず，X_1 が Banach 空間であることを示す．X が X_1 と X_2 の代数的直和であることから，X_1 は X の線形部分空間である．そのため，X_1 が閉集合になること (すなわち，閉部分空間であること) を示せば定理 5.1.4 より，X_1 が Banach 空間であることを示せる．よって，X_1 の任意の点列 $(x_n) \subset X_1$ に対し，極限 x^* が X_1 にも属することを示せば良い．x_n は X_1 に属することから $Px_n = x_n$ になる．そのうえで，P が連続であることから，$Px_n = x_n$ の極限は $Px^* = x^*$ となる．よって，$Px^* = x^* \in X_1$ から，X_1 は閉集合となるため，Banach 空間となる．

次に，X_2 が Banach 空間であることを示す．そのために，まず，$I - P$ が連続になることを示す．P が連続であることから，$x_n \to x^*$ となる任意の $x_n \in X$ に対して，$Px_n \to Px^*$ となることに注意すると

$$(I - P)x_n = x_n - Px_n \to x^* - Px^* = (I - P)x^*$$

となるため，$I - P$ も連続となる．あとは，P と X_1 のときと同様の議論をすればよい．すなわち，X_2 の任意の点列 $(x_n) \subset X_2$ に対し，極限 x^* が X_2 にも属することを示せば良い．$I - P$ が連続であることから，$(I - P)x_n = x_n$ の極限は $(I - P)x^* = x^*$ となる．よって，$(I - P)x^* = x^* \in X_2$ から，X_2 は閉集合となるため，Banach 空間となる．

「P が連続 $\Leftarrow X_1, X_2$が共に Banach 空間 の証明」

Banach 空間 X の $x_n \to x^*$ となる任意の点列 (x_n) に対し，$Px_n \to y$ としたとき，$y = Px^*$ となることを示せば良い．上記の点列を用いて X_1 の点列 (Px_n) を作成すると，X_1 が Banach 空間であることから，閉集合であるため点列 (Px_n) の極限 y は X_1 にも属する．よって，

$$y = Py$$

となる．

さらに，同様に X_2 の点列 $((I - P)x_n)$ を作成すると，その極限は

$$(I - P)x_n = x_n - Px_n \to x^* - y$$

となり，X_2 も Banach 空間であることから，閉集合であるため点列 $((I-P)x_n)$ の極限 $x^* - y$ は X_2 にも属する．そのうえで，Banach 空間 X が X_1 と X_2 の代数的直和であることから，$X_1 \cap X_2 = \{0\}$ に注意すると $P(x^* - y) = 0$ となる．よって

$$0 = P(x^* - y) = Px^* - Py = Px^* - y$$

となるため,

$$Px^* = y$$

となる.

□

最後に，位相的直和を定義しましょう.

定義 6.1.3 (位相的直和)**.** Banach 空間 X が X_1 と X_2 の代数的直和であるとする．その上，X_1 と X_2 が共に Banach 空間であるとき，Banach 空間 X が X_1 と X_2 の位相的直和であるという.

もちろん，位相的直和は射影の連続性を用いて定義することも可能です.

6.2 射影を用いた Banach 空間上のガウスの消去法

本節では，有界な線形作用素 $L \in \mathcal{B}(X)$ と $g \in X$ で構成される問題 (6.1) に対してガウスの消去法で解 ϕ が存在するための条件と解を求める方法を紹介します．まず，Banach 空間 X が X_1 と X_2 の代数的直和であるとし，$x \in X$ について

$$x = x_1 + x_2,\ x_1 \in X_1, x_2 \in X_2$$

とし，P を $\mathcal{D}(P) = X$

$$Px = x_1,\quad (I-P)x = x_2$$

を満たす射影とします．特段，強い仮定を設けずに議論は進めますが，コンピュータで解くことを意識する際には，X_1 がコンピュータで解ける有限次元部分，X_2 がコンピュータで解けない無限次元部分と心の中に留めておく良いです.

まず，はじめに，Banach 空間 X が X_1 と X_2 の代数的直和であることから，$x = x_1 + x_2, x_1 \in X_1, x_2 \in X_2$ が一意に表せることを利用し，射影を使って問題 (6.1) を変形させます:

$$L\phi = g \Leftrightarrow \begin{cases} PL\phi = Pg \\ (I-P)L\phi = (I-P)g \end{cases}$$

さらに，解 ϕ も射影を使って次のように分解します:

$$\phi = \phi_1 + \phi_2,\quad \phi_1 := P\phi,\quad \phi_2 := (I-P)\phi$$

この分解した解を利用すると

$$\begin{cases} PL(\phi_1 + \phi_2) = Pg \\ (I-P)L(\phi_1 + \phi_2) = (I-P)g \end{cases}$$

となります．ここで，4 つの線形作用素をそれぞれ

$$\begin{aligned} T :=& PL|_{X_1} : X_1 \to X_1, & B :=& PL|_{X_2} : X_2 \to X_1 \\ C :=& (I-P)L|_{X_1} : X_1 \to X_2, & D :=& (I-P)L|_{X_2} : X_2 \to X_2 \end{aligned} \tag{6.2}$$

と定義すると，次のように変形できます:

$$\begin{cases} T\phi_1 + B\phi_2 = Pg \\ C\phi_1 + D\phi_2 = (I-P)g \end{cases}$$

この問題はある種，連立一次方程式として見ることができるため，作用素行列

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix} : X_1 \times X_2 \to X_1 \times X_2$$

を定義すると

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

のように作用素行列の方程式に帰着できます．そのために，いくつかの仮定を加えてこの方程式をガウスの消去法 (正確に言うとブロックガウスの消去法) と同じ手順で計算すると次の定理が得られます:

定理 6.2.1. X, X_1, X_2, L, g, P, ϕ_1, ϕ_2, T, B, C, D をすべて本節で定義したものとする．線形作用素 T を全単射であると仮定する．線形作用素 S を

$$S := D - CT^{-1}B : \ X_2 \to X_2 \tag{6.3}$$

とする．もし，S が全単射ならば,

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} T^{-1} + T^{-1}BS^{-1}CT^{-1} & -T^{-1}BS^{-1} \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix} \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

となり，有界線形作用素 L は全単射である．

証明． 方程式

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

に対して，T が全単射であることから左から

$$\begin{pmatrix} T^{-1} & 0 \\ 0 & I_{X_2} \end{pmatrix}$$

を掛けると

$$\begin{pmatrix} I_{X_1} & T^{-1}B \\ C & D \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} T^{-1} & 0 \\ 0 & I_{X_2} \end{pmatrix} \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

となる．次に，左から

$$\begin{pmatrix} I_{X_1} & 0 \\ -C & I_{X_2} \end{pmatrix}$$

を掛けると

$$\begin{pmatrix} I_{X_1} & T^{-1}B \\ 0 & S \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} I_{X_1} & 0 \\ -C & I_{X_2} \end{pmatrix}\begin{pmatrix} T^{-1} & 0 \\ 0 & I_{X_2} \end{pmatrix}\begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$
$$= \begin{pmatrix} T^{-1} & 0 \\ -CT^{-1} & I_{X_2} \end{pmatrix}\begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

となる．次に，S が全単射であることから左から

$$\begin{pmatrix} I_{X_1} & 0 \\ 0 & S^{-1} \end{pmatrix}$$

を掛けると

$$\begin{pmatrix} I_{X_1} & T^{-1}B \\ 0 & I_{X_2} \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} T^{-1} & 0 \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix}\begin{pmatrix} Pg \\ (I-P)g \end{pmatrix} \tag{6.4}$$

となる．最後に，

$$\begin{pmatrix} I_{X_1} & -T^{-1}B \\ 0 & I_{X_2} \end{pmatrix}$$

を左から掛けると

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} I_{X_1} & -T^{-1}B \\ 0 & I_{X_2} \end{pmatrix}\begin{pmatrix} T^{-1} & 0 \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix}\begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$
$$= \begin{pmatrix} T^{-1}+T^{-1}BS^{-1}CT^{-1} & -T^{-1}BS^{-1} \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix}\begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

が得られる．これは，

$$\begin{pmatrix} T^{-1}+T^{-1}BS^{-1}CT^{-1} & -T^{-1}BS^{-1} \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix}\begin{pmatrix} T & B \\ C & D \end{pmatrix} = \begin{pmatrix} I_{X_1} & 0 \\ 0 & I_{X_2} \end{pmatrix}$$

を意味する．逆に

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix}\begin{pmatrix} T^{-1}+T^{-1}BS^{-1}CT^{-1} & -T^{-1}BS^{-1} \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} I_{X_1} & T^{-1}B+T^{-1}BS^{-1}CT^{-1}B-T^{-1}BS^{-1}D \\ 0 & -S^{-1}CT^{-1}B+S^{-1}D \end{pmatrix}$$
$$= \begin{pmatrix} I_{X_1} & T^{-1}B-T^{-1}BS^{-1}(D-CT^{-1}B) \\ 0 & S^{-1}(D-CT^{-1}B) \end{pmatrix} = \begin{pmatrix} I_{X_1} & 0 \\ 0 & I_{X_2} \end{pmatrix}$$

となるため，定義 5.2.4 と定理 5.2.1 から

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix}$$

は単射である．また，任意の $g_1 \in X_1$ と $g_2 \in X_2$ に対して，

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$

は

$$\begin{aligned} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} &= \begin{pmatrix} I_{X_1} & -T^{-1}B \\ 0 & I_{X_2} \end{pmatrix} \begin{pmatrix} T^{-1} & 0 \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix} \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix} \\ &= \begin{pmatrix} T^{-1} + T^{-1}BS^{-1}CT^{-1} & -T^{-1}BS^{-1} \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \end{aligned}$$

となる解

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \in X_1 \times X_2$$

を持つため，

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix} : X_1 \times X_2 \to X_1 \times X_2$$

は全射でもある．

続いて，L が単射であることを示す．定理 5.2.2 から，$L\phi = 0$ において，解が $\phi = 0$ だけであることを示せば良い．X が X_1 と X_2 の代数的直和であることから，解 ϕ は $P\phi$ と $(I-P)\phi$ に一意に分解できる．その上，

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix} \begin{pmatrix} P\phi \\ (I-P)\phi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

となる．ここで，

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix}$$

が全単射であることから，解は $P\phi = 0$, $(I-P)\phi = 0$ のみである．よって，$L\phi = 0$ において，解は $\phi = 0$ のみである．

最後に，L が全射であることを示す．任意の $g \in X$ に対して $L\phi = g$ を満たす解 $\phi \in X$ が存在すれば，定義 5.2.3 から L が全射であることがいえる．X が X_1 と X_2 の代数的直和であることから，解 ϕ は $P\phi$ と $(I-P)\phi$ に一意に分解できる．その上，

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix} \begin{pmatrix} P\phi \\ (I-P)\phi \end{pmatrix} = \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

となる．ここで，

$$\begin{pmatrix} T & B \\ C & D \end{pmatrix}$$

が全単射であることから，解 $(P\phi, (I-P)\phi) \in X_1 \times X_2$ は常に存在する．よって，$\phi = P\phi + (I-P)\phi \in X$ であるため，任意の $g \in X$ に対して解 $\phi \in X$ は存在する．

□

定理 6.2.1 は冒頭にも記載した通り，線形・非線形問わず，無限次元問題の解の品質保証に利用します．では，定理 6.2.1 の十分条件である S が全単射，というのは L の全単射性にどれくらい重要なのでしょうか? 簡単にいうと，必要条件になるのか?という疑問が生まれてきます．それは，次の定理で解決されます:

定理 6.2.2. X, X_1, X_2, L, g, P, ϕ_1, ϕ_2, T, B, C, D をすべて本節で定義したものとする．線形作用素 T を全単射であると仮定する．そのとき，

$$S \text{ が全単射} \Leftrightarrow L \text{ が全単射}$$

証明 . 「S が全単射 $\Rightarrow$ L が全単射」は定理 6.2.1 でいえているため,「S が全単射 $\Leftarrow$ L が全単射」のみ示す．L が全単射であることから，$L\phi = 0$ を満たす解は $\phi = 0$ のみである．また，X が X_1 と X_2 の代数的直和であることから，解 $\phi = 0$ は $P\phi = 0$ と $(I-P)\phi = 0$ に一意に分解できる．その上，T が全単射であることを利用すると (6.4) が得られるため，以下のようになる:

$$\begin{aligned} L\phi = 0 &\Leftrightarrow \begin{pmatrix} T & B \\ C & D \end{pmatrix} \begin{pmatrix} P\phi \\ (I-P)\phi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow \begin{pmatrix} I_{X_1} & T^{-1}B \\ 0 & S \end{pmatrix} \begin{pmatrix} P\phi \\ (I-P)\phi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

上記の方程式を満たす解が，$P\phi = 0$ と $(I-P)\phi = 0$ のみのため，

$$\begin{pmatrix} I_{X_1} & T^{-1}B \\ 0 & S \end{pmatrix}$$

は単射である．ここで，S が単射でないと仮定して，矛盾を示す．S が単射でないことから，

$$S\phi_2 = 0$$

を満たす 0 以外の解 $\phi_2 \in X_2$ が存在する．そのうえで，

$$\phi_1 = -T^{-1}B\phi_2$$

とすると (ϕ_1, ϕ_2) は方程式

$$\begin{pmatrix} I_{X_1} & T^{-1}B \\ 0 & S \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

の解となる．しかし，$L\phi = 0$ を満たす解は $\phi = 0$ のみであり，

$$\begin{pmatrix} I_{X_1} & T^{-1}B \\ 0 & S \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

を満たす解は 0 のみであるため矛盾する．よって，S は単射である．

続いて，L が全単射のとき，S が全射であることを示す．L が全単射であるため，任意の $g \in X$ に対して

$$\phi = L^{-1}g$$

となる解 $\phi \in X$ が一意に存在する．よって，代数的直和による分解の一意性から $P\phi \in X_1$ と $(I-P)\phi \in X_2$ が一意に存在する．そのうえで，

$$L\phi = g \Leftrightarrow \begin{pmatrix} T & B \\ C & D \end{pmatrix} \begin{pmatrix} P\phi \\ (I-P)\phi \end{pmatrix} = \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$
$$\Leftrightarrow \begin{pmatrix} I_{X_1} & T^{-1}B \\ 0 & S \end{pmatrix} \begin{pmatrix} P\phi \\ (I-P)\phi \end{pmatrix} = \begin{pmatrix} T^{-1} & 0 \\ -CT^{-1} & I_{X_2} \end{pmatrix} \begin{pmatrix} Pg \\ (I-P)g \end{pmatrix}$$

となる．さらに，第二式

$$S(I-P)\phi = -CT^{-1}Pg + (I-P)g$$

となる．ここで，任意の $g \in X$ に対して $L\phi = g$ となる解 $\phi \in X$ が存在し，$\phi = P\phi + (I-P)\phi$ のように一意に分解できることに注意すると，任意の $g_2 \in X_2$ に対しても $L\phi = g_2$ となる解 $\phi \in X$ が存在し，$\phi = P\phi + (I-P)\phi$ となる $(I-P)\phi \in X_2$ が存在する．そのうえで，$Pg_2 = -$ より

$$S(I-P)\phi = -CT^{-1}Pg_2 + (I-P)g_2 = g_2$$

となる解 $(I-P)\phi \in X_2$ が存在するため，S は全射である．

□

6.3 準直交射影と $\|L^{-1}\|_{\mathcal{B}(X)}$ の評価方法

本節では Banach 空間 X が X_1 と X_2 の代数的直和になる際に，代数的直和から誘導される射影 P と $I-P$ に対して，さらに制限をつけた際の $\|L^{-1}\|_{\mathcal{B}(X)}$ の評価方法を紹介します．その制限というのが次に定義する準直交射影です．準直交射影はこの本の特有の言い回しであり，一般的な用語ではないので注意してください．

定義 6.3.1 (準直交射影)**.** Banach 空間 X がその線形部分空間 X_1,X_2 の代数的直和であるとする．そのとき，$x \in X$ について

$$x = x_1 + x_2,\ x_1 \in X_1, x_2 \in X_2$$

とし，$\mathcal{D}(P) = X$ となる X 上の射影作用素 P を

$$Px = x_1,\ \ (I-P)x = x_2$$

とする．そのとき，任意の $x \in X$ に対して

$$\|x\|_X^2 = \|Px\|_X^2 + \|(I-P)x\|_X^2$$

が成立するとき P を準直交射影と呼ぶ.

定義 6.3.1 を端的に言うとピタゴラスの定理 $\|u\|_X^2 = \|Pu\|_X^2 + \|(I-P)u\|_X^2$ が成立するような射影のことを準直交射影[1]と呼ぶようにしています.

[1]本書ではまだ定義していませんが Hirbert 空間において内積から作成した直交射影ならばピタゴラスの定理が成立するために，準直交射影になります．Banach 空間上の空間同士の直交関係は，内積の代わりに汎関数や共役対を利用して定義することができます．しかし，Banach 空間のみで直交射影を定義し，ピタゴラスの定理まで成立する，といった証明を私は見たことがありません．また,「準直交射影 $\Rightarrow$ 直交射影」という証明も見たことがありません．その一方で，本書で利用したいのは，必ずしも直交射影である必要はなく，射影 P の制限としてピタゴラスの定理 $\|u\|_X^2 = \|Pu\|_X^2 + \|(I-P)u\|_X^2$ のみ課されていれば十分です．そこで，本書では直交射影は定義せずに，ピタゴラスの定理 $\|u\|_X^2 = \|Pu\|_X^2 + \|(I-P)u\|_X^2$ のみを課した準直交射影を定義しています．Hirbert 空間では「直交射影 $\Leftrightarrow$ 準直交射影」が成立する気もするので，興味がある方は証明してみてください.

2.1 節で定義した有限次元のベクトルの 2 ノルムを利用することで，$\|u\|_X^2 = \|Pu\|_X^2 + \|(I-P)u\|_X^2$ は

$$\|u\|_X = \left\| \begin{pmatrix} \|Pu\|_{X_1} \\ \|(I-P)u\|_{X_2} \end{pmatrix} \right\|_2$$

とも書けることに注意してください.

それでは，まず，$\|L^{-1}\|_{\mathcal{B}}(X)$ の上界の評価を示します:

定理 6.3.1. 定理 6.2.1 の記号を利用する．Banach 空間 X は X_1 と X_2 の位相的直和であるとする．さらに，P を準直交射影とし，線形作用素 T と S はそれぞれ全単射であるとする．そのとき,

$$\|L^{-1}\|_{\mathcal{B}(X)} \leq \left\| \begin{pmatrix} \|T^{-1} + T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} & \|T^{-1}BS^{-1}\|_{\mathcal{B}(X_2,X_1)} \\ \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)} & \|S^{-1}\|_{\mathcal{B}(X_2)} \end{pmatrix} \right\|_2$$

となる (行列の 2 ノルムは 2.1 節で定義されています).

証明 . 作用素ノルムの定義 (5.1) から

$$\|L^{-1}\|_{\mathcal{B}(X)} = \sup_{g \in X \setminus \{0\}} \frac{\|L^{-1}g\|_X}{\|g\|_X}$$

となる．$\phi = L^{-1}g$ とおき，$\phi_1 = P\phi$, $\phi_2 = (I-P)\phi$ とすると，P が準直交射影であることから

$$\begin{aligned} \|L^{-1}\|_{\mathcal{B}(X)} &= \sup_{g \in X \setminus \{0\}} \frac{\|\phi\|_X}{\|g\|_X} \\ &= \sup_{g \in X \setminus \{0\}} \frac{\left\| \begin{pmatrix} \|\phi_1\|_{X_1} \\ \|\phi_2\|_{X_2} \end{pmatrix} \right\|_2}{\|g\|_X} \end{aligned}$$

となる．そのうえで，定理 6.2.1 から

$$\phi_1 = (T^{-1} + T^{-1}BS^{-1}CT^{-1})Pg - T^{-1}BS^{-1}(I-P)g$$

と

$$\phi_2 = -S^{-1}CT^{-1}Pg + S^{-1}(I-P)g$$

となるため,

$$\|\phi_1\|_{X_1} \leq \|T^{-1} + T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)}\|Pg\|_{X_1} + \|T^{-1}BS^{-1}\|_{\mathcal{B}(X_2,X_1)}\|(I-P)g\|_{X_2}$$

と

$$\|\phi_2\|_{X_2} \leq \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)}\|Pg\|_{X_1} + \|S^{-1}\|_{\mathcal{B}(X_2)}\|(I-P)g\|_{X_2}$$

となる (ここで，X が X_1 と X_2 の位相的直和ではなく，代数的直和のままだと X_1 と X_2 が Banach 空間ではない可能性がある出てくるため作用素ノルムで括り出せなくなります)．よって，

$$\begin{aligned}
&\left\| \begin{pmatrix} \|\phi_1\|_{X_1} \\ \|\phi_2\|_{X_2} \end{pmatrix} \right\|_2 \\
\leq &\left\| \begin{pmatrix} \|T^{-1}+T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)}\|Pg\|_{X_1} + \|T^{-1}BS^{-1}\|_{\mathcal{B}(X_2,X_1)}\|(I-P)g\|_{X_2} \\ \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)}\|Pg\|_{X_1} + \|S^{-1}\|_{\mathcal{B}(X_2)}\|(I-P)g\|_{X_2} \end{pmatrix} \right\|_2 \\
= &\left\| \begin{pmatrix} \|T^{-1}+T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} & \|T^{-1}BS^{-1}\|_{\mathcal{B}(X_2,X_1)} \\ \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)} & \|S^{-1}\|_{\mathcal{B}(X_2)} \end{pmatrix} \begin{pmatrix} \|Pg\|_{X_1} \\ \|(I-P)g\|_{X_2} \end{pmatrix} \right\|_2 \\
\leq &\left\| \begin{pmatrix} \|T^{-1}+T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} & \|T^{-1}BS^{-1}\|_{\mathcal{B}(X_2,X_1)} \\ \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)} & \|S^{-1}\|_{\mathcal{B}(X_2)} \end{pmatrix} \right\|_2 \left\| \begin{pmatrix} \|Pg\|_{X_1} \\ \|(I-P)g\|_{X_2} \end{pmatrix} \right\|_2 \\
= &\left\| \begin{pmatrix} \|T^{-1}+T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} & \|T^{-1}BS^{-1}\|_{\mathcal{B}(X_2,X_1)} \\ \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)} & \|S^{-1}\|_{\mathcal{B}(X_2)} \end{pmatrix} \right\|_2 \|g\|_X
\end{aligned}$$

となる．よって，

$$\|L^{-1}\|_{\mathcal{B}(X)} \leq \left\| \begin{pmatrix} \|T^{-1}+T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} & \|T^{-1}BS^{-1}\|_{\mathcal{B}(X_2,X_1)} \\ \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)} & \|S^{-1}\|_{\mathcal{B}(X_2)} \end{pmatrix} \right\|_2$$

となる．

□

続いて，$\|L^{-1}\|_{\mathcal{B}(X)}$ の下界評価を行います:

定理 6.3.2. 定理 6.2.1 の記号を利用する．Banach 空間 X は X_1 と X_2 の位相的直和であるとする．さらに，P を準直交射影とし，線形作用素 T と S はそれぞれ全単射であるとする．そのとき，

$$\|L^{-1}\|_{\mathcal{B}(X)} \geq \|T^{-1}\|_{\mathcal{B}(X_1)} - \|T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} - \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)}$$

となる．

証明． 作用素ノルムの定義より

$$\|T^{-1}\|_{\mathcal{B}(X_1)} = \sup_{g\in X_1} \frac{\|T^{-1}g\|_{X_1}}{\|g\|_{X_1}}$$

となる．そのうえで，$g_1 \in X_1$ を上記の sup を満たす元とすると

$$\|T^{-1}\|_{\mathcal{B}(X_1)} = \frac{\|T^{-1}g_1\|_{X_1}}{\|g_1\|_{X_1}}$$

となる．続いて，作用素ノルムの定義と $g_1 \in X_1 \subset X$ であることに注意すると

$$\begin{aligned}
\|L^{-1}\|_{\mathcal{B}(X)} = &\sup_{g\in X\setminus\{0\}} \frac{\|L^{-1}g\|_X}{\|g\|_X} \\
\geq &\frac{\|L^{-1}g_1\|_X}{\|g_1\|_X}
\end{aligned}$$

となる．さらに，$g_1 \in X_1 \subset X$ から $Pg_1 = g_1$ と $(I-P)g_1 = 0$ になることに注意して定理 6.2.1 を利用すると

$$\begin{aligned} L^{-1}g_1 =& (T^{-1} + T^{-1}BS^{-1}CT^{-1} - S^{-1}CT^{-1})Pg_1 \\ =& (T^{-1} + T^{-1}BS^{-1}CT^{-1} - S^{-1}CT^{-1})g_1 \end{aligned}$$

となる．よって

$$\begin{aligned} \|L^{-1}\|_{\mathcal{B}(X)} \geq& \frac{\|(T^{-1} + T^{-1}BS^{-1}CT^{-1} - S^{-1}CT^{-1})g_1\|_X}{\|g_1\|_X} \\ \geq& \frac{\|T^{-1}g_1\|_X}{\|g_1\|_X} - \frac{\|T^{-1}BS^{-1}CT^{-1}g_1\|_X}{\|g_1\|_X} - \frac{\|S^{-1}CT^{-1}g_1\|_X}{\|g_1\|_X} \\ \geq& \frac{\|T^{-1}g_1\|_X}{\|g_1\|_X} - \|T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} - \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)} \\ =& \|T^{-1}\|_{\mathcal{B}(X_1)} - \|T^{-1}BS^{-1}CT^{-1}\|_{\mathcal{B}(X_1)} - \|S^{-1}CT^{-1}\|_{\mathcal{B}(X_1,X_2)} \end{aligned}$$

を得る．

□

第7章　無限次元非線形問題の解法のエッセンス～Aが全単射の場合～

本章では，X, Y を Banach 空間とし，線形作用素 $A \in \mathcal{B}(X, Y)$ と非線形作用素 $N : X \to Y$ とし

$$F(u) := Au - N(u)$$

とした際，抽象的な非線形問題

$$\text{Find } u \in X \text{ s.t. } F(u) = 0$$

に対し，解の品質を保証する方法を紹介します．ここでは定理 6.2.1 をどのように利用して基本定理の系 5.4.2 の η や K を計算するのか外観を紹介します．具体的には A や N が決定しないと計算できない部分までを導入する予定です．

この章では以下の仮定や定義を利用します:

- A は全単射とする．
- X_N を X の有限次元部分空間とし，X が X_N と X_c の位相的直和となる X_c が存在する．
- $P_N : X \to X_N$ を任意の $x \in X$ に対し，$x = x_1 + x_2,\ x_1 \in X_N,\ x_2 \in X_c$ とした際の射影 $Px = x_1,\ (I - P)x = x_2$ とする．
- P_N は準直交射影とする，すなわち，$\|u\|_X^2 = \|P_N u\|_X^2 + \|(I - P)u\|_X^2 \forall u \in X$ が成立する．
- $u_N \in X_N$ を有限次元問題

  $$\text{Find } u_N \in X_N \text{ s.t. } P_N A^{-1} F(u_N) = 0$$

  を満たす真の解とし，系 5.4.2 の $\hat{u} = u_N$ として利用する．
- 非線形作用素 N は u_N で Fréchet 微分可能とし，$N'[u_N] \in \mathcal{B}(X)$ と表記する
- $F'[u_N] := A - N'[u_N]$ とする．

7.1　$F'[\hat{u}]$ の全単射性の確認方法

系 5.4.2 を利用するために，まず，$F'[u_N]$ が全単射であることを確かめなければいけません．そのために，A が全単射であることを利用して

$$F'[u_N]\phi = g \Leftrightarrow A^{-1}F'[u_N]\phi = A^{-1}g$$

とし，定理 6.2.1 の L を

$$L := A^{-1}F'[u_N] \in \mathcal{B}(X)$$

として利用します．そのとき，例えば，(6.2) の B は

$$B := P_N(A^{-1}F'[u_N])|_{X_c} = P_N(I_X - A^{-1}F'[u_N])|_{X_c} = P_N|_{X_c} - P_N A^{-1}F'[u_N]|_{X_c}$$

となる．その上，X は X_N と X_c の位相的直和になるため，$X_N \cap X_c = \{0\}$ から $P_N|_{X_c} = 0$ となるため $B = -P_N A^{-1}F'[u_N]|_{X_c}$ となる．(6.2) の T, C, D も同様に計算すると

$$\begin{aligned} T :=& P_N A^{-1}F'[u_N]|_{X_N} : X_N \to X_N, & B :=& -P_N A^{-1}N'[u_N]|_{X_c} : X_c \to X_N \\ C :=& -(I - P_N)A^{-1}N'[u_N]|_{X_c} : X_N \to X_c, & D :=& I_{X_c} - (I - P_N)A^{-1}N'[u_N]|_{X_c} : X_c \to X_c \end{aligned}$$

となる．ここで $(I - P_N)|_{X_c}$ は X_c 上の恒等作用素 I_{X_c} となる．

T と (6.3) で定義される S が全単射であれば，定理 6.2.1 から $A^{-1}F'[u_N]$ が全単射となります．

T は有限次元 X_N から X_N への有限次元作用素になるため，多くの場合，T が全単射であることを確認する方法は，行列の正則性を確認する問題に帰着されます．そのために，定理 3.1.1 や定理 3.3.3 内に記載されている行列の正則性を保証する十分条件を確認すれば良いです．行列の正則性の問題に帰着する方法は，具体的な問題によって変わるため，具体例で紹介します．

一方で，S は無限次元部分 X_c から X_c への無限次元部分の作用素になるために，コンピュータだけで解くことはできません．そこで，Neumann 級数の定理 5.2.8 を利用します．仮定から X が X_N と X_c の位相的直和であるため，X_c は Banach 空間となります．そのうえで，X_c 上の作用素 S に対して Neumann 級数の定理 5.2.8 を利用すると

$$\|I_{X_c} - S\|_{\mathcal{B}(X_c)} < 1$$

であれば，S は全単射となります．そこで，$\|I_{X_c} - S\|_{\mathcal{B}(X_c)}$ の評価が重要になります．その上で，変形すると

$$\|I_{X_c} - S\|_{\mathcal{B}(X_c)} = \|I_{X_c} - (D - CT^{-1}B)\|_{\mathcal{B}(X_c)} = \|(I - P_N)A^{-1}N'[u_N]|_{X_c} + CT^{-1}B\|_{\mathcal{B}(X_c)}$$

となります．そこで，

$$\begin{aligned} \|(I - P_N)A^{-1}N'[u_N]\phi_c\|_X \leq& C_1\|\phi_c\|_X, \ \forall \phi_c \in X_c \\ \|CT^{-1}B\phi_c\|_X \leq& C_2\|\phi_c\|_X, \ \forall \phi_c \in X_c \end{aligned}$$

のような定数 C_1 と C_2 を求められれば，$C_1 + C_2 < 1$ を確認すれば良いです．C_1 と C_2 の具体的な計算方法は問題によって変わるため，具体例の際に紹介します．ここで C_1 や C_2 は値が小さくなるのか?という疑問がわきます．想定としては，有限次元部分 X_N が十分に近似できているのであれば，$I - P_N$ が小さくなる，ということを見込んでいます．特に，有限次元部分 X_N の次元が上がれば，C_1 や C_2 という値が小さくなることを想定しています．

さらに，Neumann 級数の定理 5.2.8 の十分条件が確認取れれば

$$\|S^{-1}\|_{\mathcal{B}(X_c)} \leq \frac{1}{1 - (C_1 + C_2)}$$

と評価できることも利用します．

上記 T と S が共に全単射であることがいえれば，定理 6.2.1 から $A^{-1}F'[u_N]$ が全単射がいえます．そのうえで，$A \in \mathcal{B}(X, Y)$ も全単射であるため，本節の目的であった $F'[u_N]$ の全単射性もいえます (この証明は難しくないのでチャレンジしてみましょう!)．

7.2　系 5.4.2 の定数 η の計算方法

7.1 節により系 5.4.2 の $F'[\hat{u}]$ の全単射性まで確認がとれました．次は，いよいよ系 5.4.2 の

$$\|F'[\hat{u}]^{-1}F(\hat{u})\|_X \leq \eta$$

の計算方法です．まず,

$$\phi := F'[\hat{u}]^{-1}F(\hat{u})$$

とおくと，$\|F'[\hat{u}]^{-1}F(\hat{u})\|_X = \|\phi\|_X \leq \eta$ となる．そのとき，A と $F'[\hat{u}]$ が全単射であることから

$$\begin{aligned}&\text{Find } \phi \in X \text{ s.t. } F'[\hat{u}]\phi = F(\hat{u})\\ \Leftrightarrow&\text{Find } \phi \in X \text{ s.t. } A^{-1}F'[\hat{u}]\phi = A^{-1}F(\hat{u})\end{aligned}$$

と書けます．そのうえで，$L = A^{-1}F'[\hat{u}]$ とおけば，7.1 節と同じ状況になります．7.1 節では T と S の全単射性を既に確認しており，定理 6.2.1 を利用していましたね．同様に，定理 6.2.1 を利用すると

$$\begin{pmatrix} P_N\phi \\ (I-P_N)\phi \end{pmatrix} = \begin{pmatrix} T^{-1}+T^{-1}BS^{-1}CT^{-1} & -T^{-1}BS^{-1} \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix} \begin{pmatrix} PA^{-1}F(u_N) \\ (I-P)A^{-1}F(u_N) \end{pmatrix}$$

を得ることができます．さらに，仮定より $P_NA^{-1}F(u_N) = 0$ となるため,

$$\begin{aligned}\begin{pmatrix} P_N\phi \\ (I-P_N)\phi \end{pmatrix} &= \begin{pmatrix} T^{-1}+T^{-1}BS^{-1}CT^{-1} & -T^{-1}BS^{-1} \\ -S^{-1}CT^{-1} & S^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ (I-P_N)A^{-1}F(u_N) \end{pmatrix}\\ &= \begin{pmatrix} -T^{-1}BS^{-1}(I-P_N)A^{-1}F(u_N) \\ S^{-1}(I-P_N)A^{-1}F(u_N) \end{pmatrix}\end{aligned}$$

のように得られます．よって，

$$\begin{aligned}\|F'[\hat{u}]^{-1}F(\hat{u})\|_X^2 =&\|\phi\|_X^2 = \|P_N\phi\|_{X_h}^2 + \|(I-P_N)\phi\|_{X_c}^2\\ =&\| - T^{-1}BS^{-1}(I-P_N)A^{-1}F(u_N)\|_{X_h}^2 + \|S^{-1}(I-P_N)A^{-1}F(u_N)\|_{X_c}^2\\ \leq&\|T^{-1}BS^{-1}\|_{\mathcal{B}(X_c,Xh)}^2\|(I-P_N)A^{-1}F(u_N)\|_{X_c}^2 + \|S^{-1}\|_{\mathcal{B}(X_c)}^2\|(I-P_N)A^{-1}F(u_N)\|_{X_c}^2\\ \leq&\left(1+\|T^{-1}B\|_{\mathcal{B}(X_c,X_N)}^2\right)\|S^{-1}\|_{\mathcal{B}(X_c)}^2\|(I-P_N)A^{-1}F(u_N)\|_{X_c}^2\\ \leq&\left(\frac{1}{1-(C_1+C_2)}\right)^2\left(1+\|T^{-1}B\|_{\mathcal{B}(X_c,X_N)}^2\right)\|(I-P_N)A^{-1}F(u_N)\|_{X_c}^2\end{aligned}$$

から

$$\|F'[\hat{u}]^{-1}F(\hat{u})\|_X \leq \frac{1}{1-(C_1+C_2)}\sqrt{1+\|T^{-1}B\|_{\mathcal{B}(X_c,X_N)}^2}\|(I-P_N)A^{-1}F(u_N)\|_{X_c}$$

となります.

また，$\|T^{-1}B\|_{\mathcal{B}(X_c,X_N)}$ は C_2 の計算に含まれているため，計算ができると思います．そのために，ここで新たな評価は

$$\|(I-P_N)A^{-1}F(u_N)\|_{X_c} \leq \delta$$

となる δ の計算になります．ここで，$F(u_N)$ は残差と呼びます．$F(u_N)$ は問題によっては簡単に計算できますが，問題によっては計算が複雑になる場合もあります．特に，有限要素法を使う場合は u_N が区分的な関数になるため，領域全体では微分が計算ができない，ということさえあり得ます.

7.3 系 5.4.2 の定数 K の計算方法

前節により系5.4.2の定数ηの計算まででできました．これにより，閉球$\bar{B}(0,2\eta)=\{v\in X\mid \|v\|_X\leq 2\eta\}$を作成できました．続いて系 5.4.2 の

$$\|F'[\hat{u}]^{-1}\left(F'[\hat{u}]-F'[\hat{u}+v]\right)\|_{\mathcal{B}(X)}\leq K,\ v\in\bar{B}\left(0,2\eta\right)$$

を満たす定数 K の計算です．定数 K の計算は基本的には $N'[u_N]$ の具体的な形がわかったあとに主となる計算が必要になります．$N'[u_N]$ の具体的な形がわかる前までの計算は以下の通りになります:

$$\begin{aligned}&\|F'[\hat{u}]^{-1}\left(F'[\hat{u}]-F'[\hat{u}+v]\right)\|_{\mathcal{B}(X)}\\ =&\|F'[\hat{u}]^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)\|_{\mathcal{B}(X)}\\ =&\sup_{g\in X\setminus\{0\}}\frac{\|F'[\hat{u}]^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)g\|_{\mathcal{B}(X)}}{\|g\|_X}\end{aligned}$$

とし，$\phi:=F'[\hat{u}]^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)g$ と置くと

$$\begin{aligned}&F'[\hat{u}]\phi=\left(N'[\hat{u}]-N'[\hat{u}+v]\right)g\\ \Leftrightarrow&A^{-1}F'[\hat{u}]\phi=A^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)g\\ \Leftrightarrow&L\phi=A^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)g\\ \Leftrightarrow&\phi=L^{-1}A^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)g\end{aligned}$$

となるため,

$$\begin{aligned}\|F'[\hat{u}]^{-1}\left(F'[\hat{u}]-F'[\hat{u}+v]\right)\|_{\mathcal{B}(X)}=&\|L^{-1}A^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)\|_{\mathcal{B}(X)}\\ \leq&\|L^{-1}\|_{\mathcal{B}(X)}\|A^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)\|_{\mathcal{B}(X)}\end{aligned}$$

となります．そのうえで，$\|L^{-1}\|_{\mathcal{B}(X)}$ の計算は定理 6.3.1 を利用することができます．その一方で，$\|A^{-1}\left(N'[\hat{u}]-N'[\hat{u}+v]\right)\|_{\mathcal{B}(X)}$ は $N'[u_N]$ の具体的な形が必要になります．

ここでは定理 6.2.1 を直接利用せずに $\|L^{-1}\|_{\mathcal{B}(X)}$ を使っています．もちろん定理 6.2.1 を利用した評価方法を考えることも可能です．しかし，K の計算に関して言えば，η ほど直接計算するメリットは少なくなるため，多少の過大評価をして楽に計算する方法を紹介しています．