大阪大学基礎工学部
平成 24 年度第 1 学期 木曜日 5 時限 (B103)

数値計算 (090440)

永原 正章

目次

第 1 章

数値計算ソフトウェア

1.1 Scilab

数値計算はコンピュータを用いることを前提として，さまざまな計算法や近似法を探る学問である．したがって，実際にコンピュータを用いて計算してみるのは良い勉強法であり，理解も深まる．コンピュータを用いて数値計算を行うためには，プログラミングが必要である．プログラミング言語として C 言語や FORTRAN は非常に良くできた言語であるが，数値計算のプログラムを一から組み立てるのは大変である．

そこで，数値計算専用のソフトウェアを使用するのが良い．数値計算ソフトウェアには，MATLAB や Mathematica[*1]などがあるが，学生にとっては非常に高価である．したがって，本書では Scilab （「サイラブ」と発音する）というフリーウェアを推薦する．Scilab は，INRIA (Institute National de Recherche en Informatique et Automatique; フランス国立情報学自動制御研究所) で開発されたフリーウェアである．常にバージョンアップがほどこされており，計算結果もかなり信頼でき，非常に良くできたソフトである．また，グラフの表示も非常に簡単である．

Scilab は次の WEB ページよりダウンロードできる．

`http://www.scilab.org`

ぜひダウンロードして使ってみて欲しい．

1.2 プログラムの実行方法

本書では，いろいろな Scilab プログラムを掲載している．ここでは，それらの実行方法を述べる．Scilab のコマンドや文法，プログラミングなどについては [41, 32, 33] など

[*1] Mathematica は数値計算もできるが数学の厳密な計算もこなす非常に優れたソフトウェアである．

を参照のこと．また次の WEB ページも参考になる．

- Scilab 情報ブログ
  `http://scilabinfo.wordpress.com/`
- Scilab 日本語ページ
  `http://www.geocities.jp/rui_hirokawa/scilab/`
- 精度保証付き数値計算入門
  `http://www.oishi.info.waseda.ac.jp/~oishi/sir/note.html`

1.2.1　Scilab プログラムの実行

本書に掲載しているプログラム（例えば，p.35 に掲載のプログラム）を実行するための方法をここでは示す．まず適当なエディタを立ち上げ，プログラムを記述する．Scilab には専用のエディタ Scipad が付属しているので，これを使用してもよい．Scipad は Scilab コンソールの「アプリケーション」メニューから「エディタ」を選べば起動する[*2]．または，コマンドウィンドウで scipad と打ち込んでも起動する．

エディタでプログラムを記述できたら，ファイル名を foo.sce として保存する．foo の部分は任意に決めてもかまわないが，拡張子は必ず .sce とする．

保存されたプログラムファイル foo.sce を実行するには，コマンドウィンドウの「ファイル」メニューから「実行」を選択し[*3]，実行したいファイル foo.sce を選択する．または，コマンドウインドウで，

```
--> exec('foo.sce');
```

と実行しても良い．なお，ファイル foo.sce が保存されているディレクトリ（例えば，これを C:/scilab/work/ とする）がカレントディレクトリ（現在，作業を行っているディレクトリ）でない場合，次のようにカレントディレクトリを変更してから，プログラムファイルを実行する．

```
--> chdir('c:\scilab\work\');
--> exec('foo.sce');
```

なお，Scipad でプログラムを記述した場合は，「実行」メニューの「Scilab へロード」を選べば，プログラムを実行できる．

[*2] バージョン 5.2 以前では，Editor メニューを選べば Scipad が起動する．
[*3] バージョン 5.2 以前では，File メニューから Exec ... を選択する．

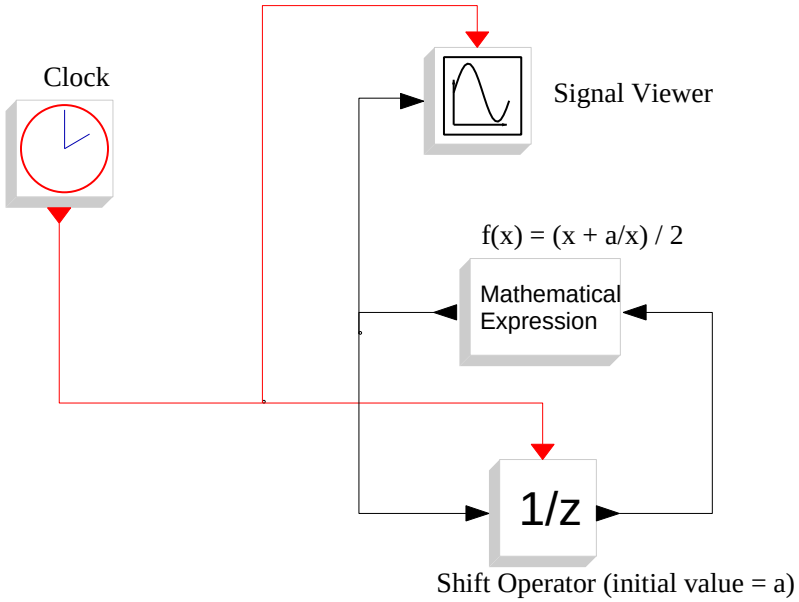


図 1.1　Xcos によるシミュレーション

1.2.2　Xcos によるシミュレーション

図 1.1 のようなブロック線図によるシミュレーションは Scilab に付属の Xcos （バージョン 5.2 以前は Scicos）というソフトウェアにより簡単に行うことができる．以下に図 1.1 に示された Xcos によるシミュレーションの実行手順を示す[*4]．なお，このシミュレーションの意味は 3 章を参照のこと．

1. Scilab を起動する．
2. メニュー「アプリケーション」より「Xcos」を選択する．「無題」という名の新しい Window とパレットブラウザーが立ち上がる．
3. パレットブラウザーの左メニューより「イベント・ハンドリング」をクリックし，選択する．
4. この中から，時計の絵のブロック `CLOCK_c`（これをアクティベーションクロックと呼ぶ）をクリックし，そのままドラッグ＆ドロップで，2. で立ち上がった「無題」

[*4] 実行手順は Scilab バージョン 5.2 以降の使用を想定している．

の Window に載せる.

5. パレットブラウザーの左メニューより「出力／表示」をクリックし，選択する．Sinks を選ぶ.
6. この中から，図 1.1 の Signal Viewer と同じブロック (CSCOPE) を選び，4. と同じようにドラッグ&ドロップで配置する.
7. パレットブラウザーの左メニューより「ユーザ定義関数」をクリックし，選択する.
8. Mathematical Expression と書かれたボックス EXPRESSION をドラッグ&ドロップで配置する.
9. パレットブラウザーの左メニューより「離散時間システム」をクリックし，選択する.
10. 1/z と書かれたボックス DOLLAR_f をドラッグ&ドロップで配置する.
11. 以上でブロックの配置は終わりである．次に各ブロックのパラメータを設定する.
    - クロックのブロックをダブルクリックし，Period を 1 に，Init time を 0 にする.
    - Signal Viewer はそのままで良い.
    - Mathematical Expression ブロックをダブルクリックし，number of inputs を 2 に，scilab expression を (u1+2/u1)/2 にする．use zero-crossing を 1 にする.
    - 1/z をダブルクリックし，initial condition を 10 にする． Inherit は 0 にする.
12. 図 1.1 を見ながら，ブロック間を配線する．ブロックとブロックを配線するためには，発信元のブロックのポート（三角形）をクリックし，そのままドラッグしてリンク先のブロックのポートの上で離せばよい．なお，見栄えよく配線するためには，配線するときに時々クリックボタンを押し，リンク線を曲げると良い．また，配線からリンク線を分岐させるためには,
    - 分岐させたい配線をクリックし,
    - そのままドラッグし,
    - リンク先のブロックの目的のポートの上で離す（配線がつながる）.

    なお，ブロックの入出力の位置を左右逆にするためには，そのブロックをクリックして選択し，キーボードのコントロールキー (ctrl) を押しながら r を押せばよい.
13. 次に，シミュレーション時間を決める．メニュー「シミュレーション」から「セットアップ」を選び,「最終統合時間」を 20 にする.
14. 最後に，再生ボタン (メニューの下の ▷ のボタン) を押し，シミュレーションを実行する.

第 2 章

数値計算と誤差

数値計算は多くの場合，厳密な計算の近似計算である．したがって，数値計算と誤差との間には切っても切れない関係がある．ここでは，数値計算とそれに起因する誤差について解説する．

2.1 厳密解と近似

積分で定義された次の関数を考える．

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{2.1}$$

これは，Gauss **誤差関数** (Gauss error function) と呼ばれ，確率統計学で特に重要な関数である．任意に与えられた正の実数 x に対するこの関数の値の計算，すなわち Gauss **関数** (Gaussian function) e^{-t^2} の 0 から x までの積分は，初等関数では表すことができないことが知られている．いっぽう，Gauss 誤差関数 (2.1) に関しては，Taylor 展開を用いた次の級数展開が知られている [44, 第 III 巻].

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!(2n+1)} \tag{2.2}$$

任意の $x \in \mathbb{R}$ に対して，この展開式は (2.1) の右辺の積分と厳密に等しい．すなわち (2.2) は (2.1) の右辺の積分に対する厳密解であると言える．一方，x として具体的に数値を与えて，$\mathrm{erf}(x)$ の値を計算したい場合，(2.2) では無限回の足し算を実行しなければならず，コンピュータでその厳密な値を計算することは有限の時間では不可能である．したがって (2.2) のように無限回足し合わせるのではなく，自然数 N を与えて，

$$\widetilde{\mathrm{erf}}(x) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{N} \frac{(-1)^n x^{2n+1}}{n!(2n+1)}$$

表 2.1　厳密解と近似

厳密解	近似
$\sqrt{2}$	1.414
π	3
$\dfrac{1}{3}$	0.3
$\displaystyle\sum_{n=0}^{\infty}\frac{1}{n}$	$\displaystyle\sum_{n=0}^{N}\frac{1}{n}$
$\displaystyle\lim_{x\to 0}\frac{\sin x}{x}$	$\dfrac{\sin\epsilon}{\epsilon},\ 0<\epsilon\ll 1$
$f'(x)$	$\dfrac{f(x+h)-f(x-h)}{2h},\ 0<h\ll 1$
$\displaystyle\int_0^1 f(x)dx$	$\displaystyle\sum_{n=0}^{N} f(x_i)(\Delta x)_i$

を計算することが考えられる．N が有限なら一般に $\mathrm{erf}(x)\neq\widetilde{\mathrm{erf}}(x)$ であるが，N が十分大きいなら，$\widetilde{\mathrm{erf}}(x)$ の値は $\mathrm{erf}(x)$ に近いと考えられる．この $\widetilde{\mathrm{erf}}(x)$ は $\mathrm{erf}(x)$ の**近似値** (approximation) である．

この Gauss 誤差関数の例に限らず，多くの工学的な問題ではその厳密解が簡単な数値で与えられる場合は少ない．例えば，表 2.1 に示したような近似が必要な場合が多い．このように近似値を得る方法を探る学問を**数値計算** (numerical computation)，または**数値解析** (numerical analysis) と呼ぶ．数値計算では，ただ単に近似値を得る方法を提供するだけでなく，近似値を

- 精度よく
- 速く

求める方法を探究する．このうち，精度に関しては，次節以降で述べる誤差の解析が非常に重要である．

2.2　誤差の原因

誤差という言葉は，実験や測定においてよく使われる．実験データや測定データにおいて，誤差は

誤差 = 真値 − 測定値

で定義され，大きくわけて次の 2 つに分類される [19].

系統誤差: 測定機器の不備や測定者の過失，未熟さなどに伴う人為的な誤差.
偶然誤差: いかに熟練者でも制御しえない偶然的に発生する誤差.

系統誤差 (systematic error) は，測定機器をより高精度なものにしたり，測定者を訓練することにより抑えることができる．いっぽう**偶然誤差** (random error) は，測定回数を増やすことにより改善される．すなわち，何度か測定を繰り返し，その測定値の平均をとることで偶然誤差を減らすことが可能となる．これは，確率論における**中心極限定理** (central limit theorem) [51, 8] が理論的な根拠となっている．

では，数値計算における誤差はどうだろうか．通常，数値計算で問題となるのは系統誤差である．例えば，次のような誤差（ミス）が生じる．

例 1 プログラムのバグによる誤差
例 2 アルゴリズムの間違いや不安定なアルゴリズムを使用したことによる誤差
例 3 反復法の初期値設定のミスによる誤差
例 4 数値を有限桁に丸めたり，操作を有限回で打ち切ったりしたときに生じる誤差

例 1 は少し長いプログラムであればほぼ必ず発生する．これを避ける方法は，プログラムをきれいにわかりやすく書くことである[*1]．一方，例 2 は典型的には，数値計算用ソフトウェアに付属のサブルーチンパッケージを盲目的に信じて（すなわちブラックボックスとして）使用することによって生じる．例えば，実数の解のみを返すアルゴリズムを用いてある方程式の解を求め，「解なし」という結果が得られたとする．この計算結果から「この方程式は実数解を持たない」と結論づけることは間違いである可能性がある．なぜなら，そのアルゴリズムはある有限区間の上でしか解を探索しないのかもしれないし，解に到達するまえに繰り返し回数上限に達したかもしれないからである．すなわち，アルゴリズムの中身を知らなければ，上のような誤った結論をしてしまう可能性がある．数値計算の理論を勉強する大きな目的のひとつは，数値計算アルゴリズムのしくみを知り，このような過ちを極力なくすことである．例 3 は初期値を真の解に近い位置に置くことで解決するが，数学的には非常に難しい問題である．この問題に関しては，第 5 章にて詳しく述べる．また，例 4 はコンピュータが有限の桁数で有限回の操作しか実行できないことが原因であり，**数値誤差** (numerical error) と呼ばれる．本章では，この数値誤差のうち丸め誤差の原因であるコンピュータの数値表現について述べた後，数値誤差の現象について詳しく調べる．

[*1] きれいでわかりやすいプログラムの書き方を勉強するには，カーニハン (B. Kernighan) とパイク (R. Pike) の著書 [17] が非常に参考になる.

練習問題 1 数値計算に偶然誤差は存在するか？

2.3 コンピュータにおける数値の表現

2.3.1 IEEE 754

コンピュータでは有限桁の 2 進数しか扱うことができない．例えば $\sqrt{2}$ や円周率 π などを数値としてコンピュータ上で扱うためには，それらを有限桁の 2 進数で表現する必要がある．数値の 2 進数表現にはさまざまな形式があるが，現在のほとんどすべてのコンピュータにおいては，IEEE 754 と呼ばれる IEEE **標準規格** (IEEE Standard) にもとづく表現法が採用されている．ここで IEEE （米国電気電子学会）とは，電気電子工学や制御工学，情報工学，通信工学などに関する世界最大の学会であり[*2]，会誌の発行や学術会議の開催のほか，それらの分野における機器や方式等の標準化の活動を行っている．IEEE 802 (LAN に関する規格）や IEEE 1394 （シリアルバスの規格）は普段からよく目にする．

2.3.2 IEEE 754 の浮動小数点数表現

IEEE 754 で採用されている数値の表現形式は**浮動小数点数** (floating-point number) である．これは，次の形式で表される有限桁の数である．

$$(-1)^c \times S \times 2^e = (-1)^c \times 1.d_1 d_2 \cdots d_{p-1} \times 2^e \tag{2.3}$$

それぞれの項の定義は以下のとおりである．

符号ビット: c を**符号ビット** (sign bit) と呼び，0 または 1 の値をとる．$c = 0$ なら正の数を，$c = 1$ なら負の数を表す．

仮数: $S = 1.d_1 d_2 \cdots d_{p-1}$ を**仮数** (significand) と呼び，各 d_i は 0 または 1 の値をとる．p は仮数のビット長であり，**単精度** (single precision) の場合は $p = 24$ ビット，**倍精度** (double precision) の場合は $p = 53$ ビットと決められている[*3]．仮数 S は必ず $1_2 \leq S < 10_2$ を満たす[*4]．これを満たす浮動小数点数を**正規化数** (normalized number) と呼ぶ．

[*2] ちなみに筆者も IEEE の会員である．

[*3] ただし，最初の 1 は固定であるので，計算機の内部表現では 23 ビット（単精度）または 52 ビット（倍精度）である．表 2.2 を見よ．

[*4] 添え字の 2 は 2 進数を表す．すなわち，$10_2 = 2$ である．

表 2.2 IEEE 754 の浮動小数点数表現の精度．数字の単位はビット．

	仮数 S	指数 e	符号 c	合計	C 言語
単精度	23	8	1	32	`float`
倍精度	52	11	1	64	`double`

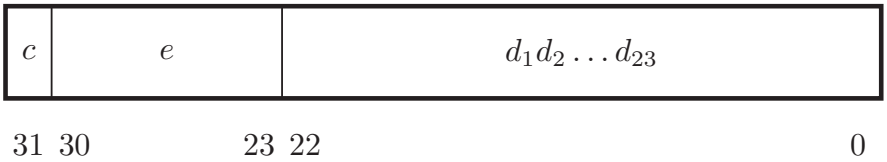


図 2.1 IEEE754 単精度浮動小数点数の 2 進数表現

指数: e は**指数** (exponent) と呼ばれ，単精度では，$e_{\min} = -126$ から $e_{\max} = 127$ （8 ビット相当）の範囲，倍精度では $e_{\min} = -1022$ から $e_{\max} = 1023$ （11 ビット相当）の範囲の整数値である．

表 2.2 に IEEE 754 の浮動小数点数表現をまとめる．単精度で表現できる絶対値が最小の数 m は $\pm 1.0\ldots0_2 \times 2^{e_{\min}}$ であり，この値は 10 進数で

$$m = 2^{e_{\min}} = 2^{-126} \approx 1.2 \times 10^{-38}$$

となる．また絶対値が最大の数 M は $\pm 1.1\ldots1_2 \times 2^{e_{\max}}$ であり，10 進数であらわすと

$$M = 2^{e_{\max}}(2 - 2^{-23}) \approx 2^{e_{\max}+1} = 2^{128} \approx 3.4 \times 10^{38}$$

となる．

練習問題 2 倍精度のときの最小数 m と最大数 M を求めよ．

2.3.3 IEEE 754 の 2 進数表現

ここでは，上で述べた IEEE 754 形式における浮動小数点数の 2 進数表現について述べる．単精度の場合のみについて述べるが，倍精度の場合も考え方は同じである．IEEE 754 では，符号 c，指数 e，仮数（の一部） $d_1d_2\ldots d_{p-1}$ の順序で図 2.1 のように 0 または 1 を並べる．ただし，指数 e の部分は実際の値に 127 を足した**バイアス表現** (biased representation) を用いる．すなわち指数 e の 2 進数表現は

$$00000001_2 \sim 11111110_2$$

の範囲の 2 進数となる．このバイアス表現で指数が $0000\ 0000_2$ または $1111\ 1111_2$ の場合，すなわち指数 e が $e_{\min} - 1 = -127$ および $e_{\max} + 1 = 128$ の場合は特別な意味を持つ（後述する）．

例題 1 10 進数 8.75 を IEEE 754 形式の 2 進数に変換してみよう．まず，

$$8.75 = 8 + 0.75 = 2^3 + \frac{1}{2} + \frac{1}{2^2}$$

であるので，

$$8.75 = 1000.11_2 = 1.00011_2 \times 2^3$$

と表現できる．符号はプラスであるので $c = 0$．指数は 3 であるが，バイアス表現にするために 127 を足して

$$130 = 128 + 2 = 2^7 + 2^1 = 1000\ 0000_2 + 10_2 = 1000\ 0010_2$$

となる．仮数は $S = 1.00011$ であるので，小数点以下を 23 ビットで表して，

$$d_1 d_2 \ldots d_{23} = 0001\ 1000\ 0000\ 0000\ 0000\ 000_2$$

である．以上を図 2.1 にしたがって並べると

$$0 \mid 1000\ 0010 \mid 0001\ 1000\ 0000\ 0000\ 0000\ 000_2$$

が得られる（縦棒は符号部・指数部・仮数部の区切りを表す）．

練習問題 3 10 進数 -15.125 を IEEE754 の形式（図 2.1）にしたがって単精度の 2 進浮動小数点数に変換せよ．

2.3.4　IEEE 754 の例外処理

IEEE 754 では，ある数値を 0 で割ったり，演算結果が最大数 M を超えたりした場合の**例外** (exception) を 5 つ定義している．それらは以下の通りである．

1. オーバーフロー
2. アンダーフロー
3. ゼロ割り
4. 不正
5. 不正確

オーバーフロー (overflow) とは，浮動小数点数演算結果の絶対値が最大数 M を上回る現象である．オーバーフローが生じたときは，指数が $e = e_{\max} + 1 = 128$ （バイアス表現では $1111\ 1111_2$）で仮数が $1.00\ldots0_2$ の浮動小数点数（これは $\pm\infty$ を表す）が数値の代わりに用いられる．計算途中で数値がオーバーフローを起こすと予想される場合は，適切に数値のスケールを変換する必要がある．例えば，次の計算を考える．

$$f(x, y) = \frac{x}{\sqrt{x^2 + y^2}}.$$

最初に $x^2 + y^2$ を計算した後ルートをとり，次にそれで x を割るという手順で計算を行うとする．もし数値 x と y が非常に大きく，計算途中で $x^2 + y^2$ がオーバーフローを起こす可能性がある場合は，十分小さな**スケール因子** (scale factor) $s > 0$ を用いて，

$$f(x, y) = \frac{sx}{\sqrt{(sx)^2 + (sy)^2}}$$

のように計算する工夫が必要である．

アンダーフロー (underflow) とは，演算結果の絶対値が最小数 m を下回る現象を指す．アンダーフローが生じた場合，**非正規化数** (denormalized number, subnormalized number) と呼ばれる m よりも小さい特殊な数が代わりに用いられるか，または符号にしたがって ±0 または $\pm m$ が用いられる．非正規化数とは，指数が $e = e_{\min} = -126$ で仮数が $0.d_1d_2\ldots d_{p-1}$ (d_i の少なくともひとつは 0 でない) の浮動小数点数であり，IEEE 754 の 2 進数表現では，仮数を $1.d_1d_2\ldots d_{p-1}$，指数を $0000\ 0000_2$ として表現する．この数はアンダーフローが生じた場合にのみ用いられ，通常の計算では用いられない．また ±0 は，指数を $0000\ 0000_2$，仮数を $1.00\ldots0_2$ として表現する．

ゼロ割り (divide by zero) とは文字通り数値を 0 で割ることであり，この演算結果は符号に応じて $\pm\infty$ となる．

不正 (invalid) とは，$\sqrt{-1}$ や $0 \times \infty$，$0/0$，∞/∞，$\infty - \infty$ などの演算で生じる．不正が生じた場合は，NaN (Not a Number) を返す．IEEE 754 では NaN を指数が $e = e_{\max} + 1 = 128$ （バイアス表現で $1111\ 1111_2$）で仮数が $1.d_1d_2\ldots d_{p-1}$ (d_i の少なくともひとつは 0 でない) の浮動小数点数で表現する．

不正確 (inexact) とは厳密な演算結果と異なる結果が生じた場合のことを言い，表 2.2 で示される桁数で表現しきれないとき（オーバーフローとアンダーフローを含む）のことである．特に次節で述べる丸め誤差が生じる演算はこれに相当する．

以上で出てきた特殊な数を表 2.3 にまとめる．

2.4 数値誤差

数値誤差には大きく分けて次の 2 つがある．

表 2.3　IEEE 754 における特殊な数．$f = d_1 d_2 \ldots d_{p-1}$ とする．± の符号は符号ビット c により定義する．

指数	指数のバイアス表現	仮数と符号	意味
$e_{\min} - 1$	$0000\ 0000_2$	$\pm 1.f,\ f = 0_2$	± 0
$e_{\min} - 1$	$0000\ 0000_2$	$\pm 1.f,\ f \neq 0_2$	$\pm 0.f \times 2^{e_{\min}}$
$e = e_{\max} + 1$	$1111\ 1111_2$	$\pm 1.f,\ f = 0_2$	$\pm\infty$
$e = e_{\max} + 1$	$1111\ 1111_2$	$\pm 1.f,\ f \neq 0_2$	NaN

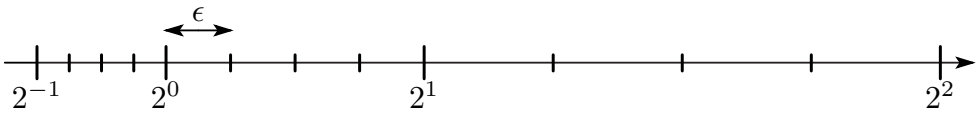


図 2.2　数値の丸め ($p = 3$ の場合)．ϵ はマシンイプシロン．

1. 実数値を有限桁の数で近似する際に発生する丸め誤差
2. 極限操作や無限個の数の和などを有限で近似する際に発生する打切り誤差

1 は数値の近似に関する誤差，2 は操作の近似に関する誤差である．

2.4.1　浮動小数点数と丸め誤差

前節で述べたようにコンピュータの内部では有限の桁数しか扱うことができない．したがって，桁の多い数や無理数を扱うためには，それらをある桁数の数で近似する必要がある．これを**丸め** (rounding) と呼び，丸めによって発生する誤差を**丸め誤差** (round-off error) と呼ぶ．

実数を丸めるという操作は，数直線を離散化することに他ならない．すなわち，図 2.2 のように数直線を離散化すると，これら離散点上の値のみをコンピュータは扱うことができ，離散点上にない実数は最も近い離散点の数で近似される．これが丸めである．また，図 2.2 は浮動小数点数表現の特徴のひとつを示している．すなわち，離散点は等間隔ではなく，2^i と 2^{i+1} の間で間隔が $2^i \times 2^{-p+1} = 2^{i-p+1}$ と変化する．

この丸めの操作は，信号処理で**量子化** (quantization) [25] と呼ばれる操作に相当する．量子化の言葉を用いると，図 2.2 で表される離散化は**非一様量子化** (non-uniform quantization) と呼ばれる操作である．一方，**固定小数点数** (fixed-point number) の表現では，図 2.2 の離散点が等間隔に並び，**一様量子化** (uniform quantization) に対応する．固定小数点数の表現は，後で述べるように丸めの相対誤差が大きくなる欠点があるが，演算処理を高速に動作させることが可能であり，また LSI 上に実装したときの消費電力も

少なくて済む．したがって，携帯電話やシリコンオーディオなどの小型機器で用いられる DSP (Digital Signal Processor，信号処理用の高速演算装置) ではしばしば固定小数点数が用いられる [48].

図 2.2 において，$2^0 = 1$ とその右隣の点 $2^0 + 2^{-2} = 1.25 = 1.01_2$ との距離 ϵ を**マシンイプシロン** (machine epsilon) と呼ぶ．一般にマシンイプシロンは，

$$\epsilon = 2^{-p+1}$$

とあらわされる．これより，IEEE 754 の浮動小数点数表現でのマシンイプシロンは，単精度 ($p = 24$) の場合 $\epsilon = 2^{-24+1} \approx 1.192 \times 10^{-7}$，倍精度 ($p = 53$) の場合 $\epsilon = 2^{-53+1} \approx 2.220 \times 10^{-16}$ となることがわかる．

このマシンイプシロンを用いれば，丸め誤差を見積もることができる．実数 x を考える．ただし，$m \leq |x| \leq M$ とする．ここで，m と M はそれぞれ表現できる最小数および最大数である．この x に対して，浮動小数点数に丸めた数を $\widetilde{x}$ とおくと，

$$\widetilde{x} = x(1 + \delta), \quad |\delta| \leq \frac{\epsilon}{2}$$

と表すことができる．これを用いれば丸め誤差は

$$|x - \widetilde{x}| = |\delta x| \leq \frac{\epsilon}{2} \cdot |x| \tag{2.4}$$

と評価できる．この式において $\epsilon/2$ を**丸めの単位** (unit round-off) と呼ぶ．上の式の両辺を $|x|$ で割ると ($|x| \geq m > 0$ であることに注意) 次の評価式が得られる．

$$\left| \frac{x - \widetilde{x}}{x} \right| \leq \frac{\epsilon}{2}.$$

この式の左辺を**相対誤差** (relative error) と呼ぶ．すなわち，浮動小数点数への丸めの相対誤差は丸めの単位以下であることがわかる．一方，固定小数点数では，$|x|$ が小さくなるほど相対誤差は一般に大きくなる．

自分の使用しているコンピュータのマシンイプシロンを調べるには，次のアルゴリズムを実行すればよい [9].

```
eps = 1;
do eps = 0.5*eps; while (eps + 1 > 1)
```

2.4.2 桁落ち

実数 x と y ($x > y$ とする) に対して，次の引き算を計算することを考える．

$$z = x - y.$$

これをコンピュータ上で行う場合，前節で述べたように x と y を丸める必要がある．実数 x, y を丸めた浮動小数点数をそれぞれ $\widetilde{x}, \widetilde{y}$ とおく．また，丸めた数 $\widetilde{x}, \widetilde{y}$ による演算結果を $\widetilde{z}$ とおく．すなわち，

$$\widetilde{z} = \widetilde{x} - \widetilde{y}.$$

この結果の相対誤差を計算してみよう．相対誤差を e_r とおくと，

$$e_r = \left| \frac{z - \widetilde{z}}{z} \right| = \frac{|x - y - (\widetilde{x} - \widetilde{y})|}{x - y} \leq \frac{|x - \widetilde{x}| + |y - \widetilde{y}|}{x - y} \leq \frac{\epsilon}{2} \cdot \frac{|x| + |y|}{x - y}$$

が得られる．ここで，最後の不等式に (2.4) を用いた．これより，$|x|$ と $|y|$ がマシンイプシロン ϵ に比べて十分大きく（すなわち，$|x| + |y| \gg \epsilon$ のとき），また x と y が非常に近い場合，すなわち $x - y \approx \epsilon$ となったときに，相対誤差 e_r が ϵ に比べて非常に大きくなる可能性を示している．実際，ほとんどの場合，非常に近い数どうしを引き算すると，有効桁数が損失してしまい，損失した桁の部分（ビット）には自動的に 0 が挿入されるので，相対誤差がマシンイプシロンに比べ非常に大きくなる．この現象を**桁落ち** (loss of significant digits) という．桁落ちを防ぐためには，大きな数どうしの引き算を極力避けることである．例えば，3 辺の長さが a, b, c の三角形の面積 A は Heron **の公式** (Heron's formula) により

$$A = \sqrt{s(s-a)(s-b)(s-c)}, \quad s = \frac{(a+b+c)}{2}$$

で与えられるが，a が $b + c$ とほぼ等しい非常に扁平な三角形の場合は，$s \approx a$ となり $s - a$ の部分で桁落ちが起こる可能性がある．このような場合は，上式のかわりに

$$A = \frac{\sqrt{(a + (b+c))(c - (a-b))(c + (a-b))(a + (b-c))}}{4}, \quad a \geq b \geq c$$

の計算公式を使用するとよいことが知られている [9]．桁落ちを防ぐにはこのような巧妙なテクニックが必要である．

練習問題 4 次の関数を考える．

$$f(x) = 1 - \cos x.$$

この関数の値をコンピュータで計算したい．しかし，x が非常に小さい場合，桁落ちが起こる可能性がある．小さな x に対して桁落ちが起きないような関数 $f(x)$ の計算公式を求めよ．

2.4.3 情報落ち

Napier **の数** (Napier's constant) e を次の公式を用いて求めることを考える．

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

コンピュータで計算させるために，まず十分大きな N を用いて，上式を近似する．

$$e \approx \sum_{n=0}^{N} \frac{1}{n!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \cdots + \frac{1}{N!}$$

このとき，N を大きくすれば上の近似値は真値 e に近づいていくだろうか？実はコンピュータを使って $n = 1$ から順番に足していくと，ある n 以降は誤差が減少しない．この原因をここでは考える．

一般に非常に大きな数と非常に小さな数を足し合わせると，その結果はほとんど変化しない．例えば，有効数字が 3 桁の計算機で 1.00×10^5 と 1.00×10^{-5} を足し合わせると，

$$1.00 \times 10^5 + 1.00 \times 10^{-5} = 100000 + 0.00001 = 100000.00001 = 1.0000000001 \times 10^5$$

となり，計算結果が 3 桁に丸められて 1.00×10^5 となる．すなわち，足し算をしても結果は変化しない．これを**情報落ち** (loss of information) と呼ぶ．

情報落ちを回避するためには，上記の級数を小さいものから足していく．すなわち，次のように計算する．

$$\begin{aligned} S_1 &= \frac{1}{N!} \\ S_2 &= S_1 + \frac{1}{(N-1)!} \\ &\vdots \\ S_N &= S_{N-1} + 1 \end{aligned}$$

最終的に計算される S_N は情報落ちがなく e のよい近似となっている．さらに精度をよくするには，同程度の大きさの項を二つずつペアを組んで加算する方法や Kahan **の総和公式** (Kahan summation formula) と呼ばれる方法などを用いればよい [9, 39]．

2.4.4 浮動小数点数の演算における丸め誤差

半径 $1/3$ の円の面積 S を求める問題を考える．この問題の厳密解は円周率 π を用いて，

$$S = \pi \times \left(\frac{1}{3}\right)^2 = \frac{\pi}{9} = 0.34906585\ldots$$

と求められる．ここで，有効数字が 1 桁の計算機を使用すると仮定し，円周率を 3，半径を 0.3 に丸めて面積を計算すると，面積の近似値 $\widetilde{S}$ は

$$\widetilde{S} = 3 \times (0.3)^2 = 0.27$$

となり，さらにこれを丸めるので，結果は 0.3 となる．この計算からわかるように，丸め誤差が存在するときに演算を繰り返すと，何度も丸め誤差が入る．ここでは，丸め誤差が四則演算などの演算結果にどう影響するかを考察する．

実数 x と y に対する演算 $\phi(x, y)$ を考える．例えば，足し算であれば，

$$\phi(x, y) = x + y$$

割り算であれば，

$$\phi(x, y) = x/y$$

である．もっと複雑な演算，例えば

$$\phi(x, y) = (x^2 + y^2)/(2xy)$$

なども可能である．とにかく，この演算をコンピュータを用いて行うことを考える．コンピュータ内部では数値は有限桁の数で表されるので，必ず丸め誤差が生じる．まず実数である x と y を浮動小数点数 $\widetilde{x}$ と $\widetilde{y}$ に丸める必要がある．ここで，丸めの操作を Q とおく[*5]．すなわち，

$$\widetilde{x} = Qx, \quad \widetilde{y} = Qy$$

とする．このように丸めた上で，$\widetilde{x}$，$\widetilde{y}$ に対して演算 ϕ を行う（演算結果を z とおく）．

$$z = \phi(\widetilde{x}, \widetilde{y}) = \phi(Qx, Qy).$$

最後に演算結果 z を丸める必要がある．丸められた z を $\widetilde{z}$ とおく．すなわち，

$$\widetilde{z} = Qz = Q\phi(\widetilde{x}, \widetilde{y}) = Q\phi(Qx, Qy).$$

このとき，演算後の丸め誤差は次のように計算できる．

$$\begin{aligned} z - \widetilde{z} &= \phi(x, y) - Q\phi(\widetilde{x}, \widetilde{y}) \\ &= \{\phi(x, y) - \phi(\widetilde{x}, \widetilde{y})\} + \{\phi(\widetilde{x}, \widetilde{y}) - Q\phi(\widetilde{x}, \widetilde{y})\} \\ &= \{\phi(x, y) - \phi(\widetilde{x}, \widetilde{y})\} + \left\{\phi(\widetilde{x}, \widetilde{y}) - \widetilde{\phi}(\widetilde{x}, \widetilde{y})\right\}, \end{aligned}$$

ただし，$\widetilde{\phi} := Q\phi$ とおいた．これより，

$$|z - \widetilde{z}| \leq |\phi(x, y) - \phi(\widetilde{x}, \widetilde{y})| + \left|\phi(\widetilde{x}, \widetilde{y}) - \widetilde{\phi}(\widetilde{x}, \widetilde{y})\right|$$

[*5] この Q には量子化 (Quantization) の意味を込めている．

が成り立ち，演算 ϕ による丸め誤差は，x と y の丸めによる誤差（第 1 項）と演算結果の丸めによる誤差（第 2 項）との和で上からおさえられることがわかる．第 2 項の誤差は，浮動小数点数どうしの演算（たとえば掛け算や割り算など）でも生じる可能性のある誤差であり，**発生誤差** (generated error) と呼ばれる．

2.4.5 打切り誤差

打切り誤差 (truncating error) とは，コンピュータで計算するために，微分や積分などの極限操作を離散化したり，無限個の数の和を有限回で打ち切ったりしたときに生じる誤差である．

例えば関数の微分を考えてみよう．微分可能な関数 $f(x)$ の微分 $f'(x)$ は次式で定義される．

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$$

この微分をコンピュータで計算するためには，極限の操作を離散化しなければならない．そこで十分小さな $h > 0$ を用いて，$f'(x)$ の近似値 $\delta_h f(x)$ を

$$\delta_h f(x) := \frac{f(x+h) - f(x-h)}{2h} \tag{2.5}$$

で定義する．ここで，h を**ステップ幅** (step size) と呼び，上記のような近似を**中心差分近似** (central-difference approximation) と呼ぶ．中心差分近似による打切り誤差を ε_0 とおくと，

$$\varepsilon_0 = \delta_h f(x) - f'(x) \tag{2.6}$$

が得られる．関数 f を C^3 級とすると，Taylor **の定理** (Taylor's theorem) [23] から[*6]，$\theta_+ \in (x, x+h)$, $\theta_- \in (x-h, x)$ が存在して，

$$f(x \pm h) = f(x) \pm h f'(x) + \frac{h^2}{2} f''(x) \pm \frac{h^3}{6} f'''(\theta_\pm)$$

と書ける．この式を (2.5) に代入し，(2.6) を用いることにより，打切り誤差は

$$\varepsilon_0 = \frac{h^2}{12} \left(f'''(\theta_+) + f'''(\theta_-) \right)$$

と計算される．すなわち，ステップ幅 h を小さくすれば，打切り誤差はステップ幅 h の 2 乗のオーダーで減少することがわかる．

[*6] Taylor の定理については付録を参照のこと．

しかしこの性質は，丸め誤差を考慮に入れると成り立たない．簡単のため，x と h は有限桁の浮動小数点数とする．まず $f(x+h)$ と $f(x-h)$ の演算に丸めが入る．それらを $\widetilde{f}(x+h)$ および $\widetilde{f}(x-h)$ とおく．すなわち，

$$\widetilde{f}(x+h) = Qf(x+h), \quad \widetilde{f}(x-h) = Qf(x-h).$$

ここで Q は丸めの操作を表す．これらの丸め誤差をそれぞれ ε_1 および ε_2 とおく．すなわち，

$$\varepsilon_1 = \widetilde{f}(x+h) - f(x+h), \quad \varepsilon_2 = \widetilde{f}(x-h) - f(x-h).$$

これらの誤差を考慮すると，中心差分近似 $\delta_h f(x)$ は

$$\begin{aligned}\delta_h \widetilde{f}(x) &:= \frac{\widetilde{f}(x+h) - \widetilde{f}(x-h)}{2h} \\ &= \frac{(f(x+h)+\varepsilon_1) - (f(x-h)+\varepsilon_2)}{2h} = \delta_h f(x) + \frac{\varepsilon_1 - \varepsilon_2}{2h}\end{aligned}$$

と計算される．さらに，この $\delta_h \widetilde{f}(x)$ が丸められ $Q\delta_h \widetilde{f}(x) = \widetilde{\delta}_h \widetilde{f}(x)$ となり，このときの丸め誤差を ε とおくと，

$$\widetilde{\delta}_h \widetilde{f}(x) = \delta_h \widetilde{f}(x) + \varepsilon = \delta_h f(x) + \frac{\varepsilon_1 - \varepsilon_2}{2h} + \varepsilon = f'(x) + \varepsilon_0 + \frac{\varepsilon_1 - \varepsilon_2}{2h} + \varepsilon$$

が得られる．したがって打切り誤差と丸め誤差を含めた全体の誤差の絶対値 E は

$$E = \left| \widetilde{\delta}_h \widetilde{f}(x) - f'(x) \right| = \left| \varepsilon_0 + \frac{\varepsilon_1 - \varepsilon_2}{2h} + \varepsilon \right| \tag{2.7}$$

となる．

上で述べたように h が十分小さいなら，打切り誤差 ε_0 は十分小さい．また，ε は有界である．しかし，$\frac{\varepsilon_1 - \varepsilon_2}{2h}$ の項は，$\varepsilon_1 - \varepsilon_2$ が h の減少にともなって小さくならない限り，$h \to 0$ で発散する．したがって，h は大きすぎても小さすぎてもだめで，ちょうど良い値を探す必要がある．ひとつの方法は，関数 f が具体的に与えられたとき，h の関数として E を求め，E を最小化する $h > 0$ を見つければよい（文献 [5] の第 7 章を参照せよ）．

なお，上のように連続のものを離散近似したときに生じる打切り誤差を**離散化誤差** (discretization error) とも言う．

練習問題 5 微分可能な関数 f の微分 f' に対して，

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}, \quad h > 0$$

を**前進差分近似** (forward-difference approximation) と呼び，

$$f'(x) \approx \frac{f(x) - f(x-h)}{h}, \quad h > 0$$

を**後退差分近似** (backward-difference approximation) と呼ぶ．前進差分近似および後退差分近似に対し，(2.7) のような誤差評価を求めよ．

練習問題 6 上記のような微分係数の計算以外に，打切り誤差が問題となるような数値計算の例をあげよ．

2.5　さらに勉強するために

2.2 節で述べた系統誤差や偶然誤差は [19] を参考にした．この本は，実験における測定誤差の統計的な解析のわかりやすい参考書である．2.3 節で取り上げた IEEE 754 に関しては，オリジナルの文献 [27] のほか，サーベイ論文 [9, 13] も参考になる．これらの文献は，

`http://grouper.ieee.org/groups/754/`

からダウンロードできる[*7]．また日本語では，[39] にわかりやすい解説がある．次の WEB ページでは，10 進数を入力すると IEEE 754 形式の 2 進浮動小数点数（単精度および倍精度）が出力され，便利である．

`http://babbage.cs.qc.cuny.edu/IEEE-754/Decimal.html`

2.4.1 節で述べた量子化については，[25] が参考になる．また量子化は，JPEG や MPEG などのデータ圧縮にも関連し，そこで用いられる手法は数値計算とも密接な関係がある．詳しくは，[29] を参照されたい．2.4.5 節の内容は [5] の第 7 章を参考にした．微分計算における最適なステップ幅 h の選び方についても同書を参照せよ．

[*7] 一部の文献は，IEEE のアカウントが無いとダウンロードできない．この機会に IEEE に入会しよう．学生の場合，IEEE の会費は驚くほど安い．

第 3 章

反復法のブロック線図表現

この章では，数値計算法，特に反復法の解析に便利なブロック線図について解説する．数値計算に限らず理論的な研究では複雑な数式を扱うことが多い．複雑な数式に対して何らかのイメージを与え，その数式の意味を把握することは非常に重要である．そのイメージ化に役立つ方法の一つにブロック線図がある．複雑な数式に出会ったとき，このブロック線図を用いて数式を理解してみるのは有用な方法である．

またブロック線図は信号処理や制御工学と関連が深い．したがって，ブロック線図を用いることにより，数値計算の問題を工学の問題として考えることが可能となる．そして，信号処理や制御工学で培われてきた膨大な理論や技術を援用することにより，数値計算の世界に新しいアイデアを持ち込むことができるという利点もある．

3.1 反復法

前章で述べたように，数値計算では，厳密解を得ることが（ほぼ）不可能な問題の近似解をより速くより精密に求めることが目標となる．厳密解を求めることが難しい問題には，大規模な連立方程式や非線形方程式，行列の固有値問題などさまざまな種類の問題があるが，それらの数値計算の多くは**反復法** (iterative method) と呼ばれる手法で記述される．反復法とは，上記のような簡単には解けない難しい問題を単純な問題の繰り返しで解く（近似解を求める）方法である．

反復法を数学的に定式化すると次のようになる．すなわち，ある集合 X 上の写像 $\phi : X \to X$ と初期値 $x_0 \in X$ を与えて[*1]，

$$x[n+1] = \phi(x[n]), \quad n = 0, 1, 2, \ldots, \quad x[0] = x_0 \tag{3.1}$$

[*1] 集合 X としては，たとえば実軸上の閉区間や $\mathbb{R}^N$ における単位球，$N \times N$ の対称行列全体などを考える．

により漸近的に厳密解に収束させる方法である．ここで，厳密解 x^* は集合 X 内に存在すると仮定する．反復法による数値計算法とは，(3.1) で定義される $x[n]$ が $n \to \infty$ で厳密解 x^* に収束するように関数 ϕ をうまく選ぶことに他ならない．また，その速さや精密さも関数 ϕ の選び方に依存する．

例題 2 ($\sqrt{a}$ を求める反復法) ここでは反復法による近似がどのようなものかを具体的な例題で確かめてみる．実数 $a > 0$ に対して，$\sqrt{a}$ の近似値を求める反復法は以下で与えられる．

$$x[n+1] = \phi(x[n]), \quad \phi(x) = \frac{1}{2}\left(x + \frac{a}{x}\right) \tag{3.2}$$

もしこの反復法がある値 $x^* \in [0, \infty)$ に収束したとする．このとき，

$$x^* = \phi(x^*) = \frac{1}{2}\left(x^* + \frac{a}{x^*}\right)$$

が成り立ち，これを変形し，$x^* > 0$ に注意すると $x^* = \sqrt{a}$ が得られる．すなわち，上の反復法は収束すればそれは $\sqrt{a}$ となる．

練習問題 7 初期値を $x_0 = a$ とし，a に自分の好きな数字（正の実数）を入れ，(3.2) の反復法により $\sqrt{a}$ の近似値を求めよ．

(3.2) のような反復法で問題となるのは，主に次の 3 つである．

- 反復法は厳密解に収束するのか？
- 収束するとして，その速さは？
- 反復法を有限回で打ち切ったときの誤差は？

これらを解析するためには，さまざまな数学的手法が必要となり，次章以降で詳しく述べる．その数学的解析の強力な道具となるのが，次で述べるブロック線図表現である．

3.2　関数とブロック線図

図 3.1 の電気回路を考えよう．この電気回路において，x を入力電圧，y を出力電圧とする．ここで入力電圧 x を周波数 $\omega \geq 0$，振幅 $A_x > 0$，位相 $\phi_x \in [0, 2\pi)$ の交流電圧とすると出力電圧 y は入力電圧と同じ周波数 ω の正弦波となり，その振幅 A_y と位相 ϕ_y は次式で与えられる．

$$A_y = A_x \cdot \left|\frac{1}{RCj\omega + 1}\right| = \frac{A_x}{\sqrt{(RC\omega)^2 + 1}}$$
$$\phi_y = \phi_x + \arg\left(\frac{1}{RCj\omega + 1}\right) = \phi_x - \arctan(RC\omega)$$

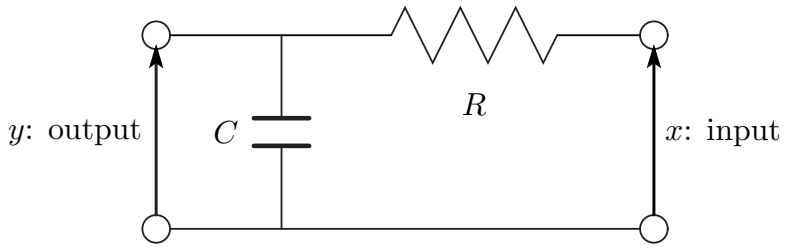


図 3.1 RC 回路

ただし，$j = \sqrt{-1}$ である[*2]．この振幅と位相は複素関数

$$g(s) = \frac{1}{RCs + 1} \tag{3.3}$$

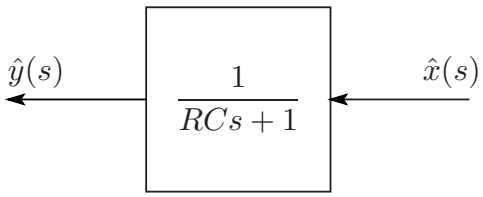
によって決定される．すなわち (3.3) の関数 $g(s)$ に $s = j\omega$ を代入し，その絶対値と偏角を計算することにより，出力電圧の振幅と位相が計算できる．(3.3) を図 3.1 の電気回路の**伝達関数** (transfer function) という．正弦波でなくても，入力 x と出力 y の Laplace **変換** (Laplace transform) [50]

$$\hat{x}(s) = \int_0^\infty x(t)e^{-st}dt, \quad \hat{y}(s) = \int_0^\infty y(t)e^{-st}dt$$

を用いると，図 3.1 の回路の入出力関係は，

$$\hat{y}(s) = \frac{1}{RCs + 1}\ \hat{x}(s)$$

と書くことができる．この関係を次のような図で表現したものが**ブロック線図** (block diagram) である．



ブロック線図を用いれば，信号の流れが一目瞭然となる利点がある．このブロック線図の「箱」は入出力関係を表し，**システム** (system) と呼ばれる．また上の例は入出力関係が線形であり，特に**線形システム** (linear system) と呼ばれる．一般に，システム f が線形システムであるとは，関数 f が入力 x に関して線形であることである．すなわち，任意の入力 x_1, x_2 と任意のスカラー a_1, a_2 に対して，

$$f(a_1x_1 + a_2x_2) = a_1f(x_1) + a_2f(x_2)$$

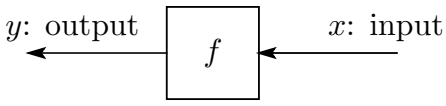
[*2] 工学，特に電気工学や制御工学では虚数単位 $\sqrt{-1}$ を i ではなく j であらわす．これは i という記号が電流を表すために使用されるためである．

が成り立つとき，システム f を線形システムという．なお，線形システムの場合，関数の後の括弧を省略し，fx と書くことが多い．

上記のブロック線図表現を一般の関数に応用してみよう．次の関数 f を考える．

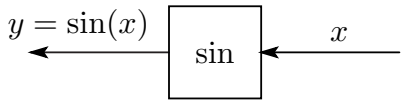
$$y = f(x) \tag{3.4}$$

この関数において，x を入力，y を出力と考え，f を入出力関係を示すシステムと考える．すると，(3.4) の関数関係は次のブロック線図で表すことができる．

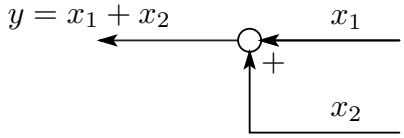


では，いくつかの関数に対して，ブロック線図を描いてみよう．

1. $y = \sin(x)$．関数 sin を入力 x，出力 y のシステムと考えると，ブロック線図は次の図となる．



2. $y = x_1 + x_2$．これは，x_1 と x_2 という 2 つの入力に対して，$y = x_1 + x_2$ という出力を持つシステムであると考える．このような足し算のブロック線図は，特に**加算点** (summing junction) と呼ばれ，次のように表される．

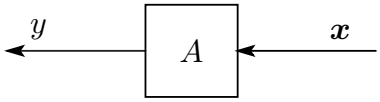


これはもちろん，箱の形でも表すことができる．すなわち，行列を用いて

$$y = x_1 + x_2 = A\boldsymbol{x}$$

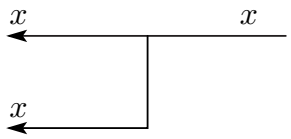
$$A = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3.5}$$

と書けるので[*3]，次のブロック線図が得られる．



[*3] 本書では，ベクトルを太文字，行列を大文字で記述する．

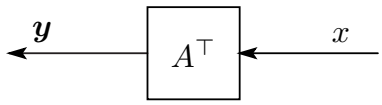
3. $\boldsymbol{y} = [x \;\; x]^\top$. これは，入力 x のコピーを作り並列に並べる操作であり，**引き出し点** (takeoff point) と呼ばれる次のブロック線図で表される．



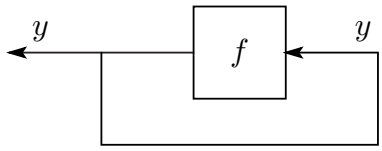
(3.5) で定義した行列 A を使えば，これも箱を用いたブロック線図で表現することができる．すなわち，

$$\boldsymbol{y} = [x \;\; x]^\top = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x = A^\top x$$

であるので，次のブロック線図が得られる．

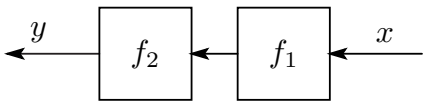


4. $y = f(y)$. これは方程式であるが，関数 f をシステムとし，y をシステム f の出力とみなすと，システム f には出力 y が再び入力される．このようなシステムを**フィードバックシステム** (feedback system) と呼び，次のブロック線図で表現される．

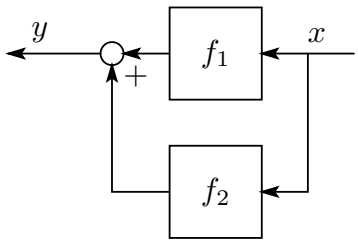


このシステム自体には入力が存在しない（システムの外側からの矢印がない）ことに注意されたい．

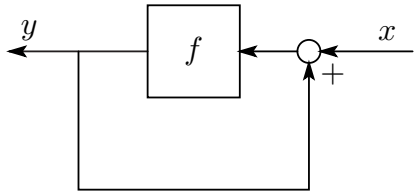
練習問題 8　1. 次のブロック線図の数式表現を求めよ．



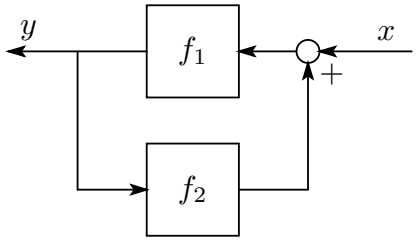
(a)



(b)

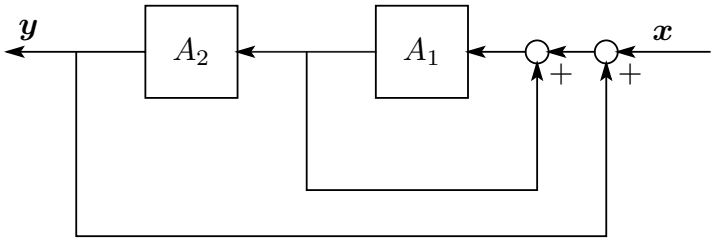


(c)



(d)

2. 次のブロック線図の数式表現を求めよ．ただし，$\boldsymbol{x}$, $\boldsymbol{y}$ はベクトル，A_1, A_2 は行列であり，$I - A_1$ および $I - A_2(I - A_1)^{-1}A_1$ は正則とする[*4]．



(e)

3. 次の数式表現のブロック線図を求めよ．
   (a) $y = f_1(x) + f_2(x)$
   (b) $y = f_1(x) + f_2(y)$
   (c) $y = f_1\left(x + f_2(x)\right)$
   (d) $y = f_1(x), \quad x = f_2(y)$

3.3　反復法のブロック線図表現

ここでは 3.1 節で述べた反復法をブロック線図で表現する方法を学ぶ．

3.3.1　信号とシステム

前節で考察した関数 $y = f(x)$ では，一つの数（またはベクトル）x に対して，一つの数（またはベクトル） y が関数 f によって定まる．入力 x をシステム f に入れて出力 y が出てくる操作は一回しか行われない．

この操作を何回も続けて行うことを考える．すなわち，まずはじめに $x[1]$ という入力を f に入れ，出力を $y[1]$ とする．

$$y[1] = f(x[1])$$

[*4] I は単位行列を表す

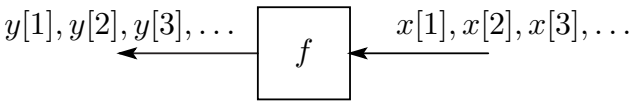
次に，また別の $x[2]$ という入力を f に入れ，出力を $y[2]$ とすると，

$$y[2] = f(x[2])$$

となる．同じことを N 回繰り返すと，N 個の入力 $\{x[1], x[2], \ldots, x[N]\}$ と N 個の出力 $\{y[1], y[2], \ldots, y[N]\}$ が得られ，その関係は次式で表される．

$$\{y[1], y[2], \ldots, y[N]\} = \{f(x[1]), f(x[2]), \ldots, f(x[N])\}$$

この関係をブロック線図で表現すると以下のようになる．



ここで添え字 $\{1, 2, \ldots, N\}$ を時刻と考える．例えば 1 秒おきにシステム f に $x[n]$ $(n = 1, 2, \ldots, N)$ を入れることを考える．そうすると，N 個の入力 $\{x[1], x[2], \ldots, x[N]\}$ と N 個の出力 $\{y[1], y[2], \ldots, y[N]\}$ は 1 秒おきに値が定義された**信号** (signal) とみなすことができる．ここで，考えている時間軸が離散的なのでこのような信号を特に**離散時間信号** (discrete-time signal) と呼び，システム f を**離散時間システム** (discrete-time system) と呼ぶ[*5]．すなわち，時刻 $n = 1$ から次々と信号 $x[n]$ $(n = 1, 2, \ldots, N)$ がシステム f に入力し，別の信号 $y[n]$ $(n = 1, 2, \ldots, N)$ が次々と得られると考えるわけである．

3.3.2 静的なシステムと動的なシステム

前節で考察したシステム f は，ある時刻 n に入力 $x[n]$ を入れると，同じ時刻に（すなわち瞬時に）$y[n]$ が出力される．また，時刻 n での出力 $y[n]$ は，その他の時刻，例えば $n+1$ や $n-2$ 等における入力（$x[n+1]$ や $x[n-2]$）には依存しない．このようなシステムは，**静的なシステム** (static system) または**記憶を持たないシステム** (memoryless sytem) と呼ばれる．

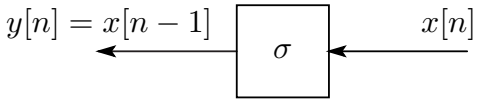
一方，次のようなシステム σ を考える[*6]．

$$y[n] = \sigma x[n] = x[n-1], \quad n = 1, 2, \ldots \tag{3.6}$$

このシステムのブロック線図表現は以下で与えられる．

[*5] 一方，3.2 節のはじめで考察した電気回路は，時間が連続的であり，**連続時間システム** (continuous-time system) と呼ばれる．連続時間システムはこれ以降ほとんど出てこないので，本書では，明記しない限り，システムといえば離散時間システムを意味するものとする．

[*6] σ は線形システムであるので，$\sigma(x[n])$ を $\sigma x[n]$ と書く．



これは，時刻 $n = 1, 2, 3, \ldots$ での入力 $x[n]$ を次の時刻 $n + 1$ まで記憶して出力するシステムで，論理回路におけるレジスタ回路（記憶回路），またはフリップフロップと同じ働きをする．数学的には，σ は**シフト作用素** (shift operator) と呼ばれる．これには，時刻を n から $n - 1$ にシフトさせるという意味合いがある．このように内部に記憶を持つシステムを**動的なシステム** (dynamical system) または**記憶を持つシステム** (system with memory) と呼ぶ．システム σ の出力のタイミングは，入力信号の時刻 $n = 0, 1, 2, \ldots$ のタイミングと同期させる必要があることに注意する（この要請は，論理回路でも同じである）．すなわち，上のブロック線図には表示されていないが，このシステムにはクロック同期回路が必要である．

また，(3.6) において，時刻 $n = 1$ の出力は $y[1] = x[0]$ であるが，$x[0]$ は入力に含まれない．ここでは，時刻 $n = 1$ においてはシステム σ の**初期値** (initial value) $x[0]$ が $y[1]$ として出力されるものと考える．すなわち，システムを動かす前（$x[1]$ が入力する前）にレジスタに記憶されているデータ $x[0]$ が時刻 $n = 1$ で出力されると考えることとする．初期値は，任意の値に決めることが可能な場合もあれば，他の要因により決まることもある．特に初期値を $x[0] = 0$ にセットすることを**リセット** (reset) という．

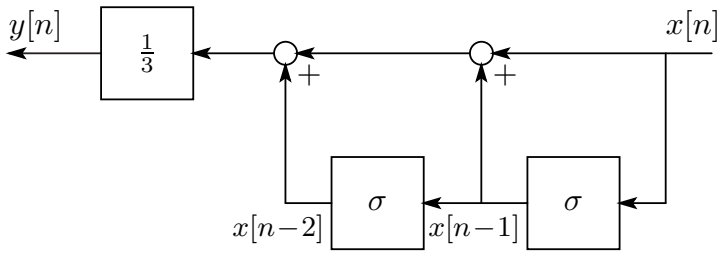
動的なシステムには，様々なものがある．例えば次のシステムも動的システムの一つである．

$$y[n] = \frac{x[n] + x[n-1] + x[n-2]}{3}, \quad n = 2, 3, 4, \ldots \tag{3.7}$$

このシステムは，**移動平均システム** (moving average system) または**移動平均フィルタ** (moving average filter) と呼ばれる．移動平均システム (3.7) は，シフト作用素 (3.6) を使って，次のように書ける．

$$\begin{aligned} y[n] &= \frac{1}{3}\left(x[n] + \sigma x[n] + \sigma\sigma x[n]\right) \\ &= \frac{1}{3}\left(1 + \sigma + \sigma^2\right) x[n], \qquad n = 2, 3, 4, \ldots \end{aligned}$$

この表現から，このシステムには 2 つのレジスタが存在し，その 2 つのレジスタにそれぞれ初期値 $x[0]$ と $x[1]$ を設定する必要があることがわかる．対応するブロック線図を以下に示す．



練習問題 9 次の動的システムを σ を用いて表現し，そのブロック線図を描け．

$$y[n] = \frac{4x[n] + 3x[n-1] + 2x[n-2] + x[n-3]}{10}, \quad n = 3, 4, 5, \ldots$$

また，このシステムに必要な初期値は何個か？

3.3.3 反復法のブロック線図

以上の準備の下で，集合 X 上の一般の写像 $\phi : X \to X$ による反復法

$$x[n+1] = \phi(x[n]), \quad n = 0, 1, 2, \ldots, \tag{3.8}$$

のブロック線図を描いてみよう．まず，上の式はレジスタ（シフト作用素） σ を用いて，次のように書ける．

$$x[n] = \phi(\sigma x[n]), \quad n = 1, 2, \ldots$$

3.3.2 節で指摘したように，このシステムには動的システム σ が存在し，初期値 $x[0]$ が必要である．この初期値を $x[0] = x_0 \in X$ に設定する．上の式を変形すると，

$$\begin{aligned} x[n] &= \phi(y[n]) \\ y[n] &= \sigma x[n], \quad n = 1, 2, \ldots \end{aligned}$$

と書けることがわかる．これより，反復法は静的なシステム ϕ と動的なシステム σ に分離できることがわかり，ブロック線図は図 3.2 のようになる．このブロック線図から以下のことがわかる．

- 反復法は静的なシステム ϕ と動的なシステム（シフト作用素） σ を持つ．
- 反復法は**フィードバック** (feedback) の構造を持つ．
- 反復法は**無限ループ** (infinite loop) である（停止しない）．

反復法を停止させるためには，ある時刻で無限ループを抜けなければならない．停止させる条件は，例えば，$x[n]$ と $x[n+1]$ との差が非常に小さい，すなわち，$|x[n] - x[n+1]|$ があらかじめ設定した値（例えば，10^{-5} など）未満になったときに停止させるという条件となる．このような条件を加えても，数列 $\{x[n]\}$ がある値に収束しなければループは

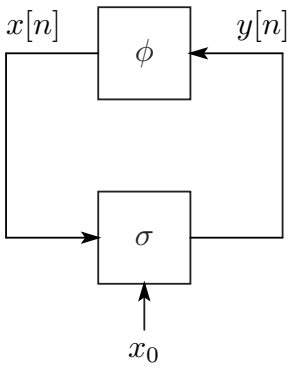


図 3.2　反復法のブロック線図

停止しない．したがって，反復法を正しく動作させるためには，$\{x[n]\}$ はある値に収束する必要がある．これは，制御工学の言葉で言うと「静的なシステム ϕ と動的なシステム σ とのフィードバックシステムは安定でなければならない」ということになる．すなわち，数値計算における反復法の収束性と制御工学におけるフィードバックシステムの安定性との間には密接な関係があることがわかる．

3.4　制御工学と数値計算

前節で述べたように，反復法における収束性は制御工学における安定性と密接な関係がある．反復法をブロック線図で図 3.2 のように表すと，収束性の条件として「システム ϕ の入出力比はあまり大きくはできない」ということが直感的にわかる．また，収束の速度を速めたり，丸め誤差に強い反復法を作ることは，うまく ϕ を選ぶこともさることながら，σ の部分をうまく設計することも有効であることがわかる．このような視点は，ブロック線図を描くことによってはじめて得られることであり，反復法をブロック線図で表現することの有効性の一例である．なお，ブロック線図に基づいた安定性や収束性の詳しい説明は，5 章を参照されたい．

3.5　Xcos でのシミュレーション

フリーソフト Scilab に付属する Xcos を使えば，反復法をフィードバックシステムで書き換えたときのシミュレーションを簡単に行うことができる．図 3.3 に $\sqrt{a}$ の近似値を求める反復法（例題 2 を参照）

$$x[n+1] = \frac{1}{2}\left(x[n] + \frac{a}{x[n]}\right), \quad n = 0, 1, 2, \ldots$$

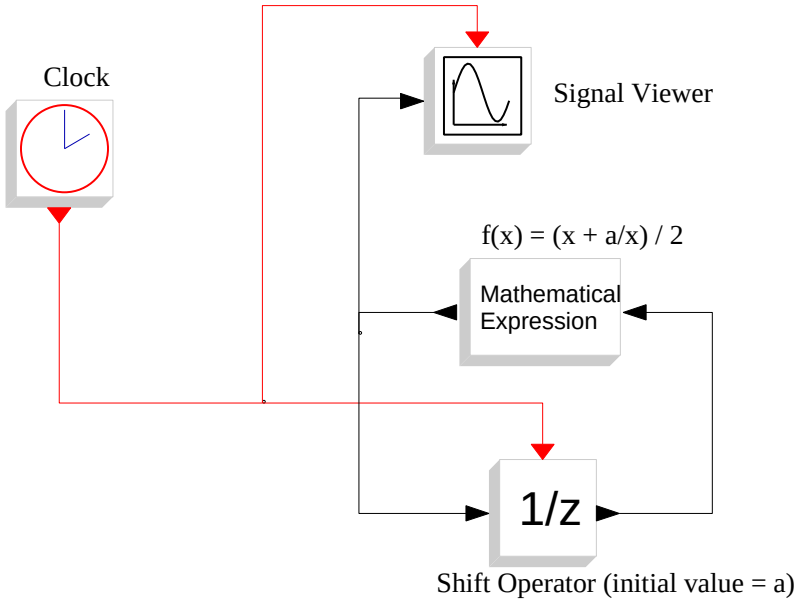


図 3.3 Xcos による反復法のシミュレーション

の Xcos によるシミュレーションを示す．ここで `Mathematical Expression` と書かれたブロックは，反復法における写像

$$\phi(x) = \frac{1}{2}\left(x + \frac{x}{a}\right)$$

を示している．また $1/z$ と書かれたブロックはシフト作用素 σ を表す．この z は σ の逆作用素であり，

$$zx[n] = x[n+1], \quad n = 0, 1, 2, \ldots$$

で定義される（シフト作用素 σ の定義 (3.6) と比べよ）．他のブロックの説明や設定方法は第 1 章の 1.2.2 節を参照してほしい．

このシミュレーションでは，この章で紹介したブロック線図の概念が使われていることがわかる．また，前節までの議論で省略したクロックによる同期も Xcos のシミュレーションでは必要である．このシミュレーションで，上記の反復法により $\sqrt{a}$ に収束していることや収束の速さなどを確認してほしい．また，a の値を負にして（すなわち $\sqrt{a}$ が虚数になる場合に）シミュレーションし， $x[n]$ がどのような振る舞いをするのかを確かめてみるのも面白い．

3.6　さらに勉強するために

第 3.2 節で述べた Laplace 変換による電気回路の解析は，[7, 26] が参考になる．残念ながら日本語の本で上記の 2 冊ほどわかりやすい本は存在しない（電気回路の教科書 [7] には日本語の訳本がある）．ブロック線図や動的システム，フィードバックなどに関しては，上記の [26] のほか，制御工学の教科書 [6, 36] などがわかりやすい．

第 4 章

非線形方程式の反復解法

この章では，方程式 $f(x) = 0$ の近似解を求めるために，初期値 $x[0]$ を与えて次の繰り返し計算

$$x[n+1] = \phi(x[n]), \quad n = 0, 1, 2, \ldots$$

を行うことを考える．この繰り返し計算によって生成される数列 $\{x[n]\}$ が方程式 $f(x) = 0$ の真の解 x^* に収束すれば，十分大きな N に対して $x[N]$ は方程式の近似解となる．

4.1 方程式の反復解法

工学では，しばしば以下の方程式を満たす x を求めることが必要となる．

$$f(x) = 0. \tag{4.1}$$

もし $f(x)$ が 2 次多項式などの簡単な式で表されるのなら，この方程式を満たす解を求めることは易しい．すなわち，四則演算などの簡単な式変形によって厳密解は求まる．しかし，工学であらわれる多くの方程式では，簡単な式の変形だけで解が求まることはきわめてまれである．例えば，5 次以上の多項式の根はそのような式変形では求めることが出来ない．このような場合には，コンピュータを用いた近似解法が必要となる．コンピュータにより方程式の近似解を求めるためには，本章で述べる**反復法** (iteration method) が有効である．方程式 (4.1) の近似解を求める反復法の手順は以下のとおりである．

1. 方程式 $f(x) = 0$ を $x = \phi(x)$ の形に等価変形する．
2. 初期値 $x[0]$ を選ぶ（解のおおよその数値がわかっている場合には，それに近い値に設定する）．
3. 十分大きな N をとり，反復法 $x[n+1] = \phi(x[n])$, $n = 0, 1, 2, \ldots, N$ を実行する．

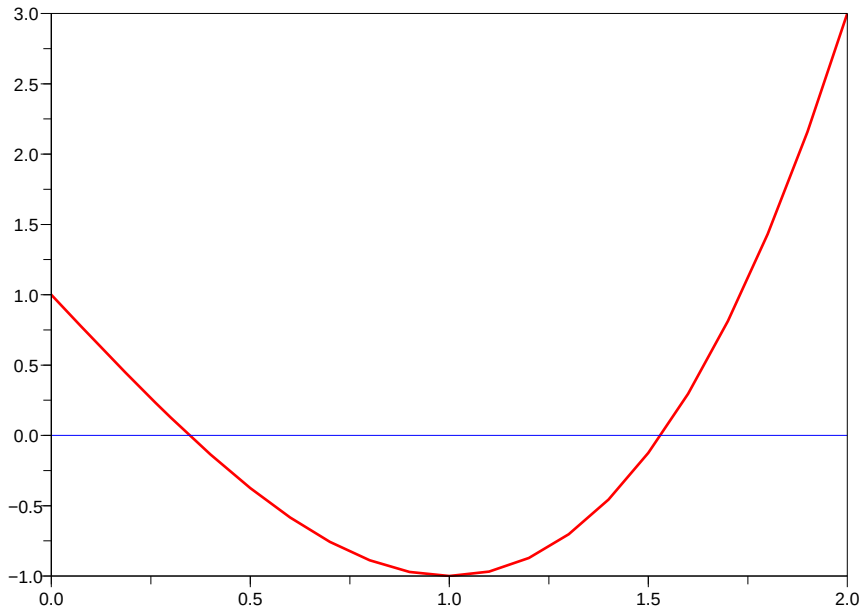


図 4.1　$y = x^3 - 3x + 1$ のグラフ

4. 得られた $x[N+1]$ を方程式 $f(x) = 0$ の近似解とする．

上記の反復法でなぜうまくいくのかを考えてみよう．上の反復法は収束し，

$$\lim_{n\to\infty} x[n] = a \in \mathbb{R}$$

であると仮定する．また，関数 $\phi(x)$ は $x = a$ で連続であるとする．このとき，$x[n+1] = \phi(x[n])$, $n = 0, 1, 2, \ldots$ であったので，$n \to \infty$ の極限で

$$a = \lim_{n\to\infty} x[n+1] = \lim_{n\to\infty} \phi(x[n]) = \phi\left(\lim_{n\to\infty} x[n]\right) = \phi(a)$$

が成り立つ．すなわち，a は方程式 $x = \phi(x)$ の解であり，すなわち方程式 $f(x) = 0$ の解であることがわかる．したがって十分大きな N をとれば上の反復法で得られる $x[N+1]$ は方程式 $f(x) = 0$ の近似解であることがわかる．なお，上記のように $a = \phi(a)$ を満たす a を写像 ϕ の**不動点** (fixed-point) と呼ぶ．

次に簡単な例題で反復法を実行してみよう．次の方程式を考える．

$$f(x) = x^3 - 3x + 1 = 0 \tag{4.2}$$

図 4.1 に $y = x^3 - 3x + 1$ のグラフを示す．このグラフより，方程式 $f(x) = 0$ の真の

解は，区間 $(0,1)$ と $(1,2)$ にそれぞれ一つづつ存在することがわかる．なおこの事実は，図 4.1 のようなグラフを描かなくとも，微分積分学の**中間値の定理** (intermediate value theorem)[*1] によって示すことができる．

練習問題 10 方程式 (4.2) が区間 $(0,1)$ と区間 $(1,2)$ にそれぞれ解を持つことを中間値の定理を用いて示せ．

反復法でこれらの近似値を求めるために，方程式 (4.2) を次のように変形する．

$$x = \frac{1}{3}\left(x^3 + 1\right). \tag{4.3}$$

すなわち，

$$\phi(x) = \frac{1}{3}\left(x^3 + 1\right)$$

とおく．初期値を $x[0] = 0.5$ とし，反復法

$$x[n+1] = \frac{1}{3}\left(x[n]^3 + 1\right), \quad n = 0, 1, 2, \ldots \tag{4.4}$$

を実行する．この反復法の Scilab のプログラムを以下に示す．

```
// Initial Value
x0 = 0.5;

// Number of Iteration
N = 19;

// Iteration
x = x0;
for n = 0: N
  printf('x[%d] = %f\n',n,x)
  x = (x^3+1)/3;
end
```

このプログラムの実行結果を図 4.2 に示す．この結果より，初期値 $x[0] = 0.5$ から出発した反復法により，$0.347296 \in (0,1)$ 付近の解に収束していることがわかる．

次に，初期値を $x[0] = 2$ としたときの実行結果を図 4.3 に示す．この結果より，反復

[*1] 中間値の定理に関しては，付録を参照のこと．

```
x[0] = 0.500000          x[10] = 0.347296
x[1] = 0.375000          x[11] = 0.347296
x[2] = 0.350911          x[12] = 0.347296
x[3] = 0.347737          x[13] = 0.347296
x[4] = 0.347350          x[14] = 0.347296
x[5] = 0.347303          x[15] = 0.347296
x[6] = 0.347297          x[16] = 0.347296
x[7] = 0.347296          x[17] = 0.347296
x[8] = 0.347296          x[18] = 0.347296
x[9] = 0.347296          x[19] = 0.347296
```

図 4.2　反復法 (4.4) の実行結果．初期値は $x[0] = 0.5$.

```
x[0] = 2.000000e+000
x[1] = 3.000000e+000
x[2] = 9.333333e+000
x[3] = 2.713457e+002
x[4] = 6.659590e+006
x[5] = 9.845125e+019
x[6] = 3.180845e+059
x[7] = 1.072769e+178
x[8] = 1.#INF00e+000
x[9] = 1.#INF00e+000
```

図 4.3　反復法 (4.4) の実行結果．初期値は $x[0] = 2$.

法により生成された数列は発散することがわかる[*2]．これら Scilab の実験により，反復法は**初期値の位置によって，方程式の解に収束したり発散したりする**ことがわかる．

次に方程式 (4.2) に対して，(4.3) のように変形するのではなく，$x > 0$ として

$$x = \frac{3x - 1}{x^2} = \phi(x)$$

と変形し反復法

$$x[n+1] = \frac{3x[n] - 1}{x[n]^2}, \quad n = 0, 1, 2, \ldots \tag{4.5}$$

[*2] `x[8]` と `x[9]` に現れた `#INF` は ∞ を表し，オーバーフローが起こったことを示している．2.3.4 節を参照のこと．

```
x[0] = 1.500000              .
x[1] = 1.555556              .
x[2] = 1.515306              .
x[3] = 1.544287          x[33] = 1.532090
x[4] = 1.523326          x[34] = 1.532088
x[5] = 1.538438          x[35] = 1.532089
x[6] = 1.527517          x[36] = 1.532089
x[7] = 1.535395          x[37] = 1.532089
x[8] = 1.529705          x[38] = 1.532089
x[9] = 1.533812          x[39] = 1.532089
```

図 4.4 反復法 (4.5) の実行結果．初期値は $x[0] = 1.5$.

を実行してみよう．初期値を $x[0] = 1.5$ として，Scilab で反復法を実行すると図 4.4 の結果が得られた．今度は，初期値 $x[0] = 1.5$ に対して，$1.532089 \in (1, 2)$ 付近の解に収束していることがわかる．しかし，収束の速さは前の例に比べて遅い．この実験から，反復法は ϕ **の作り方によって収束性や収束の速さが異なる**ことがわかる．

練習問題 11 1. 中間値の定理を用いて，次の方程式の解が少なくとも 1 つ存在する区間を求めよ．

(a) $x^3 + 2x^2 + 10x - 20 = 0$

(b) $x^2 = e^x$

(c) $x = \cos x$

(d) $x^x = 2$

2. Scilab のプログラムを使用し，上の方程式の近似解を反復法により求めよ．
3. 関数電卓で任意の正の数を入力し，ルート（$\sqrt{}$）ボタンを連打すると，1 に収束していくことを反復法を使って説明せよ．
4. 上の問題を参考に，$x = \cos x$ の近似解を関数電卓を使って簡単に求める方法を考えよ．

4.2 代表的な反復法

ここでは，方程式 $f(x) = 0$ を解くための代表的な反復法を紹介し，その収束の速さを議論する．一般的な反復法の収束性の解析は次章にて行う．

4.2.1　2 分法

前節で調べた 3 次方程式

$$f(x) = x^3 - 3x + 1 = 0$$

を考える．この $f(x)$ に $x = 1, x = 2$ を代入すると，

$$f(1) = -1 < 0, \quad f(2) = 3 > 0$$

であるので，中間値の定理から方程式 $f(x) = 0$ は区間 $(1, 2)$ に少なくとも一つ解を持つことがわかる．この区間の端点を $a[0] = 1, b[0] = 2$ とおく．方程式の解の一番粗い近似値 $x[0]$ として，区間の中点，すなわち

$$x[0] = \frac{a[0] + b[0]}{2} = \frac{1 + 2}{2} = \frac{3}{2}$$

を採用する．このとき，方程式 $f(x) = 0$ の真の解 x^* と近似値 $x[0]$ との誤差（近似誤差）は，区間 $(1, 2)$ の幅が 1 であるので

$$|x^* - x[0]| \leq \frac{1}{2}$$

と評価できる．次に， $x = x[0] = 3/2$ のときの $f(x)$ の値を求めてみよう．

$$f\left(x[0]\right) = f\left(\frac{3}{2}\right) = \frac{27}{8} - \frac{9}{2} + 1 = -\frac{1}{8} < 0$$

であるので，ふたたび中間値の定理から，方程式 $f(x) = 0$ の解は区間 $(3/2, 2)$ に存在することがわかる．すなわち，解の存在する区間が半分に狭められたわけである．この区間の端点を $a[1] = 3/2, b[1] = 2$ とおき，新しい近似値 $x[1]$ を区間 $(3/2, 2)$ の中点，すなわち

$$x[1] = \frac{a[1] + b[1]}{2} = \frac{3/2 + 2}{2} = \frac{7}{4}$$

にとる．このとき，方程式 $f(x) = 0$ の真の解 x^* と近似値 $x[1]$ との近似誤差は

$$|x^* - x[1]| \leq \frac{1}{2}|x^* - x[0]| \leq \frac{1}{4}$$

と評価される．以下同様のことを繰り返すと，真の解 x^* の存在する区間がどんどん狭められ，第 n ステップ目での近似解は，

$$x[n] = \frac{a[n] + b[n]}{2}$$

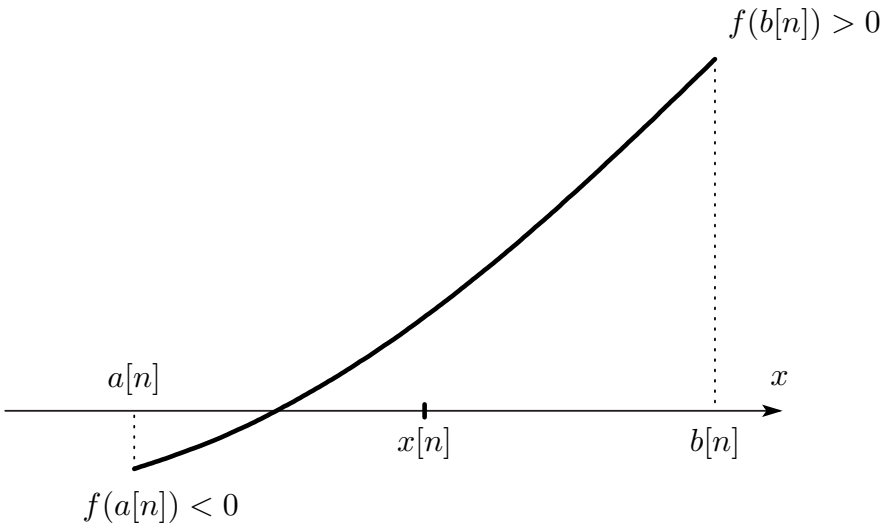


図 4.5　2 分法: 方程式 $f(x) = 0$ の近似解 $x[n]$ は $a[n]$ と $b[n]$ の中点である.

となり，真の解 x^* と近似解 $x[n]$ との誤差は

$$|x^* - x[n]| \leq \left(\frac{1}{2}\right)^n |x^* - x[0]| \leq \left(\frac{1}{2}\right)^{n+1}$$

と評価される．これより

$$\lim_{n\to\infty} x[n] = x^*$$

すなわち，数列 $\{x[n]\}$ は真の解 x^* に収束することがわかる．このような反復法を **2 分法** (bisection method) と呼ぶ．図 4.5 に 2 分法の近似解 $x[n]$ のとり方を示す．また，2 分法のアルゴリズムを以下に示す.

アルゴリズム 1 (2 分法) 閉区間 $[a, b]$ 上の連続関数 $f(x)$ が与えられ,

$$f(a) < 0, \quad f(b) > 0$$

を満たすとする[*3].

1. $a[0] := a$, $b[0] := b$ とおく.
2. 繰り返し回数 N を設定する.
3. $n = 0$ から $n = N$ の間，以下を繰り返す.
   (a) $x[n] := \frac{a[n]+b[n]}{2}$
   (b)
   - もし $f(x[n]) = 0$ ならば $n := N$.
   - もし $f(x[n]) < 0$ ならば $a[n+1] := x[n]$, $b[n+1] := b[n]$.

[*3] もし $f(a) > 0$ かつ $f(b) < 0$ のときは，$-f(x)$ を新たに $f(x)$ とおけばよい.

- もし $f(x[n]) > 0$ ならば $a[n+1] := a[n]$, $b[n+1] := x[n]$.

(c) $n := n + 1$.

4. $x[N]$ が方程式 $f(x) = 0$ の近似解であり，真の解との誤差は $2^{-(N+1)}(b-a)$ 以下となる．

2 分法における第 n ステップでの近似誤差を

$$\varepsilon[n] = |x^* - x[n]|, \quad n = 0, 1, 2, \ldots$$

とおくと，上のアルゴリズムより 2 分法では，任意の $n \geq 0$ に対して

$$\varepsilon[n+1] \leq \frac{1}{2}\varepsilon[n]$$

が成り立つ．一般に，ある整数 $n_0 \geq 0$ と実数 $C \in (-1, 1)$ が存在して，任意の $n \geq n_0$ に対して

$$\varepsilon[n+1] \leq C\varepsilon[n]$$

が成り立つとき，その反復法は 1 **次収束** (linear convergence) であるという．すなわち 2 分法は 1 次収束の反復法であるといえる．また，ある整数 $n_0 \geq 0$ と実数 $C \in (-1, 1)$, $p \geq 1$ が存在して，任意の $n \geq n_0$ に対して

$$\varepsilon[n+1] \leq C\varepsilon[n]^p, \quad n = 0, 1, 2, \ldots \tag{4.6}$$

が成り立つとき，その反復法は，p **次収束** (p-th order convergence) であるといい，特に $p > 1$ のとき，**超 1 次収束** (superlinear convergence) であるという．また，p を**収束の次数** (order of convergence) と呼ぶ．なお，$p > 1$ のとき，$|\varepsilon[n]| > 1$ だとすると $|\varepsilon[n]|^p$ は $|\varepsilon[n]|$ より大きくなり，不等式 (4.6) は意味を持たない．したがって，$p > 1$ のときの p-次収束性は厳密解の近傍でのみ議論される．

4.2.2 はさみうち法

2 分法は単純に区間を中点で分割する方法であるが，端点での $f(x)$ の値を考慮して，分割する点を変更するほうが収束は速くなることが期待できる．

2 分法では，第 n ステップ目の近似解 $x[n]$ は $a[n]$ と $b[n]$ の中点にとった（図 4.5 を参照）．この近似解 $x[n]$ を，図 4.6 のように 2 点 $(a[n], f(a[n]))$ と $(b[n], f(b[n]))$ を結ぶ直線と x 軸との交点にとる．真の解の付近で $f(x)$ が滑らかであれば（すなわち，直線に近ければ），この方法は 2 分法よりも収束が速い．これを**はさみうち法** (regula falsi) という．

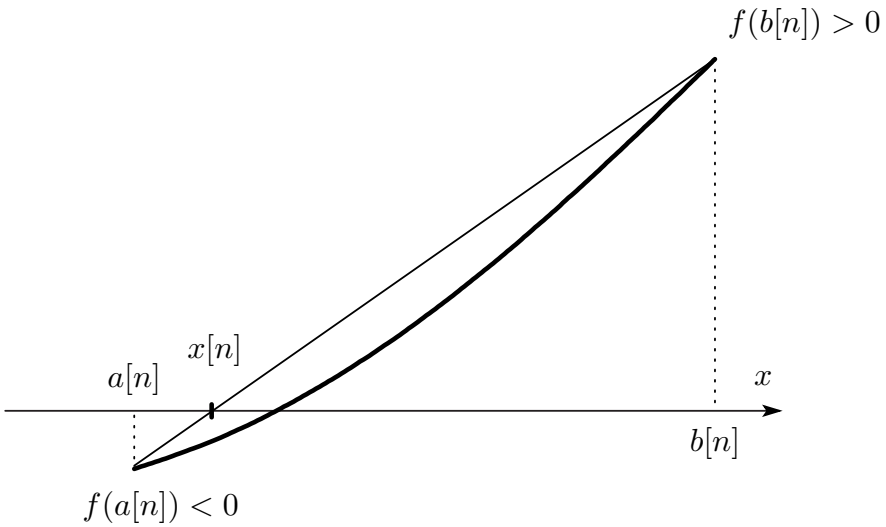


図 4.6 はさみうち法: 方程式 $f(x) = 0$ の近似解 $x[n]$ は点 $(a[n], f(a[n]))$ と点 $(b[n], f(b[n])$ を結ぶ直線と x 軸との交点である.

第 n ステップの近似解 $x[n]$ は次のように計算できる.

$$x[n] = \frac{f(b[n]) \cdot a[n] - f(a[n]) \cdot b[n]}{f(b[n]) - f(a[n])}, \quad n = 0, 1, 2, \ldots$$

これにもとづき, はさみうち法のアルゴリズムは以下で与えられる.

アルゴリズム 2 (はさみうち法) 閉区間 $[a, b]$ 上の連続関数 $f(x)$ が与えられ,

$$f(a) < 0, \quad f(b) > 0$$

を満たすとする.

1. $a[0] := a$, $b[0] := b$ とおく.
2. 繰り返し回数 N を設定する.
3. $n = 0$ から $n = N$ の間, 以下を繰り返す.
   (a) $x[n] := \frac{f(b[n]) \cdot a[n] - f(a[n]) \cdot b[n]}{f(b[n]) - f(a[n])}$
   (b)
   - もし $f(x[n]) = 0$ ならば $n := N$.
   - もし $f(x[n]) < 0$ ならば $a[n+1] := x[n]$, $b[n+1] := b[n]$.
   - もし $f(x[n]) > 0$ ならば $a[n+1] := a[n]$, $b[n+1] := x[n]$.

   (c) $n := n + 1$.
4. 方程式 $f(x) = 0$ の近似解は $x[N]$ となる.

4.2.3 割線法

2 分法，はさみうち法はともに，$f(a)\cdot f(b)<0$ となる 2 点 a,b を見つける必要があった．しかし，このような a, b を見つけることは必ずしも容易ではない．例えば，次の関数を考える．

$$f(x)=3x^2+\frac{1}{\pi^4}\log[(\pi-x)^2]+1 \tag{4.7}$$

この関数は $x=\pi$ に特異点を持ち，$x=\pi$ 付近にゼロ点，すなわち $f(x)=0$ となる点が存在する．そして，この関数が負の値をとるのは $(\pi-10^{-667},\pi+10^{-667})$ の区間だけである [21]．すなわち，この非常に狭い区間以外では，常に $f(x)\geq 0$ となる．このような関数に対して，$f(a)\cdot f(b)<0$ となる a, b を見つけるのは非常に困難であることがわかるだろう．

割線法 (secant method) は，2 分法やはさみうち法のように $f(a)\cdot f(b)<0$ となるような a と b を見つける必要のない方法であり，以下の繰り返し計算で $f(x)=0$ の近似解を計算する．

$$x[n+1]=\frac{f(x[n])\cdot x[n-1]-f(x[n-1])\cdot x[n]}{f(x[n])-f(x[n-1])},\quad n=1,2,\ldots \tag{4.8}$$

ただし，初期値として 2 点 $x[0]$ と $x[1]$ $(x[0]\neq x[1])$ をあらかじめ与えておく．割線法によって近似解が得られる様子を図 4.7 に示す．割線法は，もし収束すればその収束の次数は，**黄金比** (golden ratio) $(1+\sqrt{5})/2\approx 1.63$ となることが知られており [22]，超 1 次収束であることがわかる．ただし割線法は 2 分法やはさみうち法のように解を囲みこまず，また反復法が収束しない場合もあることに注意する．

4.2.4 Newton 法

割線法では，2 点 $(x[n-1],f(x[n-1]))$ と $(x[n],f(x[n]))$ を結んで出来る直線と x 軸との交点を $x[n+1]$ として近似解を得た．ここで $x[n]$ と $x[n-1]$ を近づけると，直線は曲線 $f(x)$ の $x[n]$ における接線となる．この接線と x 軸との交点を次の近似解 $x[n+1]$ とするのが Newton **法** (Newton's method) である．

Newton 法の反復法を求めよう．割線法の計算式 (4.8) を以下のように変形する．

$$\begin{aligned}x[n+1]&=\frac{f(x[n])\cdot x[n-1]-f(x[n-1])\cdot x[n]}{f(x[n])-f(x[n-1])}\\&=x[n]-f(x[n])\frac{x[n]-x[n-1]}{f(x[n])-f(x[n-1])},\quad n=1,2,\ldots\end{aligned}$$

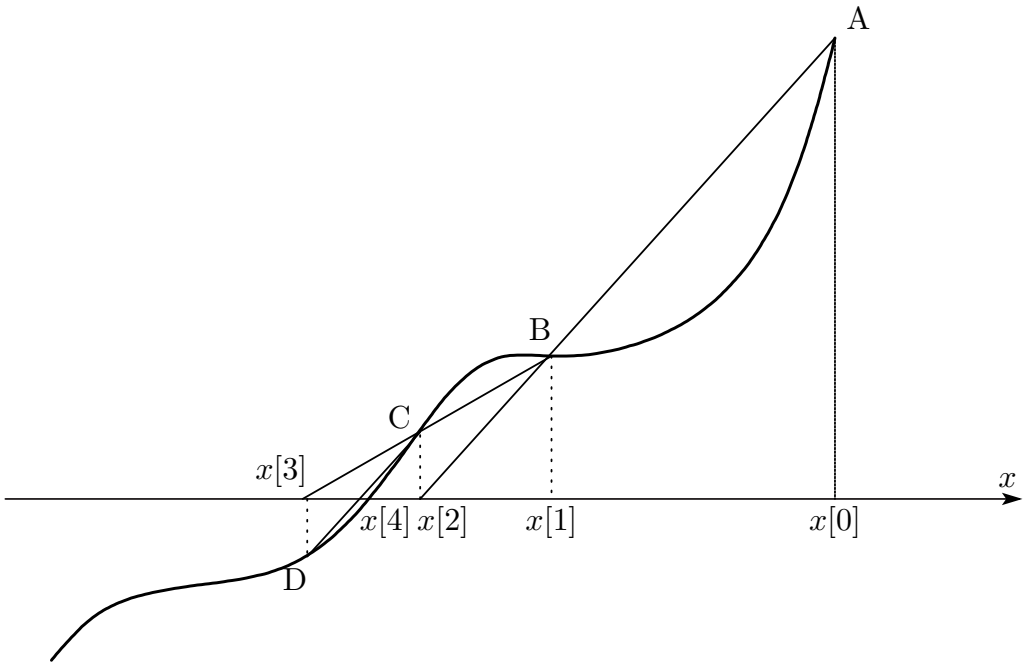


図 4.7 割線法: 初期値として $x[0]$ と $x[1]$ を与え，A と B を結ぶ直線と x 軸との交点を $x[2]$ とする．次に，B と C を結ぶ直線と x 軸の交点を $x[3]$，C と D を結ぶ直線と x 軸との交点を $x[4]$ とし，これを繰り返して近似解 $x[n]$ を得る．

ここで，関数 $f(x)$ はその定義域において微分可能かつ $f'(x) \neq 0$ とする．上式で $x[n-1] \to x[n]$ とすることにより，以下の反復法が得られる．

$$x[n+1] = x[n] - \frac{f(x[n])}{f'(x[n])}, \quad n = 0, 1, 2, \ldots$$

ただし，$x[0]$ を初期値として与える．図 4.8 に Newton 法による解の近似の様子を示す．

Newton 法の収束は非常に速く 2 **次収束** (quadratic convergence) であることが知られている[*4]．ただし，割線法と同様に，Newton 法は必ずしも収束するわけではない．

4.3 Scilab プログラム

本節では，前節で紹介した 2 分法，はさみうち法，割線法，Newton 法の Scilab プログラムを示す．これらのプログラムを用いて，収束の速さを比べたり，関数や初期値を変更して，いろいろな方程式を解いてみてほしい．

[*4] Newton 法の収束の速さは次章で詳しく述べる

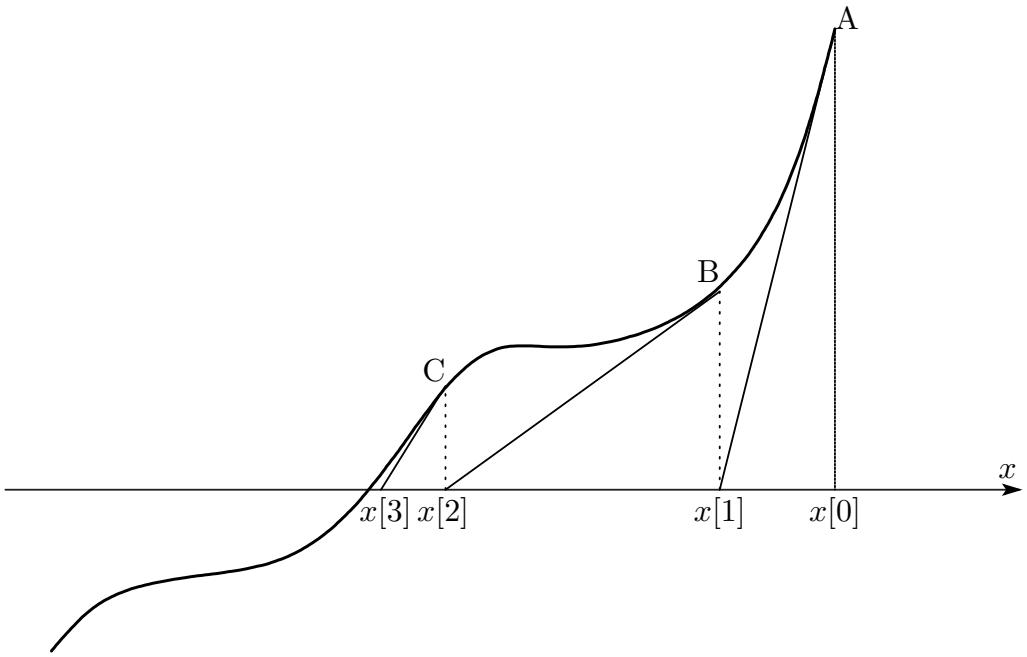


図 4.8　Newton 法: 初期値として $x[0]$ を与え，点 A において曲線 $f(x)$ の接線を引く．その接線と x 軸との交点を $x[1]$ とする．次に点 B において曲線 $f(x)$ の接線を引き，x 軸との交点を $x[2]$ とおく．$x[3]$ も同様にして得られる．

4.3.1　2 分法のプログラム

```
// Bisection Method
// f(x) = x^3 -3x +1

// Initial Interval
a = 1; b = 2;

// Number of Iteration
N = 19;

// Iteration
for n = 0: N
  x = (a + b)/2;
  printf('x[%d] = %f\n',n,x)
  fx = x^3 - 3*x + 1;
```

```
  if fx == 0
    n = N;
  elseif fx < 0
    a = x;
  else
    b = x;
  end
end
```

4.3.2 はさみうち法のプログラム

```
// Regula Falsi
// f(x) = x^3 -3x +1

// Initial Interval
a = 1; b = 2;

// Number of Iteration
N = 19;

// Iteration
for n = 0: N
  fa = a^3 - 3*a + 1;
  fb = b^3 - 3*b + 1;
  x = (fb*a - fa*b)/(fb - fa);
  printf('x[%d] = %f\n',n,x)
  fx = x^3 - 3*x + 1;
  if fx == 0
    n = N;
  elseif fx < 0
    a = x;
  else
    b = x;
  end
end
```

4.3.3　割線法のプログラム

```
// Secant method
// f(x) = x^3 -3x +1

// Initial Value x[0]=a, x[1]=b
a = 1; b = 2;

// Number of Iteration
N = 19;

// Iteration
for n = 0: N
  fa = a^3 - 3*a + 1;
  fb = b^3 - 3*b + 1;
  if fb - fa == 0
    n = N;
  else
    x = (fb*a - fa*b)/(fb - fa);
    printf('x[%d] = %f\n',n,x)
    a = b;
    b = x;
  end
end
```

4.3.4　Newton 法のプログラム

```
// Newton's method
// f(x) = x^3 -3x +1
// f'(x) = 3x^2 -3

// Initial Value x[0]
x = 2;
```

```
// Number of Iteration
N = 19;

// Iteration
for n = 0: N
  fx = x^3 - 3*x + 1;
  dfx = 3*x^2 - 3;
  if dfx == 0
    n = N;
  else
    x = x - fx/dfx;
    printf('x[%d] = %f\n',n,x)
  end
end
```

4.4 Newton 法のブロック線図表現

ここでは，前節で述べた Newton 法のブロック線図表現を考える．Newton 法の反復法

$$x[n+1] = x[n] - \frac{f(x[n])}{f'(x[n])}, \quad n = 0, 1, 2, \ldots$$

を次のように変形する．

$$\begin{aligned} x[n] &= \sigma x[n] - \sigma y[n], \\ y[n] &= g(x[n]), \quad n = 1, 2, 3, \ldots, \\ g(x) &:= \frac{f(x)}{f'(x)} \end{aligned} \tag{4.9}$$

これより，Newton 法のブロック線図は図 4.9 のように描ける．この図において，(a) と (b) は同じものを表す．なぜなら，(4.9) より，

$$(1-\sigma)x[n] = -\sigma y[n]$$

したがって，

$$x[n] = \frac{\sigma}{\sigma - 1} y[n]$$

となるからである．シフト作用素 σ に対して，このような代数的な計算を正当化する数学的理論がいわゆる**演算子法** (operational calculus) である．

ここで，システム $\sigma/(\sigma-1)$ の意味を考えてみよう．(4.9) より，

$$\begin{aligned} x[n]-x[n-1] &= -y[n-1] \\ x[n-1]-x[n-2] &= -y[n-2] \\ &\vdots \\ x[1]-x[0] &= -y[0] \end{aligned}$$

これらを辺々加えると次式が得られる．

$$x[n] = x[0] - \sum_{k=0}^{n-1} y[k]$$

ここで，$x[0]$ はレジスタ σ の初期値である[*5]．これより，システム $\sigma/(\sigma-1)$ は入力 $y[n]$ を次々と足していく（正確には，引いていく）システムであることがわかる．このような性質から，システム $\sigma/(\sigma-1)$ を**加算器** (integrator) と呼ぶ．

以上の考察から，Newton 法は，非線形システム $g(x) = f(x)/f'(x)$ と加算器 $\sigma/(\sigma-1)$ とのフィードバックシステムとみなせることがわかる．

図 4.10 に Xcos シミュレーションのブロック線図を示す．このブロック線図において，`Mathematical Expression` に $g(x) = f(x)/f'(x)$ に対応する関数を定義し，`Shift Operator 1/z` の初期値 (initial condition) を反復法の初期値 $x[0]$ とする．この Xcos シミュレーションによって，いろいろな初期値を設定してシミュレーションを実行し，収束性や収束の速さを確かめることができる．

練習問題 12 図 4.10 の Xcos のシミュレーションを実際に行い，動作を確認せよ．また，“−1” のブロック（ゲイン）の値をいろいろ変化させて（例えば，−2 などに），シミュレーションを実行し，どのような結果が得られるかを観察せよ．

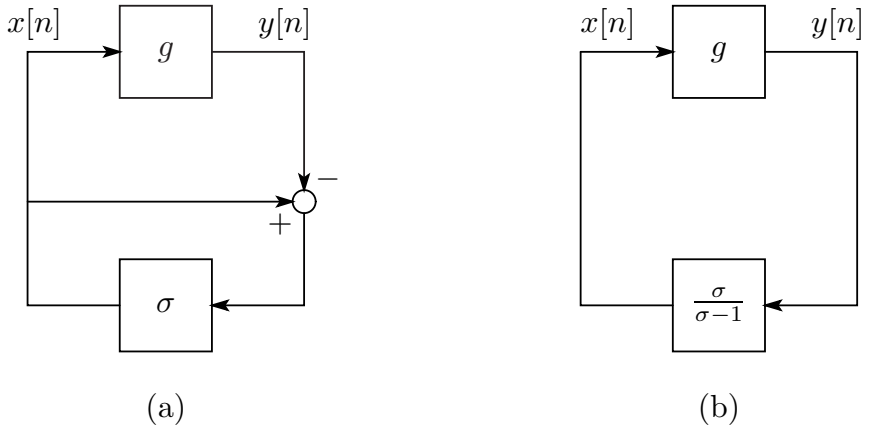


図 4.9　Newton 法のブロック線図

[*5] システム $\sigma/(\sigma-1)$ は，Newton 法のブロック線図 (a) より，レジスタ σ を一つ持つことに注意．

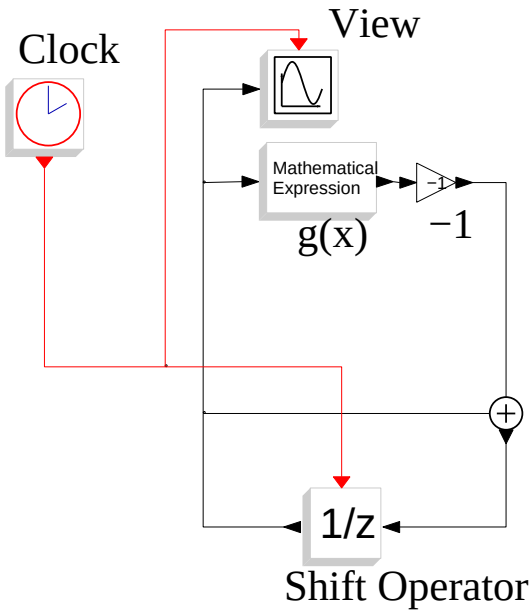


図 4.10　Xcos による Newton 法のシミュレーション

4.5　さらに勉強するために

本章で述べた代表的な反復法は，[5] や [22] を参考にした．またこれらの反復法の C 言語プログラムが [21] に掲載されている．(4.7) の病的な関数も [21, (3.0.1) 式および 9.1 節] から引用した．Scilab/Xcos に関しては，[41, 32, 33] などを参照せよ．非線型方程式 $f(x) = 0$ のうち，$f(x)$ が多項式のものを**代数方程式** (algebraic equation) という．代数方程式の反復解法に関しては，上記で挙げた [5, 22] のほか，[46, 37] などに記述がある．

第 5 章

反復法の収束性と誤差の解析

5.1 縮小写像と不動点定理

本章では，前章で考察した反復法の収束性について数理的に考察する．反復法の収束性を調べるときに，次の縮小写像の概念が非常に重要となる．

定義 1 (縮小写像) ノルム空間[*1]$(X, \|\cdot\|)$ の閉部分集合 K 上で定義された写像 ϕ が次の条件を満足するとき，ϕ を K における**縮小写像** (contraction mapping) または**非拡大写像** (nonexpansive mapping) と呼ぶ．

1. 任意の $x \in K$ に対して $\phi(x) \in K$.
2. ある $q \in [0,1)$ が存在して，任意の $x, y \in K$ に対して

$$\|\phi(x) - \phi(y)\| \leq q\|x - y\|. \tag{5.1}$$

不等式 (5.1) を Lipschitz **条件** (Lipschitz condition)，定数 q を Lipschitz **定数** (Lipschitz constant) と呼ぶ．

定義 1 の条件 2 は次の条件と等価である．すなわち，ある $q \in [0,1)$ が存在して，

$$\sup_{\substack{x,y\in K \\ x\neq y}} \frac{\|\phi(x) - \phi(y)\|}{\|x - y\|} \leq q.$$

ここで上の不等式の左辺

$$G := \sup_{\substack{x,y\in K \\ x\neq y}} \frac{\|\phi(x) - \phi(y)\|}{\|x - y\|} \tag{5.2}$$

は ϕ をシステムと解釈すると，システム ϕ の**増分ゲイン** (incremental gain) と呼ばれる量となる．これは，システム ϕ への入力の変化 $x - y$ に対して，システム ϕ の出力の変

[*1] ノルム空間に関しては付録を参照のこと．

化量の最悪値を測った量である．すなわち，システム ϕ が入力の変化に対してどれだけ**敏感** (sensitive) であるかを測る量である．この量が 1 未満であるとき，どのように入力が変化しても出力はあまり変化しない．つまり変化に動じることのないシステムであると言える．このようなシステムは入力の変化に対して**ロバスト** (robust) であるという．

練習問題 13 次の 2 つの条件は等価であることを示せ．

1. ある $q \in [0,1)$ が存在して，任意の $x, y \in K$ に対して (5.1) が成り立つ．
2. ある $q \in [0,1)$ が存在して，(5.2) で定義される G に対して $G \leq q$ が成り立つ．

また，ある $q \in [0,\infty)$ （$q < 1$ とは限らない）が存在して，任意の $x, y \in K$ に対して (5.1) が成り立つとき，$\phi(x)$ は K において Lipschitz **連続** (Lipschitz continuous) であるという．

さて，この縮小写像を用いて，反復法の収束性を調べることができる．次の定理は反復法の収束性解析において基本となる定理であり，**不動点定理** (fixed-point theorem) [*2] または**縮小写像の原理** (contraction principle) と呼ばれる．

定理 1 (不動点定理) X を完備なノルム空間 (Banach 空間) とし，K を X の閉部分集合とする．また，写像 ϕ を K における縮小写像とする．このとき，

1. 方程式 $x = \phi(x)$ は K に唯一つの解 x^* を持つ．
2. $x[0] \in K$ を初期値とする反復法

$$x[n+1] = \phi(x[n]), \quad n = 0, 1, 2, \ldots$$

によって生成されるベクトル列 $\{x[n]\}$ に対して

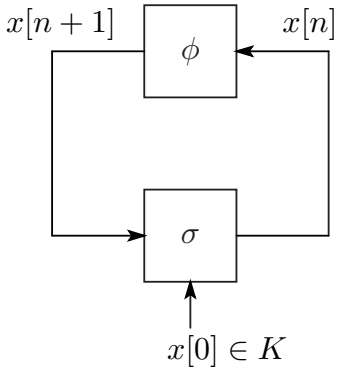
$$\lim_{n\to\infty} x[n] = x^*$$

が成り立つ．

方程式 $x = \phi(x)$ の解 x^* を写像 ϕ の**不動点** (fixed-point) と呼ぶ．点 x^* に写像 ϕ を施しても動かない，すなわち $\phi(x^*) = x^*$ であるのでこの名前が付いている．上の定理は，縮小写像 ϕ による反復法が ϕ の不動点に収束することを言っている．

この定理の証明をする前に，この定理の意味するところを考えてみよう．ϕ が縮小写像であるとは ϕ が入力の変化に対してロバストであることであった．前章で導入した反復法のブロック線図を考える．

[*2] 不動点定理と呼ばれる定理はいくつもあるが，ここで述べる定理は特に Banach **の不動点定理** (Banach's fixed-point theorem) と呼ばれるものである [43].



システム ϕ への入力 $x[n]$ は，時々刻々変化する信号である．しかし ϕ が縮小写像であれば，(5.2) で定義される増分ゲインは 1 未満であり，$x[n]$ の変化は時間が経つにつれだんだんと少なくなっていく ($q < 1$ であるから q^n は n が大きくなるに従って，だんだん小さくなる)．したがって，ある程度時間が経てば，$x[n]$ はほとんど変化しなくなるだろう．すなわち，$x[n]$ は n が十分大きいときには，ある一定値に近づくことになる．この定理は，それを数学的に保証し，さらにこの一定値が方程式 $x = \phi(x)$ の解と一致することを言っているのである．

定理 1 の証明

証明は次の手順で行う．

1. ベクトル列 $\{x[n]\}$ が K 内で収束する．すなわち, 任意の $n \geq 0$ に対して $x[n] \in K$ かつ
$$\lim_{n\to\infty} x[n] =: \alpha \in K.$$
2. 上の収束先 α が $\alpha = \phi(\alpha)$ を満たす．すなわち，$\alpha = x^*$.
3. 解の一意性.

まず，ベクトル列 $\{x[n]\}$ が K 内で収束することを示す．$x[0] \in K$ とする．ϕ は縮小写像だから，定義 1 の条件 1 より，

$$x[1] = \phi(x[0]) \in K$$

が成り立つ．したがって，もういちど定義 1 の条件 1 を使って，

$$x[2] = \phi(x[1]) \in K$$

が得られる．これを繰り返せば，任意の $n \geq 0$ に対して，$x[n] \in K$ であることが言える．また，ノルム空間 X のノルムを $\|\cdot\|$ とすると，任意の $n \geq 1$ に対して，

$$\begin{aligned}
\|x[n+1] - x[n]\| &= \|\phi(x[n]) - \phi(x[n-1])\| \\
&\leq q\|x[n] - x[n-1]\| \qquad (\phi \text{ が縮小写像であることより}) \\
&= q\|\phi(x[n-1]) - \phi(x[n-2])\| \\
&\leq q^2\|x[n-1] - x[n-2]\| \\
&\leq q^3\|x[n-2] - x[n-3]\| \\
&\quad \vdots \\
&\leq q^n\|x[1] - x[0]\|
\end{aligned}$$

が成り立つ．これを用いると，任意の自然数 m, n $(m > n)$ に対して

$$\begin{aligned}
\|x[m] - x[n]\| &= \|x[m] - x[m-1] + x[m-1] - \cdots - x[n+1] + x[n+1] - x[n]\| \\
&\leq \|x[m] - x[m-1]\| + \|x[m-1] - x[m-2]\| + \cdots + \|x[n+1] - x[n]\| \\
&\leq \left(q^{m-1} + q^{m-2} + \cdots + q^n\right)\|x[1] - x[0]\| \\
&= \frac{q^n - q^m}{1-q}\|x[1] - x[0]\| \\
&< \frac{q^n}{1-q}\|x[1] - x[0]\|
\end{aligned}$$

であることがわかる．ここで，任意に $\varepsilon > 0$ をとり，固定する．今，$0 \leq q < 1$ であるので，この ε に対して，N を十分大きくとれば，$n \geq N$ である任意の自然数 n に対して，

$$\frac{q^n}{1-q}\|x[1] - x[0]\| < \varepsilon$$

とすることができる．これより，$m > n \geq N$ なる任意の m, n に対して，

$$\|x[m] - x[n]\| < \varepsilon$$

が成り立つ．これより，ベクトル列 $\{x[n]\}$ はノルム空間 X の Cauchy 列[*3]であることがわかる．X は完備であるので，ベクトル列 $\{x[n]\}$ はあるベクトル $\alpha \in X$ に収束する．さらに，任意の n に対して，$x[n] \in K \subset X$ かつ K は閉集合であるので，

$$\alpha = \lim_{n\to\infty} x[n] \in K \tag{5.3}$$

となることがわかる．

次に，この α が $\alpha = \phi(\alpha)$ を満たすことを示す．上の証明より $\alpha \in K$ であり，また任意の $n \geq 0$ に対して $x[n] \in K$ であることに注意すると，ϕ が縮小写像であることを用い

[*3] 付録参照

て，任意の $n \geq 0$ に対して

$$\|\phi(\alpha) - \phi(x[n])\| \leq q\|\alpha - x[n]\|$$

が成り立つことがわかる．(5.3) より，$n \to \infty$ で $\|\alpha - x[n]\| \to 0$ であるので，上の不等式より，

$$\lim_{n\to\infty} \|\phi(\alpha) - \phi(x[n])\| = 0$$

すなわち

$$\lim_{n\to\infty} \phi(x[n]) = \phi(\alpha)$$

となる[*4]．これより，

$$\phi(\alpha) = \phi\left(\lim_{n\to\infty} x[n]\right) = \lim_{n\to\infty} \phi\left(x[n]\right) = \lim_{n\to\infty} x[n+1] = \alpha$$

が成り立ち，α は方程式 $x = \phi(x)$ の解 x^* であることがわかる．

最後に解の一意性を示す．2 つのベクトル x^*, y^* を方程式 $x = \phi(x)$ の解とする．すなわち，

$$x^* = \phi(x^*), \quad y^* = \phi(y^*)$$

が成り立つとする．このとき，ϕ が縮小写像であることから，

$$\|x^* - y^*\| = \|\phi(x^*) - \phi(y^*)\| \leq q\|x^* - y^*\|$$

が成り立つ．ここで，$0 \leq q < 1$ であるから，上の不等式より

$$\|x^* - y^*\| = 0$$

でなければならない．すなわち，$x^* = y^*$ となる． 【証明終】

例題 3 (縮小写像の例) 次の写像を考える．

$$\phi(x) = ax + b, \quad x \in \mathbb{R}.$$

ただし，a,b は実数とし，さらに $|a| < 1$ とする．このとき，任意の $x, y \in \mathbb{R}$ に対して

$$|\phi(x) - \phi(y)| = |ax - ay| = |a||x - y|$$

であるので，ϕ は Lipschitz 定数 $q = |a| < 1$ を持つ $\mathbb{R}$ 上の縮小写像であることがわかる．

[*4] これは $\phi(x)$ が $x = \alpha$ で連続であることを示している．さらに，ϕ が K 上で連続であることを同様に示すこともできる．すなわち縮小写像は連続写像であることがわかる．

上の例は，多変数の場合に拡張できる．次の写像を考える．

$$\boldsymbol{\phi}(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}, \quad \boldsymbol{x} \in \mathbb{R}^N.$$

ただし，$A \in \mathbb{R}^{N\times N}$, $\boldsymbol{b} \in \mathbb{R}^N$ とし，行列 $A^\top A$ のスペクトル半径[*5]が

$$\rho(A^\top A) < 1$$

を満たすとする．また，ベクトル $\boldsymbol{x} \in \mathbb{R}^N$ のノルムを Euclid ノルム $\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$ とする．このとき，写像 $\boldsymbol{\phi}$ は Lipschitz 定数

$$q = \sqrt{\rho(A^\top A)}$$

を持つ $\mathbb{R}^N$ 上の縮小写像となる．実際，任意の $\boldsymbol{x} \in \mathbb{R}^N$ に対して，$\phi(\boldsymbol{x}) = A\boldsymbol{x} \in \mathbb{R}^N$ であり，また，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して，Lipschitz 条件

$$\begin{aligned}
\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y})\|_2^2 &= \|A\boldsymbol{x} - A\boldsymbol{y}\|_2^2 \\
&= \|A(\boldsymbol{x} - \boldsymbol{y})\|_2^2 \\
&= (\boldsymbol{x} - \boldsymbol{y})^\top A^\top A(\boldsymbol{x} - \boldsymbol{y}) \\
&\le \rho(A^\top A)(\boldsymbol{x} - \boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}) \\
&= q^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2
\end{aligned}$$

が成り立つ．

5.2　微分可能な写像の不動点定理

5.2.1　1 変数の反復法

1 変数の方程式 $f(x) = 0$ の近似解を求める反復法

$$x[n+1] = \phi(x[n]), \quad n = 0, 1, 2, \ldots$$

を考える．関数 ϕ には様々な形が考えられるが，次の反復法は最も単純なものである．

$$x[n+1] = x[n] - f(x[n]), \quad n = 0, 1, 2, \ldots \tag{5.4}$$

このとき，$\phi(x) = x - f(x)$ であり，明らかに $\phi(x) = x$ と $f(x) = 0$ は同値な条件である．したがって，(5.4) は確かに方程式 $f(x) = 0$ に対する反復法である．

まず，(5.4) の収束性について考えよう．このとき，次の補題が重要である．

[*5] 行列のスペクトル半径の定義は付録を参照のこと．

補題 1 (Lipschitz 条件) 関数 ϕ は $\mathbb{R}$ の閉区間 $[a,b]$ 上で C^1 級とする．このとき，ある $q \in [0,1)$ が存在して，任意の $x \in [a,b]$ に対して

$$|\phi'(x)| \leq q \tag{5.5}$$

が成り立つならば，任意の $y, z \in [a,b]$ に対して，Lipschitz 条件

$$|\phi(y) - \phi(z)| \leq q|y - z| \tag{5.6}$$

が成り立つ．

証明

ある $q \in [0,1)$ が存在して，任意の $x \in [a,b]$ に対して (5.5) が成り立つので，

$$\max_{\eta \in [a,b]} |\phi'(\eta)| \leq q < 1 \tag{5.7}$$

が成り立つ．任意の $y, z \in [a,b]$ をとり，固定する．ここで $y \leq z$ と仮定しても一般性は失われない．また，$y = z$ のときは (5.6) は明らかに成り立つので，$y < z$ と仮定する．平均値の定理[*6]と (5.7) より，ある $\xi \in (y,z) \subset [a,b]$ が存在して

$$|\phi(y) - \phi(z)| = |\phi'(\xi)||y - z| \leq \max_{\eta \in [a,b]} |\phi'(\eta)||y - z| \leq q|y - z|$$

が成り立つ．　【証明終】

例題 4 (反復法の収束性) 次の方程式を考える．

$$f(x) = x^3 + x - 1 = 0.$$

この方程式の近似解を得るための反復法を行うために，次の写像を定義する．

$$\phi(x) = \frac{1}{x^2 + 1}.$$

この写像に対し，$f(x) = 0$ と $x = \phi(x)$ は同値であることは容易にわかる．この写像を使い，反復法

$$x[n+1] = \phi(x[n]) = \frac{1}{x[n]^2 + 1}, \quad n = 0, 1, 2, \ldots \tag{5.8}$$

をつくる．このとき，

$$|\phi'(x)| = \frac{2|x|}{(1 + x^2)^2}$$

[*6] 平均値の定理については，付録を参照のこと

となり，この右辺は $x = \pm 1/\sqrt{3}$ のとき最大値 $3\sqrt{3}/8 \approx 0.649519 < 1$ をとる．したがって，任意の $x \in \mathbb{R}$ に対して，

$$|\phi_1'(x)| \leq \frac{3\sqrt{3}}{8} < 1$$

であることがわかる．ゆえに反復法 (5.8) は，任意の初期値 $x[0]$ に対して，$x = \phi(x)$ の解，すなわち $f(x) = 0$ の解に収束する．

練習問題 14 次の写像 ϕ は領域 K における縮小写像であるかどうか調べ，縮小写像であれば Lipschitz 定数を求めよ．

1. $\phi(x) = 2x, \quad K = (-\infty, \infty)$
2. $\phi(x) = \frac{1}{2}x, \quad K = (-\infty, \infty)$
3. $\phi(x) = 2x^2, \quad K = \left[-\frac{1}{5}, \frac{1}{5}\right]$
4. $\phi(x) = \cos x, \quad K = \left[0, \frac{\pi}{3}\right]$
5. $\phi(x) = \sin x, \quad K = \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$

補題 1 を用いて，反復法 (5.4) が収束するための十分条件を得ることができる．

定理 2 (微分可能な写像の不動点定理) 実関数 $f(x)$ は，閉区間 $[a, b]$ 上で C^1 級とし，また任意の $x \in [a, b]$ に対して

$$a \leq x - f(x) \leq b \tag{5.9}$$

$$0 < f'(x) < 2 \tag{5.10}$$

が成り立つとする．このとき，$f(x) = 0$ の解 x^* が $[a, b]$ 内に唯一存在し，$x[0] \in [a, b]$ を初期値とする反復法 (5.4) は x^* に収束する．

証明

まず，$x \in [a, b]$ に対して，

$$\phi(x) = x - f(x)$$

と定義する．この両辺を x で微分すると，

$$\phi'(x) = 1 - f'(x), \quad x \in [a, b]$$

が得られる．これと (5.10) より，任意の $x \in [a, b]$ に対して，

$$-1 < \phi'(x) < 1.$$

すなわち，

$$0 \leq |\phi'(x)| < 1 \tag{5.11}$$

であることがわかる．関数 $f(x)$ は閉区間 $[a,b]$ 上の C^1 級関数であるので，$\phi'(x)$ は $[a,b]$ 上で連続となり，したがって $|\phi'(x)|$ はこの区間内で最大値を取る．一方，任意の $x \in [a,b]$ に対して (5.11) が成り立つので，

$$\max_{a \le x \le b} |\phi'(x)| < 1$$

であることがわかる．これより，$q = \max_{a \le x \le b} |\phi'(x)|$ とおくと $q \in [0,1)$ であり，任意の $x \in [a,b]$ に対して，$|\phi'(x)| \le q$ が成り立つ．ゆえに，補題 1 より，任意の $y, z \in [a,b]$ に対して，Lipschitz 条件

$$|\phi(y) - \phi(z)| \le q|y - z|$$

が成り立つ．また，(5.9) より，任意の $x \in [a,b]$ に対して $\phi(x) \in [a,b]$ であることがわかる．したがって，ϕ は縮小写像であり，不動点定理（定理 1）より $f(x) = 0$ の解 x^* は区間 $[a,b]$ に唯一つ存在し，反復法 (5.4) により生成される数列 $\{x[n]\}$ は真の解 x^* に収束することがわかる．　【証明終】

5.2.2　多変数の反復法

1 変数関数に対する補題 1 は写像 ϕ の定義域の条件を少し変更すると，多変数の場合でも成り立つ．多変数の場合というのは，連立方程式のことである．次のような N 個の変数 $x_1, x_2, \ldots, x_N$ に関する N 本の連立方程式を考えよう．

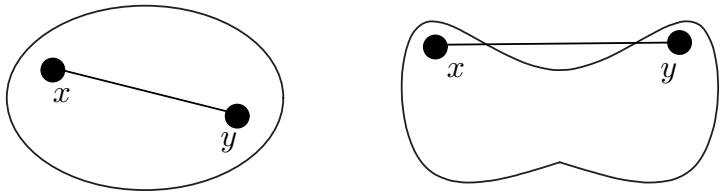
$$\begin{cases} f_1(x_1, x_2, \ldots, x_N) &= 0, \\ f_2(x_1, x_2, \ldots, x_N) &= 0, \\ & \vdots \\ f_N(x_1, x_2, \ldots, x_N) &= 0. \end{cases}$$

この連立方程式に対して，次のような等価な変形を行う．

$$\begin{cases} x_1 &= \phi_1(x_1, x_2, \ldots, x_N), \\ x_2 &= \phi_2(x_1, x_2, \ldots, x_N), \\ & \vdots \\ x_N &= \phi_N(x_1, x_2, \ldots, x_N). \end{cases}$$

ここで，次のような**ベクトル表現** (vector representation) を使って変数 $x_1, \ldots, x_N$ と関数 $\phi_1, \ldots, \phi_N$ をまとめる．

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N, \quad \boldsymbol{\phi}(\boldsymbol{x}) = \begin{bmatrix} \phi_1(x_1, x_2, \ldots, x_N) \\ \phi_2(x_1, x_2, \ldots, x_N) \\ \vdots \\ \phi_N(x_1, x_2, \ldots, x_N) \end{bmatrix} \in \mathbb{R}^N.$$



(a)　　(b)

図 5.1　(a) 凸集合 (b) 凸集合でない

すると，方程式は次のように簡単に書ける．

$$\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{x})$$

この方程式に対して，次の反復法を考える．

$$\boldsymbol{x}[n+1] = \boldsymbol{\phi}(\boldsymbol{x}[n]), \quad \boldsymbol{x}[n] \in \mathbb{R}^N, \quad n = 0, 1, 2, \ldots \tag{5.12}$$

縮小写像の原理は，$\mathbb{R}^N$ でも成り立つので，もし $\boldsymbol{\phi}$ が $\mathbb{R}^N$ の閉部分集合 K 上の縮小写像であれば，上の反復法は $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{x})$ の解 $\boldsymbol{x}^*$ に収束する．また，もし $\boldsymbol{\phi}$ が K 上で微分可能であれば，前節で考えたような収束条件も得ることができる．写像 $\boldsymbol{\phi}$ が微分可能であるときの反復法の収束条件を考えるために，まず凸集合を定義する．

定義 2 (凸集合) ベクトル空間 $\mathbb{R}^N$ の部分集合 K が**凸集合** (convex set) であるとは，任意の $\boldsymbol{x}, \boldsymbol{y} \in K$ と任意の $t \in [0, 1]$ に対して，

$$t\boldsymbol{x} + (1-t)\boldsymbol{y} \in K$$

が成り立つことである．特に，$\mathbb{R}^n$ 全体や空集合，$\mathbb{R}^n$ の線形部分空間，$\mathbb{R}^n$ の単位球などは凸集合である．

図 5.1 に凸集合と凸でない集合の例を示す．この図に示してあるように，凸集合とは集合内の任意の 2 点を結ぶ線分がその集合に含まれるような集合である．

以上の準備のもと，以下の補題が成り立つ．

補題 2 (Lipschitz 条件) ベクトル空間 $\mathbb{R}^N$ の閉凸部分集合[*7]K 上で定義され，K で C^1 級[*8]の写像 $\boldsymbol{\phi} : K \to \mathbb{R}^N$ について，ある $q \in [0, 1)$ が存在して，任意の $\boldsymbol{x} \in K$ に対して

$$\|D_{\boldsymbol{\phi}}(\boldsymbol{x})\| \leq q < 1 \tag{5.13}$$

[*7] $\mathbb{R}^N$ の閉凸部分集合とは，閉集合かつ凸集合であるような $\mathbb{R}^N$ の部分集合のことである．

[*8] ベクトル値の多変数関数 $\boldsymbol{\phi}$ が C^1 級であるとは，$\boldsymbol{\phi}$ の各要素 ϕ_i が C^1 級であることである．

が成り立つとする．ただし，$D_{\boldsymbol{\phi}}$ は写像 $\boldsymbol{\phi}$ の Jacobi **行列** (Jacobian matrix) であり，次式で定義される．

$$D_{\boldsymbol{\phi}}(\boldsymbol{x}) := \begin{bmatrix} \dfrac{\partial \phi_1}{\partial x_1} & \dfrac{\partial \phi_1}{\partial x_2} & \cdots & \dfrac{\partial \phi_1}{\partial x_N} \\ \dfrac{\partial \phi_2}{\partial x_1} & \dfrac{\partial \phi_2}{\partial x_2} & \cdots & \dfrac{\partial \phi_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial \phi_N}{\partial x_1} & \dfrac{\partial \phi_N}{\partial x_2} & \cdots & \dfrac{\partial \phi_N}{\partial x_N} \end{bmatrix}.$$

また，$\|D_{\boldsymbol{\phi}}(\boldsymbol{x})\|$ は $\mathbb{R}^N$ のベクトルのノルム $\|\cdot\|$ から誘導された行列ノルムとする[*9]．このとき，任意の $\boldsymbol{y}, \boldsymbol{z} \in K$ に対して，Lipschitz 条件

$$\|\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{\phi}(\boldsymbol{z})\| \leq q\|\boldsymbol{y} - \boldsymbol{z}\| \tag{5.14}$$

が成り立つ．

証明

任意の $\boldsymbol{y}, \boldsymbol{z} \in K$ をとり固定する．K は凸集合であるから，任意の $t \in [0,1]$ に対して，$\boldsymbol{z} + t(\boldsymbol{y} - \boldsymbol{z}) \in K$ が成り立つ．これを用いれば

$$\begin{aligned} \boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{\phi}(\boldsymbol{z}) &= \boldsymbol{\phi}(\boldsymbol{z} + t(\boldsymbol{y} - \boldsymbol{z}))|_{t=1} - \boldsymbol{\phi}(\boldsymbol{z} + t(\boldsymbol{y} - \boldsymbol{z}))|_{t=0} \\ &= \int_0^1 \frac{d}{dt}\boldsymbol{\phi}(\boldsymbol{z} + t(\boldsymbol{y} - \boldsymbol{z}))dt \end{aligned}$$

と書ける．ここで

$$\frac{d}{dt}\boldsymbol{\phi}(\boldsymbol{z} + t(\boldsymbol{y} - \boldsymbol{z})) = D_{\boldsymbol{\phi}}(\boldsymbol{z} + t(\boldsymbol{y} - \boldsymbol{z}))(\boldsymbol{y} - \boldsymbol{z})$$

が成り立つので，

$$\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{\phi}(\boldsymbol{z}) = \int_0^1 D_{\boldsymbol{\phi}}(\boldsymbol{z} + t(\boldsymbol{y} - \boldsymbol{z}))(\boldsymbol{y} - \boldsymbol{z})dt$$

となる．また，ある実数 $q \in [0,1)$ が存在して，任意の $\boldsymbol{x} \in K$ に対して，

$$\|D_{\boldsymbol{\phi}}(\boldsymbol{x})\| \leq q$$

[*9] 行列のノルムに関しては，付録を参照のこと．

が成り立つとすると，

$$\begin{aligned}\|\boldsymbol{\phi}(\boldsymbol{y})-\boldsymbol{\phi}(\boldsymbol{z})\| &= \left\|\int_0^1 D_{\boldsymbol{\phi}}(\boldsymbol{z}+t(\boldsymbol{y}-\boldsymbol{z}))(\boldsymbol{y}-\boldsymbol{z})dt\right\| \\ &\leq \int_0^1 \|D_{\boldsymbol{\phi}}(\boldsymbol{z}+t(\boldsymbol{y}-\boldsymbol{z}))\|\|(\boldsymbol{y}-\boldsymbol{z})\|dt \\ &\leq q\|\boldsymbol{y}-\boldsymbol{z}\|\int_0^1 dt \\ &= q\|\boldsymbol{y}-\boldsymbol{z}\|\end{aligned}$$

が得られる．　【証明終】

この補題の証明からわかるように，領域 K が凸集合であることは本質的である．

補題 2 を使えば，多変数の連立方程式に対する反復法の収束性を調べることが可能となる．

定理 3 (微分可能な写像の不動点定理) ベクトル空間 $\mathbb{R}^N$ の閉凸部分集合 K 上で定義され，K で C^1 級の写像 $\boldsymbol{\phi}: K \to \mathbb{R}^N$ が次の条件を満たすとする．

1. 任意の $\boldsymbol{x} \in K$ に対して $\boldsymbol{\phi}(\boldsymbol{x}) \in K$.
2. ある $q \in [0,1)$ が存在して，任意の $\boldsymbol{x} \in K$ に対して，

$$\|D_{\boldsymbol{\phi}}(\boldsymbol{x})\| \leq q < 1.$$

このとき $\boldsymbol{\phi}$ は K に唯一つの不動点 $\boldsymbol{x}^*$ を持ち，$\boldsymbol{x}[0] \in K$ を初期値とする反復法

$$\boldsymbol{x}[n+1] = \boldsymbol{\phi}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots$$

によって生成されるベクトル列 $\{\boldsymbol{x}[n]\}$ に対して

$$\lim_{n\to\infty} \boldsymbol{x}[n] = \boldsymbol{x}^*$$

が成り立つ．

練習問題 15 定理 2 の証明を参考にして，定理 3 を証明せよ．

例題 5 (連立一次方程式と Neumann 級数) 例題 3 で考えた問題をもう一度考える．$A \in \mathbb{R}^{N\times N}$, $\boldsymbol{b} \in \mathbb{R}^N$ とし，

$$\boldsymbol{\phi}(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}, \quad \boldsymbol{x} \in \mathbb{R}^N$$

とする．すなわち，方程式 $(I-A)\boldsymbol{x} = \boldsymbol{b}$ を考える．まず明らかに，任意の $\boldsymbol{x} \in \mathbb{R}^N$ に対して，$\boldsymbol{\phi}(\boldsymbol{x}) \in \mathbb{R}^N$ が成り立つ．次に写像 $\boldsymbol{\phi}(\boldsymbol{x}) = A\boldsymbol{x}$ の Jacobi 行列は

$$D_{\boldsymbol{\phi}}(\boldsymbol{x}) = A, \quad \boldsymbol{x} \in \mathbb{R}^N$$

で与えられる．もし行列 A が $\|A\| < 1$ という性質を持てば，定理 3 より，方程式 $(I-A)\boldsymbol{x} = \boldsymbol{b}$ は $\mathbb{R}^N$ に唯一つの解を持つ．すなわち，行列 $I-A$ は**正則行列** (non-singular matrix) であることがわかる．また，反復法

$$\boldsymbol{x}[n+1] = A\boldsymbol{x}[n] + \boldsymbol{b}, \quad n = 0, 1, 2, \ldots$$

は任意の初期値 $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，解 $\boldsymbol{x}^* = (I-A)^{-1}\boldsymbol{b}$ に収束することもわかる．具体的に $x[n]$ を求めてみると以下のようになる．

$$\begin{aligned}
\boldsymbol{x}[1] &= A\boldsymbol{x}[0] + \boldsymbol{b} \\
\boldsymbol{x}[2] &= A\boldsymbol{x}[1] + \boldsymbol{b} = A^2\boldsymbol{x}[0] + \boldsymbol{b} + A\boldsymbol{b} \\
\boldsymbol{x}[3] &= A\boldsymbol{x}[2] + \boldsymbol{b} = A^3\boldsymbol{x}[0] + \boldsymbol{b} + A\boldsymbol{b} + A\boldsymbol{b} \\
&\cdots \\
\boldsymbol{x}[n] &= A^n\boldsymbol{x}[0] + \boldsymbol{b} + A\boldsymbol{b} + A^2\boldsymbol{b} + \cdots + A^{n-1}\boldsymbol{b}
\end{aligned}$$

ここで，$\|A\| < 1$ より，どのような $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対しても

$$\|A^n\boldsymbol{x}[0]\| \leq \|A\|^n \|x[0]\| \to 0, \quad n \to \infty$$

が言える．ゆえに

$$\lim_{n\to\infty} \boldsymbol{x}[n] = \left(I + A + A^2 + \cdots\right)\boldsymbol{b} = \left(\sum_{n=0}^{\infty} A^n\right)\boldsymbol{b}.$$

定理 3 より上の極限ベクトルは $\boldsymbol{x}^* = (I-A)^{-1}\boldsymbol{b}$ に等しく，また，上の等式はどのようなベクトル $\boldsymbol{b} \in \mathbb{R}^N$ に対しても成り立つので，逆行列に関する公式

$$(I-A)^{-1} = \sum_{n=0}^{\infty} A^n$$

が成り立つことがわかる．これを Neumann **級数** (Neumann series) と呼ぶ．

例題 3 では $\mathbb{R}^n$ のノルムとして Euclid ノルム $\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$ を使った．このとき，

$$\|A\| = \sqrt{\rho(A^\top A)}$$

が成り立つ．すなわち，上の例は例題 3 を Euclid ノルムとは限らない場合に一般化したものである．

練習問題 16 行列 $A \in \mathbb{R}^{N\times N}$ とベクトル $\boldsymbol{b} \in \mathbb{R}^N$ で定義された写像 $\boldsymbol{\phi}(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$ に対して，その Jacobi 行列が $D_{\boldsymbol{\phi}}(\boldsymbol{x}) = A$ となることを示せ．

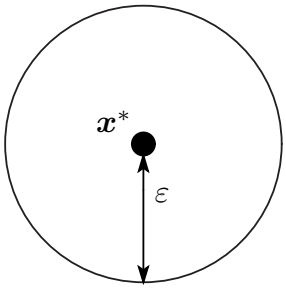


図 5.2　中心 $\boldsymbol{x}^* \in \mathbb{R}^n$，半径 $\varepsilon > 0$ の閉球 $B(\boldsymbol{x}^*, \varepsilon)$.

5.3　不動点近傍での収束条件

これまで述べた収束条件では，任意の $\boldsymbol{x} \in K$ に対して

$$\boldsymbol{\phi}(\boldsymbol{x}) \in K$$

が成り立つような写像 $\boldsymbol{\phi}$ と領域 K を選ぶ必要があった．ここでは，この条件のかわりに，領域 K に $\boldsymbol{\phi}$ の不動点 $\boldsymbol{x}^*$ が含まれていることがあらかじめわかっている場合の収束条件を考える．すなわち，不動点 $\boldsymbol{x}^*$ の近傍での収束条件を考えることになる．例えば，前章で考察した 2 分法のように，方程式 $f(x) = 0$ に対して $f(a) < 0$, $f(b) > 0$ となるような実数 a と b が見つかれば，閉区間 $[a, b]$ に $f(x) = 0$ の解が含まれることがわかる．したがって，この方程式を等価的に変形した $x = \phi(x)$ の不動点は $K = [a, b]$ に含まれる．本節では，このような場合の反復法を考える．

定理 4 (不動点近傍での収束条件 1) 写像 $\boldsymbol{\phi} : \mathbb{R}^N \to \mathbb{R}^N$ の不動点を $\boldsymbol{x}^*$ とし，$\mathbb{R}^N$ の領域 K を $\boldsymbol{x}^*$ を中心とする半径 $\varepsilon > 0$ の閉球，すなわち

$$K = B(\boldsymbol{x}^*, \varepsilon) = \{\boldsymbol{x} \in \mathbb{R}^N : \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \varepsilon\} \subset \mathbb{R}^N$$

とする（図 5.2 を参照）．また，ある $q \in [0, 1)$ が存在して，任意の $\boldsymbol{y}, \boldsymbol{z} \in K$ に対して，

$$\|\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{\phi}(\boldsymbol{z})\| \leq q\|\boldsymbol{y} - \boldsymbol{z}\| \tag{5.15}$$

が成り立つと仮定する．このとき，$\boldsymbol{x}^*$ は K 内の唯一つの不動点であり，$\boldsymbol{x}[0] \in K$ を初期値とする反復法

$$\boldsymbol{x}[n+1] = \boldsymbol{\phi}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots$$

によって生成されるベクトル列 $\{\boldsymbol{x}[n]\}$ に対して

$$\lim_{n \to \infty} \boldsymbol{x}[n] = \boldsymbol{x}^*$$

が成り立つ．

証明

任意の $\boldsymbol{x} \in K$ に対して $\boldsymbol{\phi}(\boldsymbol{x}) \in K$ を示せばよい．実際，$\boldsymbol{x}^* = \boldsymbol{\phi}(\boldsymbol{x}^*)$ と Lipschitz 条件 (5.15) を用いれば，

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{x}^*\| = \|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{x}^*)\| \leq q\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq q\varepsilon < \varepsilon$$

となり，ゆえに $\boldsymbol{\phi}(\boldsymbol{x}) \in K$ であることがわかる．したがって不動点定理（定理 1）より，定理 4 が成り立つことがわかる．【証明終】

写像 $\boldsymbol{\phi}$ が微分可能である場合は，定理 2 を多変数に拡張した次の定理により，反復法の収束性を調べることができる．

定理 5 (不動点近傍での収束条件 2) 写像 $\boldsymbol{\phi} : \mathbb{R}^N \to \mathbb{R}^N$ の不動点を $\boldsymbol{x}^*$ とし，$K = B(\boldsymbol{x}^*, \varepsilon), \varepsilon > 0$ とする．さらに $\boldsymbol{\phi}$ は次の条件を満たすとする．

1. $\boldsymbol{\phi}$ は K 上で C^1 級
2. ある $q \in [0, 1)$ が存在して，任意の $\boldsymbol{x} \in K$ に対して，

$$\|D_{\boldsymbol{\phi}}(\boldsymbol{x})\| \leq q.$$

このとき，$\boldsymbol{x}^*$ は K 内の唯一つの不動点であり，$\boldsymbol{x}[0] \in K$ を初期値とする反復法

$$\boldsymbol{x}[n+1] = \boldsymbol{\phi}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots$$

によって生成されるベクトル列 $\{\boldsymbol{x}[n]\}$ に対して

$$\lim_{n\to\infty} \boldsymbol{x}[n] = \boldsymbol{x}^*$$

が成り立つ．

証明

ある $q \in [0, 1)$ が存在して，条件 $\|D_{\boldsymbol{\phi}}(\boldsymbol{x})\| \leq q$ が成り立つと仮定する．このとき，補題 2 より，任意の $\boldsymbol{x} \in K$ に対して，

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{x}^*\| = \|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{x}^*)\| \leq q\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq q\varepsilon < \varepsilon$$

が成り立つ．すなわち，任意の $x \in K$ に対して，$\boldsymbol{\phi}(\boldsymbol{x}) \in K$ であることがわかる．したがって，定理 3 より，定理 5 が成り立つことがわかる．【証明終】

練習問題 17 次の反復法が収束するかどうかを調べよ．収束する場合，その収束先はどのような方程式の解か？

1. $x[n+1] = \frac{1}{2}x[n] + 1, \quad n = 0, 1, 2, \ldots, \quad x[0] = 1.$
2. $x[n+1] = \cos x[n], \quad n = 0, 1, 2, \ldots, \quad x[0] = 1.$
3. $$\begin{cases} x[n+1] = \dfrac{1}{2}\left(x[n] + y[n] + 1\right), \\ y[n+1] = \dfrac{1}{2}\left(y[n] + 1\right), \qquad n = 0, 1, 2, \ldots, \quad x[0] = 1, \quad y[0] = 1. \end{cases}$$

5.4　Newton 法の収束性

ここでは，4.2.4 節で導入した 1 変数の Newton 法の収束条件と収束の速さについて調べる．また多変数の場合の Newton 法についても述べ，その収束性を調べる．

5.4.1　1 変数 Newton 法

1 変数の非線形方程式 $f(x) = 0$ に対する Newton 法

$$x[n+1] = x[n] - \frac{f(x[n])}{f'(x[n])}, \quad n = 0, 1, 2, \ldots \tag{5.16}$$

を考える．ここで，方程式 $f(x) = 0$ の解を x^* とし，x^* を含む閉区間 K 上で f は C^2 級かつ任意の $x \in K$ に対して $f'(x) \neq 0$ と仮定する．反復法 (5.16) より，x^* は次の写像 ϕ の不動点であることがわかる．

$$\phi(x) = x - \frac{f(x)}{f'(x)}, \quad x \in K. \tag{5.17}$$

関数 f に対する仮定より，この写像 ϕ を用いた方程式 $x = \phi(x)$ と $f(x) = 0$ は K 上で同値である．本節では，この Newton 法の収束条件を導く．まずは簡単な例題で Newton 法の収束性を調べてみよう．

例題 6 ($\sqrt{a}$ を求める Newton 法) $a > 0$ として，$\sqrt{a}$ を求める数値計算を考える．すなわち，方程式

$$f(x) = x^2 - a = 0$$

を解くための数値計算を行う．Newton 法を使うと，反復法は

$$\begin{aligned} x[n+1] &= x[n] - \frac{f(x[n])}{f'(x[n])} \\ &= x[n] - \frac{x[n]^2 - a}{2x[n]} \\ &= \frac{1}{2}x[n] + \frac{a}{2x[n]} \end{aligned} \tag{5.18}$$

と表され,

$$\phi(x) = \frac{1}{2}x + \frac{a}{2x}$$

であることがわかる．またこのとき,

$$\phi'(x) = \frac{1}{2} - \frac{a}{2x^2}$$

となる．ここで， $x > \sqrt{a}$ とすると,

$$0 < \phi'(x) < \frac{1}{2}$$

また，$\sqrt{\frac{a}{2}} \leq x \leq \sqrt{a}$ とすると,

$$-\frac{1}{2} \leq -\phi'(x) \leq 0$$

以上より，任意の $x \in K := \left[\sqrt{\frac{a}{2}}, \infty\right)$ に対して,

$$|\phi'(x)| \leq \frac{1}{2} < 1$$

が成り立つ．さらに任意の $x \in K$ に対して,

$$\phi(x) = \frac{1}{2}\left(x + \frac{a}{x}\right) \geq \sqrt{x \cdot \frac{a}{x}} = \sqrt{a} > \sqrt{\frac{a}{2}}$$

が成り立つ．ただし途中で**相加平均と相乗平均の不等式** (inequality of arithmetic and geometric means) を用いた．ここで，相加平均と相乗平均の不等式とは，任意の実数 $a > 0, b > 0$ に対して，不等式

$$\frac{a+b}{2} \geq \sqrt{ab}$$

が成り立つことである．等号は $a = b$ のときに成り立つ．上の式より，任意の $x \in K$ に対して $\phi(x) \in K$ であることがわかる．すなわち $\phi(x)$ は K 上の縮小写像である．したがって，不動点定理（定理 1）より，初期値を $\sqrt{\frac{a}{2}}$ 以上，例えば $x[0] = a$ とすれば，Newton 法 (5.18) によって生成される数列 $\{x[n]\}$ は $\sqrt{a}$ に収束する．

練習問題 18 1. 例題 6 では，初期値を $x[0] \geq \sqrt{\frac{a}{2}}$ にとると反復法 (5.18) は $\sqrt{a}$ に収束することを示した．しかし，実際には，(5.18) の Newton 法は，初期値を $x[0] > 0$ にとると $\sqrt{a}$ に，$x[0] < 0$ にとると $-\sqrt{a}$ に収束する．これを証明せよ．

2. 次の方程式の近似解を求めるための Newton 法を導出せよ．
   (a) $3x^3 - 2x - 5 = 0$
   (b) $x = 2\sin x$
   (c) $e^{-x} = \sin x$

（d）$x = \frac{1}{2} + \sin x$

3. 与えられた $a > 0$ と自然数 $n \geq 2$ に対して，$\sqrt[n]{a}$ の近似値を求めるための Newton 法を導出し，その Newton 法が収束するための初期値の範囲を求めよ．

一般の方程式 $f(x) = 0$ に対する Newton 法の収束条件に関して以下の定理が成り立つ．

定理 6 (Newton 法の収束条件 1) 閉集合 K を方程式 $f(x) = 0$ の解 x^* を含む閉区間 $[x^* - \varepsilon, x^* + \varepsilon]$ $(\varepsilon > 0)$ とし，K 上で f は C^2 級かつ $f' \neq 0$ と仮定する．さらに，ある $q \in [0, 1)$ が存在して，任意の $x \in K$ に対して

$$\left| \frac{f''(x)f(x)}{f'(x)^2} \right| \leq q \tag{5.19}$$

が成り立つとする．このとき，任意の初期値 $x[0] \in K$ から出発する Newton 反復法 (5.16) は方程式 $f(x) = 0$ の解 x^* に収束する．

証明

(5.17) の両辺を x で微分すると，

$$\phi'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f''(x)f(x)}{f'(x)^2}$$

となる．関数 f は K 上で C^2 級であり，また K 上で $f'(x) \neq 0$ であるので，$\phi(x)$ は K 上で C^1 級となることがわかる．また (5.19) より任意の $x \in K$ に対して $|\phi'(x)| \leq q$ となることもわかる．したがって，定理 5 より，任意の初期値 $x[0] \in K$ から出発する Newton 反復法 (5.16) は方程式 $f(x) = 0$ の解 x^* に収束する．　【証明終】

上の定理において，$\varepsilon > 0$ を十分小さくとれば，上の定理の条件 (5.19) を満たす閉区間 $K = [x^* - \varepsilon, x^* + \varepsilon]$ は必ず存在する．実際，$f(x^*) = 0$，$f'(x^*) \neq 0$ より，$\phi'(x^*) = 0$ であり，また上の証明より ϕ' は K 上で連続であるので，$\varepsilon > 0$ を十分小さくとれば，不等式 (5.19) を満たす $q \in [0, 1)$ は必ず存在する．すなわち，初期値 $x[0]$ を x^* の十分近くにとれば，Newton 法は $f(x) = 0$ の解に収束することがわかる．

5.4.2　Newton 法の収束の速さについて

定理 6 の条件が成り立つような $q \in [0, 1)$ と閉区間 K が与えられているとする．このとき，任意の $x[0] \in K$ から出発する Newton 法 (5.16) に対して，

$$x[n] \in K, \quad n = 0, 1, 2, \ldots$$

が成り立ち，数列 $\{x[n]\}$ は $f(x)=0$ の厳密解 x^* に収束することが定理 6 よりわかる．ここでは，Newton 法が収束すると仮定して，その収束の速さを考察する．

Newton 法における第 n ステップでの近似解 $x[n]$ を考える．一般性を失わず，$x[n]<x^*$ と仮定する．Taylor の定理[*10] より，ある $\xi\in(x[n],x^*)$ が存在して，

$$f(x^*)=f(x[n])+f'(x[n])(x^*-x[n])+\frac{1}{2}f''(\xi)(x^*-x[n])^2 \tag{5.20}$$

が成り立つ．さらに x^* は $f(x)=0$ の解であるので，$f(x^*)=0$．これを (5.20) に代入して整理すると

$$f(x[n])=-f'(x[n])(x^*-x[n])-\frac{1}{2}f''(\xi)(x^*-x[n])^2$$

が得られる．Newton 反復法の式 (5.16) に上の等式を代入すると，

$$\begin{aligned}x[n+1]&=x[n]-\frac{-f'(x[n])(x^*-x[n])-\frac{1}{2}f''(\xi)(x^*-x[n])^2}{f'(x[n])}\\&=x^*+\frac{f''(\xi)}{2f'(x[n])}(x^*-x[n])^2\end{aligned}$$

となり，これより

$$|x[n+1]-x^*|=\left|\frac{f''(\xi)}{2f'(x[n])}\right|\cdot|x[n]-x^*|^2$$

が得られる．ここで，K は閉区間であり，f は K 上で C^2 級かつ任意の $x\in K$ に対して $f'(x)\neq 0$ であるので，

$$c:=\max_{x,y\in K}\left|\frac{f''(y)}{2f'(x)}\right|<\infty$$

である．ゆえに任意の $n=0,1,2,\ldots$ に対して，

$$|x[n+1]-x^*|\leq c|x[n]-x^*|^2 \tag{5.21}$$

が成り立つ．すなわち，Newton 法は 2 **次収束** (quadratic convergence) であることがわかる．初期値 $x[0]$ が厳密解 x^* に十分近ければ，各ステップで厳密解との誤差は 2 乗に比例して小さくなり，Newton 法の収束は非常に速い．Newton 法がよく用いられるのは，これが主な理由である．

5.4.3 多変数 Newton 法

多変数の連立方程式の近似解を求める際にも，Newton 法は有効である．N 変数の連立方程式

$$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{0}$$

[*10] Taylor の定理については付録を参照のこと．

を考える．ただし，

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N, \quad \boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_N) \\ f_2(x_1, x_2, \ldots, x_N) \\ \vdots \\ f_N(x_1, x_2, \ldots, x_N) \end{bmatrix} \in \mathbb{R}^N$$

とする．この連立方程式に対する Newton 法を**多変数 Newton 法** (multivariate Newton's method) と呼び，次式で与えられる．

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1} \boldsymbol{f}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots \tag{5.22}$$

ただし，$D_{\boldsymbol{f}}(\boldsymbol{x})$ は $\boldsymbol{f}(\boldsymbol{x})$ の Jacobi 行列，すなわち

$$D_{\boldsymbol{f}}(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_N}{\partial x_1} & \cdots & \dfrac{\partial f_N}{\partial x_N} \end{bmatrix}$$

であり，また

$$\boldsymbol{x}[n] = \begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_N[n] \end{bmatrix} \in \mathbb{R}^N, \quad n = 0, 1, 2, \ldots$$

である．

まず，多変数 Newton 法 (5.22) の収束性について調べよう．方程式 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ の解を $\boldsymbol{x}^*$ とし，領域 K を 中心 $\boldsymbol{x}^*$，半径 $\varepsilon > 0$ の閉球 $K = B(\boldsymbol{x}^*, \varepsilon)$ とする．また，関数 $\boldsymbol{f}$ は K 上で C^2 級，かつ任意の $\boldsymbol{x} \in K$ に対して，行列 $D_{\boldsymbol{f}}(\boldsymbol{x})$ は正則，すなわち逆行列が存在すると仮定する．写像 $\boldsymbol{\phi}(\boldsymbol{x})$ を次で定義する．

$$\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{x} - D_{\boldsymbol{f}}(\boldsymbol{x})^{-1} \boldsymbol{f}(\boldsymbol{x}).$$

このとき，$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ と $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{x})$ は K 上で同値であり，Newton 法 (5.22) は

$$\boldsymbol{x}[n+1] = \boldsymbol{\phi}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots$$

と書くことができる．この反復法の収束性を調べるために $\boldsymbol{\phi}(\boldsymbol{x})$ の微分，すなわち Jacobi 行列 $D_{\boldsymbol{\phi}}(\boldsymbol{x})$ を求める．多変数関数 $\boldsymbol{\phi}(\boldsymbol{x})$ を x_i $(i = 1, 2, \ldots, N)$ で偏微分すると，

$$\begin{aligned} \frac{\partial \boldsymbol{\phi}(\boldsymbol{x})}{\partial x_i} &= \boldsymbol{e}_i - \frac{\partial}{\partial x_i} \left\{ D_{\boldsymbol{f}}(\boldsymbol{x})^{-1} \boldsymbol{f}(\boldsymbol{x}) \right\} \\ &= \boldsymbol{e}_i - \frac{\partial}{\partial x_i} \left\{ D_{\boldsymbol{f}}(\boldsymbol{x})^{-1} \right\} \boldsymbol{f}(\boldsymbol{x}) - D_{\boldsymbol{f}}(\boldsymbol{x})^{-1} \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_i} \end{aligned}$$

が得られる．ただし，$\boldsymbol{e}_i$ は第 i 要素が 1，その他の要素が 0 の単位縦ベクトルである．逆行列の微分の公式，

$$\frac{\partial}{\partial x_i}\left\{M(\boldsymbol{x})^{-1}\right\} = -M(\boldsymbol{x})^{-1}\frac{\partial M(\boldsymbol{x})}{\partial x_i}M(\boldsymbol{x})^{-1} \tag{5.23}$$

を使えば，

$$\begin{aligned}\frac{\partial \boldsymbol{\phi}(\boldsymbol{x})}{\partial x_i} &= \boldsymbol{e}_i + D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\frac{\partial D_{\boldsymbol{f}}(\boldsymbol{x})}{\partial x_i}D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\boldsymbol{f}(\boldsymbol{x}) - D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_i} \\ &= \boldsymbol{e}_i + \boldsymbol{v}_i(\boldsymbol{x}) - D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_i}\end{aligned}$$

となる．ただし，$\boldsymbol{v}_i(\boldsymbol{x})$ は $\mathbb{R}^N$ のベクトル値関数であり，

$$\boldsymbol{v}_i(\boldsymbol{x}) := D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\frac{\partial D_{\boldsymbol{f}}(\boldsymbol{x})}{\partial x_i}D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\boldsymbol{f}(\boldsymbol{x}), \quad i = 1, 2, \ldots, N \tag{5.24}$$

と定義する．これらのベクトル $\boldsymbol{v}_1(\boldsymbol{x}), \ldots, \boldsymbol{v}_N(\boldsymbol{x})$ をまとめて

$$V(\boldsymbol{x}) := \left[\begin{array}{cccc} \boldsymbol{v}_1(\boldsymbol{x}) & \boldsymbol{v}_2(\boldsymbol{x}) & \ldots & \boldsymbol{v}_N(\boldsymbol{x}) \end{array}\right]$$

とおくと，

$$\begin{aligned}D_{\boldsymbol{\phi}}(\boldsymbol{x}) &= \left[\begin{array}{cccc} \dfrac{\partial \boldsymbol{\phi}(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial \boldsymbol{\phi}(\boldsymbol{x})}{\partial x_2} & \ldots & \dfrac{\partial \boldsymbol{\phi}(\boldsymbol{x})}{\partial x_N} \end{array}\right] \\ &= \left[\begin{array}{cccc} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_N \end{array}\right] + \left[\begin{array}{cccc} \boldsymbol{v}_1(\boldsymbol{x}) & \boldsymbol{v}_2(\boldsymbol{x}) & \ldots & \boldsymbol{v}_N(\boldsymbol{x}) \end{array}\right] \\ &\quad - D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\left[\begin{array}{cccc} \dfrac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_2} & \ldots & \dfrac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_N} \end{array}\right] \\ &= I + V(\boldsymbol{x}) - D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}D_{\boldsymbol{f}}(\boldsymbol{x}) \\ &= V(\boldsymbol{x})\end{aligned}$$

が得られる．定理 5 を用いると，多変数 Newton 法に対して次の定理が得られる．

定理 7 (多変数 Newton 法の収束条件) ある $q \in [0, 1)$ が存在して，任意の $\boldsymbol{x} \in K = B(\boldsymbol{x}^*, \varepsilon)$ に対して，

$$\|V(\boldsymbol{x})\| \leq q \tag{5.25}$$

が成り立つとき，初期値 $\boldsymbol{x}[0] \in K$ から出発する多変数 Newton 反復法 (5.22) は方程式 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ の解 $\boldsymbol{x}^*$ に収束する．

仮定より，方程式 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ の解 $\boldsymbol{x}^*$ に対して，$D_{\boldsymbol{f}}(\boldsymbol{x}^*)$ は正則であり，また $\boldsymbol{f}(\boldsymbol{x}^*) = \boldsymbol{0}$ であるので，

$$\boldsymbol{v}_i(\boldsymbol{x}^*) = D_{\boldsymbol{f}}(\boldsymbol{x}^*)^{-1}\frac{\partial D_{\boldsymbol{f}}(\boldsymbol{x}^*)}{\partial x_i}D_{\boldsymbol{f}}(\boldsymbol{x}^*)^{-1}\boldsymbol{f}(\boldsymbol{x}^*) = \boldsymbol{0}$$

となる．これより $D_{\boldsymbol{\phi}}(\boldsymbol{x}^*) = V(\boldsymbol{x}^*) = \mathbf{0}$ であることがわかる．また，$\boldsymbol{f}$ は領域 $K = B(\boldsymbol{x}^*, \varepsilon)$ で C^2 級であるので，$V(\boldsymbol{x})$ は K 上で連続となる．特にノルムの連続性から，$\|V_{\boldsymbol{f}}(\boldsymbol{x})\|$ は K 上で連続となる．したがって，ε を十分小さくとれば，不等式 (5.25) を満たす $q \in [0,1)$ は必ず存在する．すなわち，初期値 $\boldsymbol{x}[0]$ を $\boldsymbol{x}^*$ の十分近くにとれば，Newton 法 (5.22) は方程式 $\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0}$ の解 $\boldsymbol{x}^*$ に収束することがわかる．

なお，$\boldsymbol{f}(\boldsymbol{x})$ の “2 階微分” を

$$D^{(2)}_{i,\boldsymbol{f}}(\boldsymbol{x}) := \frac{\partial D_{\boldsymbol{f}}(\boldsymbol{x})}{\partial x_i}, \quad i = 1, 2, \ldots, N$$

とおくと，(5.24) より

$$\boldsymbol{v}_i(\boldsymbol{x}) = D_{\boldsymbol{f}}(\boldsymbol{x})^{-1} D^{(2)}_{i,\boldsymbol{f}}(\boldsymbol{x}) D_{\boldsymbol{f}}(\boldsymbol{x})^{-1} \boldsymbol{f}(\boldsymbol{x})$$

と書ける．これより，定理 7 の条件 (5.25) は，1 変数の Newton 法に対する収束条件 (5.19) の多変数への拡張であることがわかる．

練習問題 19 Newton 反復法 (5.22) に対して，$\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{x} - D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\boldsymbol{f}(\boldsymbol{x})$ とおく．任意の $\boldsymbol{x} \in K$ に対して，$D_{\boldsymbol{\phi}}(\boldsymbol{x})$ が正則ならば，$\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0}$ と $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{x})$ は K 上で同値であることを示せ．

練習問題 20 逆行列の微分の公式 (5.23) を証明せよ（ヒント： $M(\boldsymbol{x})^{-1}M(\boldsymbol{x}) = I$ の両辺を x_i で微分せよ）．

次に多変数 Newton 法の収束の速さを調べよう．まず次の補題を示す．

補題 3 K を $\mathbb{R}^N$ の凸集合とし，多変数関数 $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^N$ の Jacobi 行列 $D_{\boldsymbol{f}}(\boldsymbol{x})$ は K 上で Lipschitz 連続であるとする．すなわち，ある $r \geq 0$ が存在して，任意の $\boldsymbol{x}, \boldsymbol{y} \in K$ に対して

$$\|D_{\boldsymbol{f}}(\boldsymbol{x}) - D_{\boldsymbol{f}}(\boldsymbol{y})\| \leq r\|\boldsymbol{x} - \boldsymbol{y}\|$$

が成り立つとする．このとき，任意の $\boldsymbol{x}, \boldsymbol{y} \in K$ に対して

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y}) - D_{\boldsymbol{f}}(\boldsymbol{y} - \boldsymbol{x})\| \leq \frac{r}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2$$

が成り立つ．

証明

ベクトル $\boldsymbol{x}, \boldsymbol{y} \in K$ を任意にとり固定する．K は凸集合であるから，補題 2 の証明と同様にして，任意の $t \in [0,1]$ に対して

$$\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y}) = \int_0^1 D_{\boldsymbol{f}}\left(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})\right)(\boldsymbol{y} - \boldsymbol{x})dt$$

が成り立つ．これより，

$$\begin{aligned}\|\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{f}(\boldsymbol{y})-D_{\boldsymbol{f}}(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})\| &= \left\|\int_0^1 \left\{D_{\boldsymbol{f}}\left(\boldsymbol{x}+t(\boldsymbol{y}-\boldsymbol{x})\right)-D_{\boldsymbol{f}}(\boldsymbol{x})\right\}(\boldsymbol{y}-\boldsymbol{x})dt\right\| \\ &\le \int_0^1 \left\|D_{\boldsymbol{f}}\left(\boldsymbol{x}+t(\boldsymbol{y}-\boldsymbol{x})\right)-D_{\boldsymbol{f}}(\boldsymbol{x})\right\| \|(\boldsymbol{y}-\boldsymbol{x})\|dt \\ &\le \int_0^1 r\left\|\boldsymbol{x}+t(\boldsymbol{y}-\boldsymbol{x})-\boldsymbol{x}\right\| \|(\boldsymbol{y}-\boldsymbol{x})\|dt \\ &= \int_0^1 rt\|(\boldsymbol{y}-\boldsymbol{x})\|^2 dt \\ &= \frac{r}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2\end{aligned}$$

となることがわかる． 【証明終】

この補題を用いて，多変数 Newton 法が 2 次収束することを示すことができる．

定理 8 (多変数 Newton 法の収束の速さ) K を $\mathbb{R}^N$ の閉凸集合とし，多変数関数 $\boldsymbol{f}$: $\mathbb{R}^N \to \mathbb{R}^N$ の Jacobi 行列 $D_{\boldsymbol{f}}(\boldsymbol{x})$ は K 上で正則かつ Lipschitz 連続であるとする．このとき，多変数 Newton 法 (5.22) は K 上で 2 次収束する．すなわち，ある定数 $c>0$ が存在して，

$$\|\boldsymbol{x}[n+1]-\boldsymbol{x}^*\| \le c\|\boldsymbol{x}[n]-\boldsymbol{x}^*\|^2, \quad n=0,1,2,\ldots \tag{5.26}$$

が成り立つ．

証明

Newton 反復法の関係式 (5.22)，および $\boldsymbol{f}(\boldsymbol{x}^*)=\boldsymbol{0}$ より，

$$\begin{aligned}\boldsymbol{x}[n+1]-\boldsymbol{x}^* &= \boldsymbol{x}[n]-D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1}\boldsymbol{f}(\boldsymbol{x}[n])-\boldsymbol{x}^* \\ &= -D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1}\left\{\boldsymbol{f}(\boldsymbol{x}[n])-\boldsymbol{f}(\boldsymbol{x}^*)-D_{\boldsymbol{f}}(\boldsymbol{x}[n])(\boldsymbol{x}[n]-\boldsymbol{x}^*)\right\}\end{aligned}$$

が成り立つ．$D_{\boldsymbol{f}}(\boldsymbol{x})$ は K 上で正則であるので，

$$M := \max_{\boldsymbol{x}\in K}\|D_{\boldsymbol{f}}(\boldsymbol{x})^{-1}\| < \infty$$

である．したがって，補題 3 を使えば

$$\begin{aligned}\|\boldsymbol{x}[n+1]-\boldsymbol{x}^*\| &= \left\|D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1}\left\{\boldsymbol{f}(\boldsymbol{x}[n])-\boldsymbol{f}(\boldsymbol{x}^*)-D_{\boldsymbol{f}}(\boldsymbol{x}[n])(\boldsymbol{x}[n]-\boldsymbol{x}^*)\right\}\right\| \\ &\le \left\|D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1}\right\| \cdot \|\boldsymbol{f}(\boldsymbol{x}[n])-\boldsymbol{f}(\boldsymbol{x}^*)-D_{\boldsymbol{f}}(\boldsymbol{x}[n])(\boldsymbol{x}[n]-\boldsymbol{x}^*)\| \\ &\le \frac{Mr}{2}\|\boldsymbol{x}[n]-\boldsymbol{x}^*\|^2\end{aligned}$$

が成り立つことがわかる．したがって，$c := Mr/2$ とおけば，(5.26) が成り立つ．

【証明終】

5.4.4　Newton 法の大域的収束性

一般的な反復法

$$\boldsymbol{x}[n+1] = \boldsymbol{\phi}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots \tag{5.27}$$

を考える．この反復法が任意の $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，$\boldsymbol{\phi}$ の不動点 $\boldsymbol{x}^*$ に収束するとき，反復法 (5.27) は**大域的に収束する** (globally convergent) という．これに対して，これまで述べてきた反復法は，閉領域 K を決めて，その中での収束を議論した．このような反復法は，閉領域 K において**局所的に収束する** (locally convergent) という．反復法が不動点に大域的に収束すれば，初期値を $\mathbb{R}^N$ のどこにとっても良く，非常に便利である．ここでは，反復法（特に Newton 法）が大域的に収束する条件を考察する．

定理 1 および定理 3 において，$K = \mathbb{R}^N$ とすれば，次の定理が得られる．

定理 9 ある $q \in [0, 1)$ が存在して，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して，

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y})\| \leq q\|\boldsymbol{x} - \boldsymbol{y}\|$$

が成り立つならば，$\boldsymbol{\phi}$ は唯一つの不動点 $\boldsymbol{x}^* \in \mathbb{R}^N$ を持ち，反復法 (5.27) は大域的に $\boldsymbol{x}^*$ に収束する．

定理 10 写像 $\boldsymbol{\phi}$ は $\mathbb{R}^N$ において C^1 級であるとする．ある $q \in [0, 1)$ が存在して，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して

$$\|D_{\boldsymbol{\phi}}(\boldsymbol{x})\| \leq q$$

が成り立つならば，$\boldsymbol{\phi}$ は唯一つの不動点 $\boldsymbol{x}^* \in \mathbb{R}^N$ を持ち，反復法 (5.27) は大域的に $\boldsymbol{x}^*$ に収束する．

上の定理の条件を満たすような写像 $\boldsymbol{\phi}$ は非常に限られている．特に $\boldsymbol{\phi}(\boldsymbol{x})$ が $\boldsymbol{x}$ に対して非線形である場合，条件が満たされるかどうかを調べることも一般には非常に難しい．そこで本節では Newton 法に限定して，別のアプローチから大域的収束性を調べる．

まず，$\mathbb{R}^N$ のベクトルおよび行列に対する大小関係を定義する．$\mathbb{R}^N$ のベクトル $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^\top$ と $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^\top$ について，$x_i \leq y_i$, $i = 1, 2, \ldots, N$ が成り立つとき，$\boldsymbol{x} \preceq \boldsymbol{y}$ と書く．また $N \times N$ 行列 $A = [a_{ij}]$ と $B = [b_{ij}]$ について，$a_{ij} \leq b_{ij}$, $i, j = 1, 2, \ldots, N$ が成り立つとき，$A \preceq B$ と書く．例えば，

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \preceq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \preceq \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \preceq \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

などである．また，$\boldsymbol{x} \preceq \boldsymbol{y}$ を $\boldsymbol{y} \succeq \boldsymbol{x}$，$A \preceq B$ を $B \succeq A$ とも書く．

これらの記法のもとに多変数関数の凸性を定義する．

定義 3 (凸関数) 多変数関数 $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^N$ が $\mathbb{R}^N$ 上の**凸関数** (convex function) であるとは，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ と任意の $\alpha \in [0, 1]$ に対して，

$$\boldsymbol{f}(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \preceq \alpha\boldsymbol{f}(\boldsymbol{x}) + (1-\alpha)\boldsymbol{f}(\boldsymbol{y})$$

が成り立つことである．

関数 $\boldsymbol{f}$ の凸性に関して，次の補題が成り立つ．

補題 4 関数 $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^N$ は $\mathbb{R}^N$ で微分可能であるとする．このとき，$\boldsymbol{f}$ が $\mathbb{R}^N$ 上の凸関数であるための必要十分条件は，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して

$$\boldsymbol{f}(\boldsymbol{y}) \succeq \boldsymbol{f}(\boldsymbol{x}) + D_{\boldsymbol{f}}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) \tag{5.28}$$

が成り立つことである．

証明

まず，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して (5.28) が成り立つと仮定する．任意に選んで固定したベクトル $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$ と定数 $\alpha \in [0, 1]$ に対して，

$$\boldsymbol{z} := \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v} \in \mathbb{R}^N$$

とおく．このとき，

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{u}) &\succeq \boldsymbol{f}(\boldsymbol{z}) + D_{\boldsymbol{f}}(\boldsymbol{z})(\boldsymbol{u} - \boldsymbol{z}) \\ \boldsymbol{f}(\boldsymbol{v}) &\succeq \boldsymbol{f}(\boldsymbol{z}) + D_{\boldsymbol{f}}(\boldsymbol{z})(\boldsymbol{v} - \boldsymbol{z})\end{aligned}$$

が成り立つ．第一式に $\alpha \geq 0$，第二式に $1 - \alpha \geq 0$ を掛けて，足し合わせると，

$$\begin{aligned}\alpha\boldsymbol{f}(\boldsymbol{u}) + (1-\alpha)\boldsymbol{f}(\boldsymbol{v}) &\succeq \boldsymbol{f}(\boldsymbol{z}) + D_{\boldsymbol{f}}(\boldsymbol{z})(\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v} - \boldsymbol{z}) \\ &= \boldsymbol{f}(\boldsymbol{z}) \\ &= \boldsymbol{f}(\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v})\end{aligned}$$

が得られる．ベクトル $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$ と定数 $\alpha \in [0, 1]$ は任意であったので，$\boldsymbol{f}$ は $\mathbb{R}^N$ 上の凸関数であることがわかる．

逆に，$\boldsymbol{f}$ が $\mathbb{R}^N$ 上の凸関数であると仮定すると，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ と任意の $\alpha \in (0, 1]$ に対して，

$$\boldsymbol{f}(\alpha\boldsymbol{y} + (1-\alpha)\boldsymbol{x}) \preceq \alpha\boldsymbol{f}(\boldsymbol{y}) + (1-\alpha)\boldsymbol{f}(\boldsymbol{x})$$

これを変形すると，

$$\boldsymbol{f}(\boldsymbol{y}) \succeq \boldsymbol{f}(\boldsymbol{x}) + \frac{\boldsymbol{f}\left(\boldsymbol{x} + \alpha(\boldsymbol{x} - \boldsymbol{y})\right) - \boldsymbol{f}(\boldsymbol{x})}{\alpha}$$

写像 $\boldsymbol{f}$ は $\mathbb{R}^N$ 上で微分可能であるので，

$$\lim_{\alpha \to 0} \frac{\boldsymbol{f}\left(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x})\right) - \boldsymbol{f}(\boldsymbol{x})}{\alpha} = D_{\boldsymbol{f}}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})$$

が成り立つ．したがって，(5.28) が成り立つ．　【証明終】

以上の準備のもとで，方程式 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ に対する Newton 法

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1} \boldsymbol{f}(\boldsymbol{x}[n]) \tag{5.29}$$

の大域的収束性を考えよう．次の定理が成り立つ [20].

定理 11 (Newton 法の大域的収束性) 関数 $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^N$ を $\mathbb{R}^N$ の全域で C^1 級かつ凸であるとする．また $D_{\boldsymbol{f}}(\boldsymbol{x})$ は任意の $\boldsymbol{x} \in \mathbb{R}^N$ に対して正則かつ $D_{\boldsymbol{f}}(\boldsymbol{x})^{-1} \succeq 0$ であるとする[*11]．また方程式 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ は解を少なくとも一つ持つとする．このとき，$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ の解は唯一つ存在し，それを $\boldsymbol{x}^*$ とおくと，Newton 法 (5.29) は任意の初期値 $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，$\boldsymbol{x}^*$ に収束する．さらに，ベクトル列 $\{\boldsymbol{x}[n]\}_{n \geq 1}$ は単調非増加である．すなわち，$\boldsymbol{x}[1] \succeq \boldsymbol{x}[2] \succeq \cdots \succeq \boldsymbol{x}^*$ が成り立つ．

この定理においては，関数 $\boldsymbol{f}$ が凸であることが本質的である．非線形な関数に対する反復法の収束性の判別は一般に難しいが，その関数が凸であれば非線形であっても比較的容易である．これは，最適化理論においても同様であり，非線形最適化は一般に非常に難しいが，凸最適化は比較的容易で，盛んに研究されている[*12]．このように，非線形問題というのは，一般に非常に難しい問題であるが，その非線形が凸という性質を持てば，それは容易となる場合が多い．

練習問題 21 次の関数が凸であるかどうかを調べよ．

1. $f(x) = ax, \quad x \in \mathbb{R}$．ただし，$a$ は実数．
2. $f(x) = ax^2, \quad x \in \mathbb{R}$．ただし，$a$ は実数．
3. $\boldsymbol{f}(\boldsymbol{x}) = A\boldsymbol{x}$, $\boldsymbol{x} \in \mathbb{R}^N$．ただし，$A$ は $N \times N$ 行列．
4. $\boldsymbol{f}(\boldsymbol{x}) = \left[a_1 x_1^2,\ a_2 x_2^2, \ldots, a_N x_N^2\right]^\top$, $\boldsymbol{x} = [x_1, \ldots, x_N]^\top \in \mathbb{R}^N$．ただし，$a_i$ $(i = 1, 2, \ldots, N)$ は実数．

[*11] 行列 A に対して $A \succeq 0$ は，A の全ての要素が非負（すなわち，0 以上）であるという意味である．
[*12] 凸最適化においても Newton 法はよく用いられる

5.4.5 簡易 Newton 法

多変数の Newton 法

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1}\boldsymbol{f}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots$$

において，各ステップで Jacobi 行列 $D_{\boldsymbol{f}}(\boldsymbol{x}[\boldsymbol{n}])$ の逆行列を求める必要がある．しかし，$\boldsymbol{x}[n]$ のサイズが非常に大きい場合，この逆行列の計算に大きな時間がかかってしまうことがある．このような場合には，各ステップで Jacobi 行列の逆行列を求めることをあきらめ，初期値 $x[0]$ における逆行列 $D_{\boldsymbol{f}}(\boldsymbol{x}[\boldsymbol{0}])^{-1}$ を代わりに使う方法があり，**簡易 Newton 法** (simplified Newton method) と呼ばれる．すなわち，次の反復により $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ の近似解を求める．

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - D_{\boldsymbol{f}}(\boldsymbol{x}[0])^{-1}\boldsymbol{f}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots.$$

はじめに $D_{\boldsymbol{f}}(\boldsymbol{x}[\boldsymbol{0}])^{-1}$ を求めておけば，あとは掛け算と足し算だけで計算でき，各ステップにおける計算量は大幅に減少する．ただし，各ステップで微分の情報が使えないため，収束の速さは通常の Newton 法より遅い．なお，1 次元の簡易 Newton 法を von Mises **法** (von Mises' method) とも呼ぶ

5.5 反復法の誤差解析

一般の反復法

$$\boldsymbol{x}[n+1] = \boldsymbol{\phi}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots \tag{5.30}$$

を考える．この反復法は，任意の初期値 $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して不動点 $\boldsymbol{x}^*$ に収束すると仮定する[*13]．実際の反復法はコンピュータにより実行されるので，有限回で反復を打ち切るときに打切り誤差が生じる．また，計算過程で丸め誤差も生じる．ここではこれらの誤差が計算結果にどう影響するかを調べる．

まず，打切り誤差について考える．反復法 (5.30) を計算機で実行する場合，反復は有限回で終了しなければならない．例えば，N 回の反復で得られる近似解 $\boldsymbol{x}[N]$ は厳密解 $\boldsymbol{x}^*$ とは一致せず，一般には打切り誤差 $\|\boldsymbol{x}^* - \boldsymbol{x}[N]\|$ が発生する．この打ち切り誤差に関して，以下の定理が成り立つ．

[*13] 不動点の近傍の領域だけで考察することも可能であるが，ここでは簡単のため反復法は大域的に収束すると仮定する．

定理 12 (反復法の打ち切り誤差) 写像 $\boldsymbol{\phi}$ を Lipschitz 定数 $q \in [0,1)$ を持つ $\mathbb{R}^N$ 上の縮小写像とする．すなわち，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y})\| \leq q\|\boldsymbol{x} - \boldsymbol{y}\|$$

が成り立つとする．このとき，反復法 (5.30) を N 回で打ち切ったときの打切り誤差 $\|\boldsymbol{x}^* - \boldsymbol{x}[N]\|$ に関して次式が成り立つ．

$$\|\boldsymbol{x}^* - \boldsymbol{x}[N]\| \leq \frac{q^N}{1-q}\|\boldsymbol{\phi}(\boldsymbol{x}[0]) - \boldsymbol{x}[0]\|. \tag{5.31}$$

証明

任意の自然数 m と n （$m > n$ とする）に対して，

$$\|\boldsymbol{x}[m] - \boldsymbol{x}[n]\| < \frac{q^n}{1-q}\|\boldsymbol{x}[1] - \boldsymbol{x}[0]\| = \frac{q^n}{1-q}\|\boldsymbol{\phi}(\boldsymbol{x}[0]) - \boldsymbol{x}[0]\|$$

が成り立つ[*14]．これより，任意の $k \geq 0$ に対して，

$$\|\boldsymbol{x}[N+k] - \boldsymbol{x}[N]\| < \frac{q^N}{1-q}\|\boldsymbol{\phi}(\boldsymbol{x}[0]) - \boldsymbol{x}[0]\|$$

ここで，$\boldsymbol{x}^* = \lim_{k\to\infty} \boldsymbol{x}[N+k]$ と表せるから，上の不等式より (5.31) が成り立つことがわかる．【証明終】

次に反復法における丸め誤差の影響を考える．まず，各ステップにおいて丸め誤差は有界であると仮定する．すなわち，ある $\delta > 0$ が存在して，第 n ステップにおける丸め誤差 $\boldsymbol{d}[n]$ について

$$\sup_{n\geq 0} \|\boldsymbol{d}[n]\| \leq \delta \tag{5.32}$$

が成り立つと仮定する．ここで，初期値 $\boldsymbol{x}[0]$ は自由に決めることができるため，初期値には誤差は入らないものとする．すなわち，初期値は計算機で表現できる範囲内の数であるとする．

まず，反復法 (5.30) により $\boldsymbol{x}[1] = \boldsymbol{\phi}(\boldsymbol{x}[0])$ が得られる．この計算に丸め誤差 $\boldsymbol{d}[0]$ が入るため，

$$\widetilde{\boldsymbol{x}}[1] = \boldsymbol{\phi}(\boldsymbol{x}[0]) + \boldsymbol{d}[0]$$

となる．以下同様にして，

$$\widetilde{\boldsymbol{x}}[n+1] = \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}[n]) + \boldsymbol{d}[n], \quad n = 0, 1, 2, \ldots, \quad \widetilde{\boldsymbol{x}}[0] = \boldsymbol{x}[0] \tag{5.33}$$

が得られる．これが丸め誤差を含んだ反復法の表現である．以上の準備のもとで，以下の定理が成り立つ．

[*14] 5.1 節，定理 1 の証明を参照せよ．

定理 13 (反復法の丸め誤差) 写像 $\boldsymbol{\phi}$ を Lipschitz 定数 $q \in [0,1)$ を持つ $\mathbb{R}^N$ 上の縮小写像とする．すなわち，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y})\| \leq q\|\boldsymbol{x} - \boldsymbol{y}\|$$

が成り立つとする．このとき，丸め誤差を含んだ反復法 (5.33) を N 回で打ち切ったときの数値誤差 $\|\boldsymbol{x}^* - \widetilde{\boldsymbol{x}}[N]\|$ に関して次式が成り立つ．

$$\|\boldsymbol{x}^* - \widetilde{\boldsymbol{x}}[N]\| \leq \frac{q^N}{1-q}\|\boldsymbol{\phi}(\boldsymbol{x}[0]) - \boldsymbol{x}[0]\| + \frac{1-q^N}{1-q}\delta. \tag{5.34}$$

証明

(5.30) と (5.33) より

$$\begin{aligned}
\|\boldsymbol{x}[N] - \widetilde{\boldsymbol{x}}[N]\| &= \|\boldsymbol{\phi}(\boldsymbol{x}[N-1]) - \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}[N-1]) - \boldsymbol{d}[N-1]\| \\
&\leq \|\boldsymbol{\phi}(\boldsymbol{x}[N-1]) - \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}[N-1])\| + \|\boldsymbol{d}[N-1]\| \\
&\leq q\|\boldsymbol{x}[N-1] - \widetilde{\boldsymbol{x}}[N-1]\| + \delta \\
&\leq q\left\{q\|\boldsymbol{x}[N-2] - \widetilde{\boldsymbol{x}}[N-2]\| + \delta\right\} + \delta \\
&= q^2\|\boldsymbol{x}[N-2] - \widetilde{\boldsymbol{x}}[N-2]\| + (q+1)\delta \\
&\leq \cdots \\
&\leq q^N\|\boldsymbol{x}[0] - \widetilde{\boldsymbol{x}}[0]\| + (q^{N-1} + q^{N-2} + \cdots + q + 1)\delta \\
&= \frac{1-q^N}{1-q}\delta
\end{aligned}$$

が成り立つ．ここで，$\boldsymbol{\phi}$ が Lipschitz 定数 q を持つ縮小写像であること，および (5.32) を用いた．また，$\widetilde{\boldsymbol{x}}[0] = \boldsymbol{x}[0]$ であることも用いた．上の不等式と (5.31) を用いると，

$$\begin{aligned}
\|\boldsymbol{x}^* - \widetilde{\boldsymbol{x}}[N]\| &= \|\boldsymbol{x}^* - \boldsymbol{x}[N] + \boldsymbol{x}[N] - \widetilde{\boldsymbol{x}}[N]\| \\
&\leq \|\boldsymbol{x}^* - \boldsymbol{x}[N]\| + \|\boldsymbol{x}[N] - \widetilde{\boldsymbol{x}}[N]\| \\
&\leq \frac{q^N}{1-q}\|\boldsymbol{\phi}(\boldsymbol{x}[0]) - \boldsymbol{x}[0]\| + \frac{1-q^N}{1-q}\delta
\end{aligned}$$

が成り立つことがわかる．　【証明終】

この定理から，縮小写像 $\boldsymbol{\phi}$，Lipschitz 定数 q，初期値 $\boldsymbol{x}[0]$，丸め誤差の最大値 δ が与えられれば，数値計算の実行前に誤差を評価できることがわかる．また，δ と q が小さければ，誤差も少ないことがわかる．ここで，δ が小さくするには，精度の良い計算機を使用する必要があり，q を小さくするには良いアルゴリズムを採用する必要がある．

5.6　さらに勉強するために

本章の数値計算における反復法の収束性の解析と誤差解析は，数値計算の中心的なトピックスであり，数値計算の教科書 [46, 49, 37, 22] などにも多くの記述がある．一般的な Banach 空間における不動点定理に関しては，[38, 30, 43] などが参考になる．5.4.4 節の Newton 法の大域的収束性については，[20] を参考にした．5.4.5 節で紹介した簡易 Newton 法の収束条件に関しては，[43] を参照せよ．

第 6 章

線形方程式の反復解法

本章では，N 変数の連立線形方程式の近似解を求める方法を解説する．すなわちここでは，方程式 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{b}$ において，写像 $\boldsymbol{f}$ が変数 $\boldsymbol{x}$ に関して**線形** (linear) であるものを考える．写像 $\boldsymbol{f}$ が変数 $\boldsymbol{x}$ に関して線形であるとは，任意のベクトル $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ と任意のスカラー $\alpha, \beta \in \mathbb{R}$ に対して，

$$\boldsymbol{f}(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha\boldsymbol{f}(\boldsymbol{x}) + \beta\boldsymbol{f}(\boldsymbol{y})$$

が成り立つことである．有限次元ベクトル空間における線形写像 $\boldsymbol{f}$ は行列で表現することができる．したがって，線形方程式として本章では特に以下の方程式を考察する．

$$A\boldsymbol{x} = \boldsymbol{b}, \quad A \in \mathbb{R}^{N\times N}, \quad \boldsymbol{x} \in \mathbb{R}^N, \quad \boldsymbol{b} \in \mathbb{R}^N.$$

6.1 Newton 法と線形方程式

非線形方程式 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ の近似解を求める Newton 反復法

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - D_{\boldsymbol{f}}(\boldsymbol{x}[n])^{-1}\boldsymbol{f}(\boldsymbol{x}[n]), \quad n = 0, 1, 2, \ldots$$

を計算機で実行する場合，通常，次の二つのステップを繰り返して計算を行う．

1. 線形方程式 $D_{\boldsymbol{f}}(\boldsymbol{x}[n])\boldsymbol{y} = -\boldsymbol{f}(\boldsymbol{x}[n])$ を解き，$\boldsymbol{y}$ を求める．
2. $\boldsymbol{x}[n+1] = \boldsymbol{x}[n] + \boldsymbol{y}$ により $n+1$ ステップ目の近似解を計算する．

すなわち，各ステップにおいて線形方程式を解く必要がある．Newton 法は非常に収束が速いアルゴリズムであるが，この線形方程式を解くのに時間がかかっていては Newton 法を使うご利益はあまりない[*1]．例えば，線形代数で習う Gauss **の消去法** (Gaussian

[*1] さらに，Newton 法では関数 $\boldsymbol{f}(\boldsymbol{x})$ の微分 $D_{\boldsymbol{f}}(\boldsymbol{x})$ を陽に求める必要があるが，この微分は一般に厳密に求めることは難しい．そこで厳密に微分を求めずに Newton 法を実行するうまい方法（**準 Newton 法**

elimination) [16, 22] は，有限回の計算で線形方程式の解を厳密に求めることができるが，その計算量（足し算，引き算，掛け算，割り算の総数）は，行列 A のサイズを $N \times N$ とすると N^3 に比例する．例えば 行列のサイズが 10 倍になれば，計算時間は 1000 倍になるのである．これでは，実用的な方程式（$N = 10^5$ など）を解くことは非常に難しくなる．そこで，このようにサイズの大きな線形方程式に対しては，厳密解を求めることをあきらめ，反復法により高速に近似解を求めることが必要となる．

練習問題 22 1G (Giga $= 10^9$) FLOPS のコンピュータ（1 秒間に浮動小数点数演算を 10^9 回（$=$ 10 億回）行えるコンピュータ）で $N = 10000$ の線形方程式を Gauss の消去法で計算した場合，およそどれだけの計算時間がかかるか？ 1T (Tera $= 10^{12}$) FLOPS, 1P (Peta $= 10^{15}$) FLOPS のコンピュータではどうか？

6.2　線形方程式の反復法

線形方程式

$$A\boldsymbol{x} = \boldsymbol{b} \tag{6.1}$$

に対する反復法として次の形式を考える．

$$\boldsymbol{x}[n+1] = M\boldsymbol{x}[n] + \boldsymbol{v}, \quad n = 0, 1, 2, \ldots \tag{6.2}$$

ここで，M は $N \times N$ の行列，$\boldsymbol{v}$ は $\mathbb{R}^N$ のベクトルであり，方程式

$$\boldsymbol{x} = M\boldsymbol{x} + \boldsymbol{v} = \boldsymbol{\phi}(\boldsymbol{x}) \tag{6.3}$$

がもとの方程式 (6.1) と等価になるように M と $\boldsymbol{v}$ を選ぶ．

例題 7 次の連立方程式を考える．

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1 \\ 4x_1 + 5x_2 + 6x_3 &= 2 \\ 7x_1 + 8x_2 + 10x_3 &= 4 \end{aligned}$$

次のようにして反復法 (6.2) を作る．まず，方程式の対角成分だけを左辺に残し，残りを右辺に移項する．

$$\begin{aligned} x_1 &= -2x_2 - 3x_3 + 1 \\ 5x_2 &= -4x_1 - 6x_3 + 2 \\ 10x_3 &= -7x_1 - 8x_2 + 4 \end{aligned}$$

(quasi-Newton method) と呼ばれる）が知られている [47, 21].

これより，次の反復法を得る．

$$\begin{aligned} x_1[n+1] &= -2x_2[n] - 3x_3[n] + 1 \\ x_2[n+1] &= -\frac{4}{5}x_1[n] - \frac{6}{5}x_3[n] + \frac{2}{5} \\ x_3[n+1] &= -\frac{7}{10}x_1[n] - \frac{8}{10}x_2[n] + \frac{4}{10} \end{aligned}$$

すなわち，

$$M = \begin{bmatrix} 0 & -2 & -3 \\ -4/5 & 0 & -6/5 \\ -7/10 & -8/10 & 0 \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} 1 \\ 2/5 \\ 4/10 \end{bmatrix}$$

である．この方法を Jacobi **法** (Jacobi method) と呼ぶ．この方法では，$x_2[n+1]$ を計算するのに $x_1[n]$ を使い，$x_3[n+1]$ を計算するのに $x_1[n]$ と $x_2[n]$ を使っている．しかし，それらを計算する時点では更新された $x_1[n+1]$ や $x_2[n+1]$ を使うほうが精度がよりあがりそうである．したがって，1 行目で計算した $x_1[n+1]$ を 2 行目以降で使用し，2 行目で計算した $x_2[n+1]$ を 3 行目で使用することを考える．すなわち次のような反復法を作る．

$$\begin{aligned} x_1[n+1] &= -2x_2[n] - 3x_3[n] + 1 \\ x_2[n+1] &= -\frac{4}{5}x_1[n+1] - \frac{6}{5}x_3[n] + \frac{2}{5} \\ x_3[n+1] &= -\frac{7}{10}x_1[n+1] - \frac{8}{10}x_2[n+1] + \frac{4}{10} \end{aligned}$$

この方法を Gauss-Seidel **法** (Gauss-Seidel method) と呼ぶ．

上記の Jacobi 法や Gauss-Seidel 法を一般の方程式 (6.1) に対して構成しよう．方程式 (6.1) の行列 A を対角行列 D, 下三角行列 E および 上三角行列 F を用いて，$A = D + E + F$ と分解する．すなわち, 行列 A の第 ij 要素を a_{ij}, $i, j = 1, \ldots, N$ としたとき，

$$D = \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & a_{NN} \end{bmatrix}, \quad E = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ a_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{N1} & \ldots & a_{N,N-1} & 0 \end{bmatrix},$$

$$F = \begin{bmatrix} 0 & a_{12} & \ldots & a_{1N} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{N-1,N} \\ 0 & \ldots & 0 & 0 \end{bmatrix}$$

である．この分解を用いれば，Jacobi 法と Gauss-Seidel 法の反復法における写像 $\boldsymbol{\phi}(\boldsymbol{x})$ は次で与えられる．

Jacobi **法**　$\boldsymbol{\phi}(\boldsymbol{x}) = \underbrace{-D^{-1}(E+F)}_{M}\boldsymbol{x} + \underbrace{D^{-1}\boldsymbol{b}}_{\boldsymbol{v}}$

Gauss-Seidel **法**　$\boldsymbol{\phi}(\boldsymbol{x}) = \underbrace{-(D+E)^{-1}F}_{M}\boldsymbol{x} + \underbrace{(D+E)^{-1}\boldsymbol{b}}_{\boldsymbol{v}}$

またこれらの方法を改良し，収束を速めた次の方法もよく知られている．

$$\boldsymbol{\phi}(\boldsymbol{x}) = \underbrace{(I+\omega D^{-1}E)^{-1}\left\{(1-\omega)I - \omega D^{-1}F\right\}}_{M}\boldsymbol{x} + \underbrace{\omega(D+\omega E)^{-1}\boldsymbol{b}}_{\boldsymbol{v}}$$

これを**加速緩和法** (Successive over-relaxation method) , または SOR **法** (Successive over-relaxation method) と呼ぶ．この方法における ω は加速パラメータであり，$0 < \omega < 2$ の範囲で収束が早まるようにうまく選ぶ必要がある．

練習問題 23 Jacobi 法，Gauss-Seidel 法，SOR 法におけるそれぞれの $\boldsymbol{\phi}$ に対して，方程式 $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{x})$ が方程式 $A\boldsymbol{x} = \boldsymbol{b}$ と同値であることを示せ．

練習問題 24　次の方程式に対して，Jacobi 法および Gauss-Seidel 法による反復法により近似解を求めよ．

$$\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

6.3　反復法の収束性

前章までで見たように，反復法 (6.2) によって生成されるベクトル列 $\{\boldsymbol{x}[n]\}$ が方程式 (6.1) の厳密解 $\boldsymbol{x}^*$ に収束するかどうかが問題となる．そこでここでは，反復法 (6.2) の収束性について調べる．これ以降，$\mathbb{R}^N$ のベクトルのノルム $\|\cdot\|$ を（たとえば Euclid ノルムに）固定し，行列のノルムをそのノルムから誘導されるノルムとする．

補題 5 (縮小写像であるための必要十分条件) 方程式 (6.3) の $\boldsymbol{\phi}$ が $\mathbb{R}^N$ における縮小写像であるための必要十分条件は $\|M\| < 1$ である．

証明

$\boldsymbol{\phi}$ を $\mathbb{R}^N$ の縮小写像とする．すなわち，ある $q \in (0,1)$ が存在して，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して，

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y})\| \leq q\|\boldsymbol{x} - \boldsymbol{y}\|$$

が成り立つとする．$\boldsymbol{\phi}(\boldsymbol{x}) = M\boldsymbol{x} + \boldsymbol{v}$ であったので，

$$\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y}) = M\boldsymbol{x} + \boldsymbol{v} - (M\boldsymbol{y} + \boldsymbol{v}) = M(\boldsymbol{x} - \boldsymbol{y}). \tag{6.4}$$

これより，

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y})\| = \|M(\boldsymbol{x} - \boldsymbol{y})\| \leq q\|\boldsymbol{x} - \boldsymbol{y}\|$$

が成り立つ．ここで，$\boldsymbol{y} = \boldsymbol{0}$ とおくと，$\boldsymbol{0}$ でない任意の $\boldsymbol{x} \in \mathbb{R}^N$ に対して

$$\|M\boldsymbol{x}\| \leq q\|\boldsymbol{x}\| < \|\boldsymbol{x}\|, \quad \forall \boldsymbol{x} \in \mathbb{R}^N$$

が成り立ち，これより

$$\|M\| = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|M\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq q < 1$$

が成り立つことがわかる．

逆に $\|M\| < 1$ とする．(6.4) および行列のノルムの性質 ($\|M(\boldsymbol{x} - \boldsymbol{y}\| \leq \|M\|\|\boldsymbol{x} - \boldsymbol{y}\|$) を使えば，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ に対して，

$$\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{y})\| = \|M(\boldsymbol{x} - \boldsymbol{y})\| \leq \|M\|\|\boldsymbol{x} - \boldsymbol{y}\|$$

が成り立つことがわかる．$M = 0$ とすると $\|M\| = 0$ であり，$\boldsymbol{\phi}$ は明らかに縮小写像である．また，$M \neq 0$ とすると $\|M\| > 0$ であり，また仮定より $\|M\| < 1$．ゆえに $q := \|M\|$ が Lipschitz 定数となり，写像 $\boldsymbol{\phi}$ は Lipschitz 条件を満たす．また，明らかに $\boldsymbol{x} \in \mathbb{R}^N$ ならば $\boldsymbol{\phi}(\boldsymbol{x}) = M\boldsymbol{x} + \boldsymbol{v} \in \mathbb{R}^N$ である．以上より，$\boldsymbol{\phi}$ は $\mathbb{R}^N$ の縮小写像であることが示された．

【証明終】

この補題を用いれば，次の定理が成り立つことがわかる．

定理 14 (反復法が収束するための十分条件) もし $\|M\| < 1$ ならば，線形方程式 $\boldsymbol{x} = M\boldsymbol{x} + \boldsymbol{v}$ は唯一つの解 $\boldsymbol{x}^*$ を持ち，任意の初期ベクトル $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，反復法 (6.2) は $\boldsymbol{x}^*$ に収束する．

この定理は収束のための十分条件であり，$\|M\| < 1$ が成り立たなくても任意の初期ベクトル $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して反復法が収束することがある．実際，収束する反復法で $\|M\|$ の値を任意に大きくできるものが存在する．

練習問題 25 任意の実数 $R >$ に対して，ある行列 M が存在して，

$$\|M\| > R$$

かつ

$$\lim_{n \to \infty} M^n = 0$$

となる（すなわち反復法が収束する）ことを示せ．（ヒント：有限の $n \geq 2$ で $M^n = 0$ となるような行列を考えよ．）

反復法が収束するための必要十分条件は，一般の非線形方程式の場合は得ることが難しい．しかし，線形方程式の場合には，次のように必要十分条件が得られる．

定理 15 (反復法が収束するための必要十分条件) 線形方程式

$$\boldsymbol{x} = M\boldsymbol{x} + \boldsymbol{v} \tag{6.5}$$

は一意解を持つとする[*2]．任意の初期ベクトル $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して反復法

$$\boldsymbol{x}[n+1] = M\boldsymbol{x}[n] + \boldsymbol{v}, \quad n = 0, 1, 2, \ldots \tag{6.6}$$

が線形方程式 (6.5) の厳密解 $\boldsymbol{x}^*$ に収束するための必要十分条件は，$\rho(M) < 1$ である．ただし，$\rho(M)$ は行列 M の**スペクトル半径** (spectral radius) であり，次式で定義される．

$$\rho(M) := \max_{1 \le i \le N} |\lambda_i(M)| .$$

ここで，$\lambda_1(M), \ldots, \lambda_N(M)$ は行列 M の固有値を表す．

証明

反復法 (6.6) が任意の初期ベクトル $\boldsymbol{x}[0]$ に対して厳密解 $\boldsymbol{x}^*$ に収束すると仮定する．このとき，第 n ステップにおける近似解 $\boldsymbol{x}[n]$ と厳密解 $\boldsymbol{x}^*$ との誤差を $\boldsymbol{e}[n]$ とおくと，

$$\begin{aligned}
\boldsymbol{e}[n] &= \boldsymbol{x}[n] - \boldsymbol{x}^* \\
&= (M\boldsymbol{x}[n-1] + \boldsymbol{v}) - (M\boldsymbol{x}^* + \boldsymbol{v}) \\
&= M(\boldsymbol{x}[n-1] - \boldsymbol{x}^*) \\
&= M\boldsymbol{e}[n-1]
\end{aligned}$$

となる．これより，

$$\boldsymbol{e}[n] = M^n\boldsymbol{e}[0] = M^n(\boldsymbol{x}[0] - \boldsymbol{x}^*), \quad n = 0, 1, 2, \ldots$$

が成り立つ．ここで M の固有値を λ_i $(i = 1, 2, \ldots, N)$ とし，λ_i に属する M の固有ベクトルを $\boldsymbol{w}_i$ とおく．初期ベクトルを $\boldsymbol{x}[0] = \boldsymbol{w}_i + \boldsymbol{x}^*$ $(i = 1, 2, \ldots, N)$ ととると，

$$\begin{aligned}
\boldsymbol{e}[n] &= M^n(\boldsymbol{w}_i + \boldsymbol{x}^* - \boldsymbol{x}^*) \\
&= M^n\boldsymbol{w}_i \\
&= M^{n-1}(M\boldsymbol{w}_i) \\
&= M^{n-1}(\lambda_i\boldsymbol{w}_i) \\
&= \cdots \\
&= \lambda_i^n\boldsymbol{w}_i
\end{aligned}$$

[*2] すなわちもとの方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の行列 A は正則とする．

となることがわかる．これより，

$$\|\boldsymbol{e}[n]\| = \|\lambda_i^n \boldsymbol{w}_i\| = |\lambda_i|^n \|\boldsymbol{w}_i\|, \quad i = 1, 2, \ldots, N$$

が成り立つ．今，任意の $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，反復法 (6.6) は厳密解に収束するので，

$$\lim_{n\to\infty} \boldsymbol{e}[n] = \boldsymbol{0}$$

となる．これより，

$$\lim_{n\to\infty} \|\boldsymbol{e}[n]\| = \lim_{n\to\infty} |\lambda_i|^n \|\boldsymbol{w}_i\| = \boldsymbol{0}, \quad i = 1, 2, \ldots, N$$

となることがわかる．ここで $\boldsymbol{w}_i \neq \boldsymbol{0}$ であるので，すべての $i = 1, 2, \ldots, N$ で $|\lambda_i| < 1$ でなければならない．すなわち，$\rho(M) < 1$ である．

逆に $\rho(M) < 1$ であると仮定すると，$1 - \rho(M) > 0$ であるので，ある実数 $\varepsilon > 0$ が存在して，$1 - \rho(M) > \varepsilon$ が成り立つ．この ε に対して，ある行列ノルム $\|\cdot\|_\alpha$ が存在して，

$$\|M\|_\alpha \leq \rho(M) + \varepsilon < \rho(M) + 1 - \rho(M) = 1$$

が成り立つ[*3]．したがって，補題 5 より，この行列ノルム $\|\cdot\|_\alpha$ を誘導するベクトルノルム $\|\cdot\|_\alpha$ のもとで，$\boldsymbol{\phi}(\boldsymbol{x}) = M\boldsymbol{x} + \boldsymbol{v}$ は $\mathbb{R}^N$ の縮小写像となり，反復法 (6.6) によって生成されるベクトル列 $\{\boldsymbol{x}[n]\}$ は，任意の初期ベクトル $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，このノルムの意味で厳密解 $\boldsymbol{x}^*$ に収束する．すなわち，任意の初期ベクトル $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，

$$\lim_{n\to\infty} \|\boldsymbol{x}[n] - \boldsymbol{x}^*\|_\alpha = \boldsymbol{0}$$

が成り立つ．ノルムの連続性とノルムの公理 ($\|\boldsymbol{x}\|_\alpha = \boldsymbol{0}$ ならば $\boldsymbol{x} = \boldsymbol{0}$) より，

$$\lim_{n\to\infty} \boldsymbol{x}[n] = \boldsymbol{x}^*$$

であることがわかる． 【証明終】

この定理にもとづき，スペクトル半径が 1 未満の行列（すなわち全ての固有値の絶対値が 1 未満の行列）を Schur **安定行列** (Schur matrix) または単に，**安定行列** (stable matrix) と呼ぶことがある．次の反復法

$$\boldsymbol{x}[n+1] = M\boldsymbol{x}[n]$$

において，任意の初期ベクトル $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して，

$$\lim_{n\to\infty} \boldsymbol{x}[n] = \boldsymbol{0}$$

[*3] 付録の定理 34 を参照のこと．

となるための必要十分条件は M が Schur 安定行列であることである[*4]．この事実は，線形システムの安定性と密接に関係し，その意味で Schur 安定行列と呼ばれる．

なお，行列 M の固有値がすべて求まれば，その行列が Schur 安定行列であるかどうかは判定できるが，行列の固有値を求めなくとも，行列 M が Schur 安定行列であるかどうかを調べる方法がある．詳しくは 7.1 節を参照せよ．

なお，方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の解が一意でない場合は，任意の初期ベクトル $\boldsymbol{x}[0] \in \mathbb{R}^N$ に対して反復法 (6.6) が収束するからといって，$\rho(M) < 1$ と結論づけることはできない．なぜなら，例えば $M = I, \boldsymbol{v} = 0$ の場合，すなわち，方程式 (6.1) で $A = 0, \boldsymbol{b} = \boldsymbol{0}$ の場合，反復法は任意の初期ベクトルに対して明らかに収束するが，$\rho(M) = 1$ である．

先に述べた Jacobi 法，Gauss-Seidel 法，そして SOR 法に関しては，収束するかどうかを調べる非常に便利な定理がある．証明は [46] などを参照のこと．

定理 16 (Jacobi 法・Gauss-Seidel 法の収束条件) 方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の行列 A が対角優位行列[*5]ならば，

1. $\rho(-D^{-1}(E + F)) < 1$ が成り立つ．すなわち，Jacobi 法は，任意の初期ベクトルに対して唯一つの厳密解に収束する．
2. $\rho(-(D + E)^{-1}F) < 1$ が成り立つ．すなわち，Gauss-Seidel 法は，任意の初期ベクトルに対して唯一つの厳密解に収束する．

定理 17 (Gauss-Seidel 法の収束条件) 方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の行列 A およびその対角成分行列 D が正定値対称行列[*6]ならば，

$$\rho(-(D + E)^{-1}F) < 1$$

が成り立つ．すなわち，Gauss-Seidel 法は，任意の初期ベクトルに対して唯一つの厳密解に収束する．

定理 18 (SOR 法の収束条件) 方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の行列 A および対角成分行列 D が正定値対称行列であり，さらに $0 < \omega < 2$ かつ $(D + \omega E)$ が正則ならば，

$$\rho((I + \omega D^{-1}E)^{-1}(1 - \omega)I - \omega D^{-1}F) < 1$$

が成り立つ．すなわち，SOR 法は，任意の初期ベクトルに対して唯一つの厳密解に収束する．

[*4] 定理 15 において，$\boldsymbol{v} = \boldsymbol{0}$ の場合に相当する．

[*5] 対角優位行列の定義は付録 A.2.6 節を参照のこと．

[*6] 正定値対称行列の定義は付録 A.2.9 節を参照のこと．

例題 8 次の方程式を考える．

$$\underbrace{\begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{\boldsymbol{b}}.$$

行列 A は対角優位行列であるので，定理 16 より，この方程式に対する Jacobi 法および Gauss-Seidel 法による反復法は任意の初期ベクトルに対して厳密解 $\boldsymbol{x}^* = A^{-1}\boldsymbol{b}$ に収束する．

例題 9 次の方程式を考える．

$$\underbrace{\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{\boldsymbol{b}}.$$

行列 A は正定値対称行列であることを付録の定理 38 を用いて示す．まず行列 A を次のように分解する．

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{bmatrix}, \quad A_{11} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad A_{22} = 3$$

部分行列 A_{11} に関して

$$1 > 0, \quad 1 \times 2 - (-1) \times (-1) = 1 > 0$$

より，行列 A_{11} は正定値行列であり，さらに

$$A_{22} - A_{12}^\top A_{11}^{-1} A_{12} = 3 > 0$$

であるので，定理 38 より行列 A は正定値対称行列であることがわかる．また，行列 A の対角成分はすべて正であるので，対角成分行列 D も正定値対称行列となる．したがって，定理 17 より，上の方程式 $A\boldsymbol{x} = \boldsymbol{b}$ に対して，Gauss-Seidel 法による反復法は任意の初期ベクトルに対して厳密解 $\boldsymbol{x}^* = A^{-1}\boldsymbol{b}$ に収束する．

練習問題 26 定理 16 と定理 17 を用いて，次の方程式に対する Jacobi 法および Gauss-Seidel 法による反復法の収束性を調べよ．

$$(1) \quad \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$(2) \quad \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

6.4　共役勾配法

スライドファイル　lecture9.tex　を参照せよ

6.5　A が正則でない場合

- 射影
- proximity operator ?

6.6　誤差の影響

前節までに考えた反復法を計算機で行う場合，必ず数値の丸めによる誤差が生じる．ここでは，線形方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の反復解法における丸め誤差の影響を考察する．

まず，線形方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の係数 A と $\boldsymbol{b}$ に丸め誤差が生じた場合に，その方程式の解はどれだけの誤差を含むかを調べる．行列 A およびベクトル $\boldsymbol{b}$ の丸め誤差をそれぞれ ΔA，$\Delta\boldsymbol{b}$ とおく．丸められた方程式 $(A + \Delta A)\boldsymbol{x} = \boldsymbol{b} + \Delta\boldsymbol{b}$ の厳密解と，もとの方程式 $A\boldsymbol{x} = \boldsymbol{b}$ の厳密解との差を $\Delta\boldsymbol{x}$ とおく．すなわち，

$$(A + \Delta A)(\boldsymbol{x}^* + \Delta\boldsymbol{x}^*) = \boldsymbol{b} + \Delta\boldsymbol{b}$$

が成り立つとする．このとき相対誤差 $\|\Delta\boldsymbol{x}^*\|/\|\boldsymbol{x}^*\|$ に関して次の定理が成り立つ．

定理 19 行列 A は正則であるとし，さらに $\boldsymbol{b} \neq 0$ と仮定する．また行列 A に対する丸め誤差 ΔA は十分小さく，

$$\|\Delta A\| < \frac{1}{\|A^{-1}\|}$$

が成立しているとする．このとき次の不等式が成り立つ．

$$\frac{\|\Delta\boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} \leq \frac{k(A)}{1 - k(A)\|\Delta A\|\,\|A\|^{-1}} \left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} \right).$$

ここで $k(A)$ は行列 A の条件数[*7]であり，次式で定義される．

$$k(A) := \|A\|\|A^{-1}\|.$$

[*7] 条件数の定義は，付録 A.2.6 節を参照のこと．

証明

$A\boldsymbol{x}^* = \boldsymbol{b}$ および $(A+\Delta A)(\boldsymbol{x}^*+\Delta\boldsymbol{x}^*) = \boldsymbol{b}+\Delta\boldsymbol{b}$ より，

$$(A+\Delta A)\Delta\boldsymbol{x}^* = -\Delta A\boldsymbol{x}^* + \Delta\boldsymbol{b}$$

が成り立つ．行列 A は正則であるので，両辺に左から A^{-1} を掛けて

$$(I+A^{-1}\Delta A)\Delta\boldsymbol{x}^* = A^{-1}(-\Delta A\boldsymbol{x}^* + \Delta\boldsymbol{b})$$

を得る．ここで $A^{-1}\Delta A$ に関して，仮定より

$$\|A^{-1}\Delta A\| \le \|A^{-1}\|\,\|\Delta A\| < 1$$

が成り立つので，$(I+A^{-1}\Delta A)$ は正則であることがわかる[*8]．したがって，

$$\Delta\boldsymbol{x}^* = (I+A^{-1}\Delta A)^{-1}A^{-1}(-\Delta A\boldsymbol{x}^* + \Delta\boldsymbol{b})$$

が成り立つ．これより，

$$\|\Delta\boldsymbol{x}^*\| \le \|(I+A^{-1}\Delta A)^{-1}\|\,\|A^{-1}\|\left(\|\Delta A\| + \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{x}^*\|}\right)\|\boldsymbol{x}^*\|$$

となることがわかる．ここで，

$$\|(I+A^{-1}\Delta A)^{-1}\| \le \frac{1}{1-\|A^{-1}\Delta A\|} \le \frac{1}{1-\|A^{-1}\|\,\|\Delta A\|}$$

が成り立ち，また $A\boldsymbol{x}^* = \boldsymbol{b}$ より $\|\boldsymbol{b}\| \le \|A\|\|\boldsymbol{x}^*\|$ であり，さらに $\boldsymbol{b} \neq \boldsymbol{0}$ より，$\boldsymbol{x}^* = A^{-1}\boldsymbol{b} \neq \boldsymbol{0}$ となる．したがって，

$$\frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{x}^*\|} \le \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}\|A\|$$

が得られる．以上より次の不等式が成り立つことがわかる．

$$\begin{aligned}
\frac{\|\Delta\boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|} &\le \frac{\|A^{-1}\|}{1-\|A^{-1}\|\,\|\Delta A\|}\left(\|\Delta A\| + \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{x}^*\|}\right) \\
&\le \frac{\|A^{-1}\|\,\|A\|}{1-\|A\|\,\|A^{-1}\|\cdot\frac{\|\Delta A\|}{\|A\|}}\frac{1}{\|A\|}\left(\|\Delta A\| + \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}\|A\|\right) \\
&= \frac{k(A)}{1-k(A)\|\Delta A\|\,\|A\|^{-1}}\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}\right).
\end{aligned}$$

【証明終】

[*8] 第 5 章の例題 5 を見よ．また，付録の補題 11 も参照せよ．

この定理より，条件数 $k(A)$ が大きければ，相対誤差 $\|\Delta \boldsymbol{x}^*\|/\|\boldsymbol{x}^*\|$ の上界が大きくなることがわかる．すなわち，$k(A)$ が大きければ，丸め誤差 ΔA や $\Delta \boldsymbol{b}$ が小さくても，解の誤差が大きくなる可能性を示唆している．この性質により，一般に条件数が大きな行列に対する線形方程式は数値計算でも解くのが難しい問題である．また，行列 A が Hermite **行列** (Hermitian matrix)，すなわち $A = \bar{A}^\top$ のとき，条件数は

$$k(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\max_i |\lambda_i(A)|}{\min_i |\lambda_i(A)|}$$

で与えられる．すなわち，行列 A の最大固有値と最小固有値が大きく離れている場合，線形方程式は解きにくくなることがわかる．

次に反復法における丸め誤差の影響を考える．丸めがある場合，反復法では繰り返しによって各ステップでの丸め誤差が累積していく．これは**誤差の伝播** (propagation of error) と呼ばれる問題であり，数値計算においてその誤差の見積もりは非常に重要である．繰り返しによる誤差の見積もりは一般に難しい問題であるが，丸め誤差が常にある一定値で抑えられると仮定すると議論はいくぶん簡単になる．

線形方程式の反復法

$$\boldsymbol{x}[n+1] = M\boldsymbol{x}[n] + \boldsymbol{v}, \quad n = 0, 1, 2, \ldots$$

を考える．この反復法において，各ステップに丸め誤差 $\boldsymbol{d}[n]$ が生じるものとする．すなわち，丸め誤差を含んだ反復法を次のように定式化する．

$$\widetilde{\boldsymbol{x}}[n+1] = M\widetilde{\boldsymbol{x}}[n] + \boldsymbol{v} + \boldsymbol{d}[n], \quad n = 0, 1, 2, \ldots \tag{6.7}$$

ここで $\{\widetilde{\boldsymbol{x}}[n]\}$ は丸め誤差の影響を受けたベクトル列である．また丸め誤差 $\boldsymbol{d}[n]$ に関して，ある $\delta > 0$ が存在して

$$\sup_{n \geq 0} \|\boldsymbol{d}[n]\| \leq \delta \tag{6.8}$$

が成り立つと仮定する．このとき次の定理が成り立つ．

定理 20 (反復法の誤差評価) 丸め誤差の影響を受けたベクトル列 $\{\widetilde{\boldsymbol{x}}[n]\}$ と真の解 $\boldsymbol{x}^*$ との差を $\{\boldsymbol{e}[n]\}$ とおく．すなわち，

$$\boldsymbol{e}[n] = \widetilde{\boldsymbol{x}}[n] - \boldsymbol{x}^*, \quad n = 0, 1, 2, \ldots$$

とおく．このとき次の不等式が成り立つ．

$$\|\boldsymbol{e}[n]\| \leq \|M^n \boldsymbol{e}[0]\| + \frac{1 - \|M\|^n}{1 - \|M\|}\delta, \quad n = 0, 1, 2, \ldots \tag{6.9}$$

また $\rho(M) < 1$ ならば，$\mathbb{R}^N$ のあるノルム $\|\cdot\|_\alpha$ とそれが誘導する行列のノルム $\|\cdot\|_\alpha$ が存在して，

$$\lim_{n\to\infty} \|\boldsymbol{e}[n]\|_\alpha \leq \frac{\delta}{1 - \|M\|_\alpha} \tag{6.10}$$

が成り立つ．

証明

まず，$\boldsymbol{x}^* = M\boldsymbol{x}^* + \boldsymbol{v}$ が成り立つので，(6.7) より $n = 0, 1, 2, \ldots$ に対して，

$$\begin{aligned}
\widetilde{\boldsymbol{x}}[n+1] - \boldsymbol{x}^* &= M\widetilde{\boldsymbol{x}}[n] - \boldsymbol{x}^* + \boldsymbol{v} + \boldsymbol{d}[n] \\
&= M(\widetilde{\boldsymbol{x}}[n] - \boldsymbol{x}^*) + (M\boldsymbol{x}^* + \boldsymbol{v}) - \boldsymbol{x}^* + \boldsymbol{d}[n] \\
&= M(\widetilde{\boldsymbol{x}}[n] - \boldsymbol{x}^*) + \boldsymbol{d}[n]
\end{aligned}$$

が成り立つ．これより，$\boldsymbol{e}[n+1] = M\boldsymbol{e}[n] + \boldsymbol{d}[n]$ が成り立つことがわかる．この式に $n = 0, 1, 2, \ldots$ を代入していくと帰納法により

$$\boldsymbol{e}[n] = M^n\boldsymbol{e}[0] + \sum_{k=0}^{n-1} M^{n-k-1}\boldsymbol{d}[k], \quad n = 0, 1, 2, \ldots$$

が得られる．ノルムの性質，および (6.8) より，

$$\begin{aligned}
\|\boldsymbol{e}[n]\| &\leq \|M^n\boldsymbol{e}[0]\| + \sum_{k=0}^{n-1} \|M\|^{n-k-1}\|\boldsymbol{d}[n]\| \\
&\leq \|M^n\boldsymbol{e}[0]\| + \sum_{k=0}^{n-1} \|M\|^{n-k-1}\delta \\
&= \|M^n\boldsymbol{e}[0]\| + \frac{1 - \|M\|^n}{1 - \|M\|}\delta
\end{aligned}$$

が成り立つ．また，$\rho(M) < 1$ と仮定すると，任意の $\boldsymbol{e}[0] \in \mathbb{R}^N$ に対して，

$$\|M^n\boldsymbol{e}[0]\| \leq |\rho(M)|^n\|\boldsymbol{e}[0]\| \to \boldsymbol{0}, \quad n \to \infty$$

が成り立つ．さらに，$\rho(M) < 1$ より，あるベクトルノルム $\|\cdot\|_\alpha$ とそれが誘導する行列ノルム $\|\cdot\|_\alpha$ が存在して，$\|M\|_\alpha < 1$ が成り立つ[*9]．また上の証明は，$\mathbb{R}^N$ の任意のノルムについても言えるので，ノルム $\|\cdot\|_\alpha$ に対しても，(6.9) が成り立つことがわかる．したがって，ノルムを $\|\cdot\|_\alpha$ としたときの誤差の不等式 (6.9) において $n \to \infty$ とすることにより，(6.10) の不等式が得られる．

【証明終】

[*9] 付録の定理 34 を参照のこと．

この定理より，丸め誤差が有界であれば，収束する反復法から生成される近似解も有界となることがわかる．このような性質を Lagrange **安定性** (Lagrange stability) と呼ぶ．また，丸め誤差を反復法への入力と考えると，この性質は**有界入力-有界出力安定性** (bounded-input bounded-output stability)，または略して BIBO **安定性** (BIBO stability) と呼ばれるものである．

6.7　線形方程式に対する反復法とブロック線図表現

ここでは線形方程式に対する反復法をブロック線図で表現することを考える．まず，Jacobi 法を次のように変形する．

$$\begin{aligned}\boldsymbol{x}[n+1] &= -D^{-1}(E+F)\boldsymbol{x}[n] + D^{-1}\boldsymbol{b} \\ &= \boldsymbol{x}[n] - D^{-1}(D+E+F)\boldsymbol{x}[n] + D^{-1}\boldsymbol{b} \\ &= \boldsymbol{x}[n] - D^{-1}(A\boldsymbol{x}[n] - \boldsymbol{b})\end{aligned} \tag{6.11}$$

ここで $A = D + E + F$ を使った．同様にして，Gauss-Seidel 法および SOR 法は以下のように変形できる．

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - (D+E)^{-1}(A\boldsymbol{x}[n] - \boldsymbol{b}), \quad \text{(Gauss-Seidel)}$$
$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - \omega(D+\omega E)^{-1}(A\boldsymbol{x}[n] - \boldsymbol{b}), \quad \text{(SOR)}$$

以上より，Jacobi 法，Gauss-Seidel 法，SOR 法の反復法はすべて次の形式で記述できることがわかる．

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] + P(A\boldsymbol{x}[n] - \boldsymbol{b}). \tag{6.12}$$

また，最急降下法では，

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] - \alpha[n](A\boldsymbol{x}[n] - \boldsymbol{b}).$$

となる．

反復法 (6.12) のブロック線図表現を図 6.1 に示す．このブロック線図において，Jacobi 法の場合は $P = -D^{-1}$，Gauss-Seidel 法の場合は $P = -(D+E)^{-1}$，SOR 法の場合は $P = -\omega(D+\omega E)^{-1}$ である．

このブロック線図においては，Newton 法のブロック線図と同じように（4.4 を参照せよ），加算器が含まれていることに注意すること．フィードバックシステムの内部に存在する加算器は，一定値の入力 $\boldsymbol{b}$ に対する誤差 $A\boldsymbol{x}[n] - \boldsymbol{b}$ を漸近的にゼロにする働きがある．これは，一定値の信号

$$\boldsymbol{b}, \boldsymbol{b}, \boldsymbol{b}, \ldots$$

のモデルが加算器で表されることに対応し，**内部モデル原理** (internal model principle) と呼ばれる，制御工学では非常に重要な性質である [28]．線形方程式に対する反復法のこの

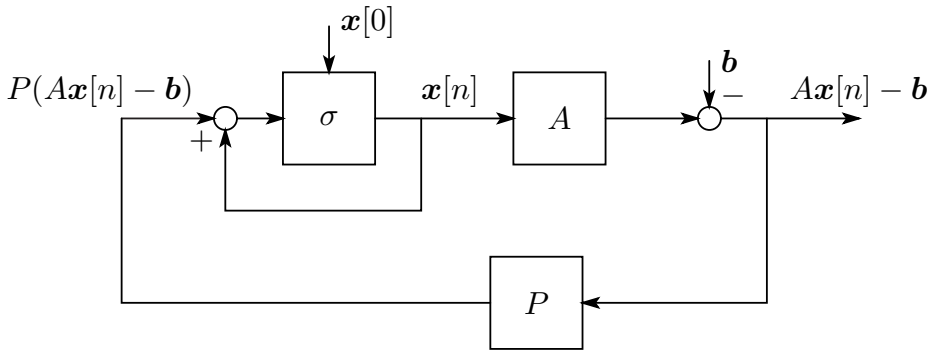


図 6.1　反復法 $\boldsymbol{x}[n+1]=\boldsymbol{x}[n]+P(A\boldsymbol{x}[n]-\boldsymbol{b})$ のブロック線図表現

ような性質も，ブロック線図を描くことにより明快になる．数値計算において，ブロック線図を使う大きな利点の一つである．

練習問題 27 式 (6.11) を参考にして，Gauss-Seidel 法では $P=-(D+E)^{-1}$，SOR 法では $P=-\omega(D+\omega E)^{-1}$ であることを示せ．

練習問題 28 図 6.1 を参考にして，Scilab (Xcos) を用いて，Jacobi 法，Gauss-Seidel 法，SOR 法のシミュレーションを実行せよ．

6.8　さらに勉強するために

線形方程式の反復解法に関しては，[46, 49] を主に参照した．線形方程式に対する反復解法は，数値計算の中でも非常に重要なトピックであり，[45, 22] などが参考になる．また，線形方程式の反復解法をブロック線図で表すアイデアは，[14, 1] で提案されている．大規模な線形方程式の数値解法については，[42, 40] に入門的な解説がある．

第 7 章

行列の固有値問題

行列の**固有値問題** (eigenvalue problem) は工学において非常に重要な問題である．例えば制御工学では，制御対象を安定に動作させるために，ある行列の固有値を移動させるという操作を行う．また，大規模なシステムの解析には，固有値を一般化した特異値と呼ばれる量が重要となる．ここでは，この固有値問題の数値計算について述べる．

ここで述べる固有値問題は以下の 2 つである．

1. 全ての固有値の存在する領域を求める．
2. 全ての固有値を求める．

7.1 固有値の存在する領域

前章でみたように，線形方程式の反復法では行列 M の固有値が複素平面上の単位円内に存在するかどうか，すなわち $\rho(M) < 1$ かどうかが反復法の収束性に重要な役割を果たした．また，例えば，次の微分方程式

$$\frac{d\boldsymbol{x}(t)}{dt} = A\boldsymbol{x}(t), \quad \boldsymbol{x}(0) \in \mathbb{R}^N$$

が**漸近安定** (asymptotically stable) である[*1]ための必要十分条件は，行列 A の固有値の実部が負，すなわち

$$\mathrm{Re}\ \lambda_i(A) < 0, \quad i = 1, 2, \ldots, N$$

が成り立つことである．これらの例でわかるように，行列の固有値がどこに存在するかを知ることが必要な場面は多い．

次は，行列の固有値の存在領域に関する定理の一つである．

[*1] 漸近安定とは，任意の初期値 $\boldsymbol{x}(0) \in \mathbb{R}^N$ に対して，微分方程式の解が $\lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{0}$ となることである．

定理 21 A を正方行列，$\|\cdot\|$ を行列ノルムとすると

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

が成り立つ．すなわち行列 A の固有値は全て，原点を中心とした半径 $\|A\|$ の円の内部（または境界）に存在する．

この定理は，行列のノルムを計算すればよいだけであるので単純ではあるが，領域が行列のノルムの取り方に依存するので，ノルムの選び方によっては領域を広く取りすぎることがある．このような任意性をなくし，より詳細に領域を求めたい場合は，Gershgorin **の定理** (Gershgorin theorem) が役に立つ．まず次の補題を示す．

補題 6 複素数を要素として持つ行列 $A = [a_{ij}] \in \mathbb{C}^{N\times N}$ を考える．このとき次が成り立つ．

$$\sigma(A) \subseteq S_R = \bigcup_{i=1}^{N} R_i, \quad R_i := \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j\neq i}}^{N} |a_{ij}| \right\} \tag{7.1}$$

ここで $\sigma(A)$ は行列 A の全ての固有値の集合とする．すなわち A の固有値を $\lambda_1, \lambda_2, \ldots, \lambda_N$ とすると，$\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ である．

証明

複素数 λ を A の固有値とする．すなわち，$\lambda \in \sigma(A)$ とする．この λ が $\lambda \in S_R$ を満たすことを示す．まず，行列 A の固有値 λ が対角成分 $a_{11}, a_{22}, \ldots, a_{NN}$ のどれかと等しい場合，明らかに $\lambda \in S_R$ である．次に，λ は A の対角成分のどれとも等しくないとする．すなわち，$\lambda \neq a_{ii}$ $(i = 1, \ldots, N)$ とする．行列 A を対角成分と非対角成分に次のように分解する．

$$A = D + E$$

$$D = \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & a_{NN} \end{bmatrix}, \quad E = \begin{bmatrix} 0 & a_{12} & \ldots & a_{1N} \\ a_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{N-1,N} \\ a_{N1} & \ldots & a_{N,N-1} & 0 \end{bmatrix}.$$

行列 $B_\lambda = A - \lambda I = (D - \lambda I) + E$ を考えると，複素数 λ は A の固有値だから，$B_\lambda = A - \lambda I$ は正則ではない．したがって，ある $\mathbf{0}$ でないベクトル $\boldsymbol{x} \in \mathbb{C}^N$ が存在して，$B_\lambda \boldsymbol{x} = \mathbf{0}$ が成り立つ．これより，$\{(D - \lambda I) + E\}\boldsymbol{x} = \mathbf{0}$ となる．また，$\lambda \neq a_{ii}$ $(i = 1, \ldots, N)$ より，$D - \lambda I$ は正則である．したがって，

$$\boldsymbol{x} = -(D - \lambda I)^{-1} E \boldsymbol{x}$$

と書け，これより，ノルムに関する不等式

$$\|\boldsymbol{x}\|_\infty \le \|(D-\lambda I)^{-1}E\|_\infty \|\boldsymbol{x}\|_\infty \tag{7.2}$$

が成り立つことがわかる．ここでベクトル $\boldsymbol{x} = [x_1, \ldots, x_N]^\top \in \mathbb{R}^N$ に対して

$$\|\boldsymbol{x}\|_\infty = \max_{i=1,\ldots,N} |x_i|$$

であり，また行列 $M = [m_{ij}] \in \mathbb{R}^{N\times N}$ に対して

$$\|M\|_\infty = \sup_{\boldsymbol{x}\in\mathbb{R}^N, \boldsymbol{x}\neq\boldsymbol{0}} \frac{\|M\boldsymbol{x}\|_\infty}{\|\boldsymbol{x}\|_\infty} = \max_{i=1,\ldots,N} \sum_{j=1}^{N} |m_{ij}|$$

である．ベクトル $\boldsymbol{x}$ は $\boldsymbol{0}$ ではないので，$\|\boldsymbol{x}\|_\infty > 0$ ので，不等式 (7.2) の両辺を $\|\boldsymbol{x}\|_\infty$ で割ると

$$1 \le \|(D-\lambda I)^{-1}E\|_\infty = \max_{i=1,\ldots,N} \sum_{\substack{j=1\\ j\neq i}}^{N} \frac{1}{|a_{ii}-\lambda|}\ |a_{ij}|$$

が得られる．これよりある番号 $i \in \{1, 2, \ldots, N\}$ が存在して，

$$1 \le \frac{1}{|a_{ii}-\lambda|} \sum_{\substack{j=1\\ j\neq i}}^{N} |a_{ij}|$$

すなわち，

$$|a_{ii}-\lambda| \le \sum_{\substack{j=1\\ j\neq i}}^{N} |a_{ij}|$$

が成り立つ．これは $\lambda \in R_i$ であることを示している．ゆえに $\lambda \in S_R$ である．

【証明終】

行列 A の固有値とその転置行列 A^T の固有値は等しいことより，次の系が得られる．

系 1 複素数を要素として持つ行列 $A = [a_{ij}] \in \mathbb{C}^{N\times N}$ を考える．このとき次が成り立つ．

$$\sigma(A) \subseteq S_C = \bigcup_{j=1}^{N} C_j, \quad C_j := \left\{ z \in \mathbb{C} : |z-a_{jj}| \le \sum_{\substack{i=1\\ i\neq j}}^{N} |a_{ij}| \right\} \tag{7.3}$$

補題 6 の R_i および系 1 の C_j は Gershgorin **の円** (Gershgorin circle) と呼ばれる．補題 6 と系 1 より次の定理が得られる．

定理 22 (Gershgorin の定理) 行列 $A \in \mathbb{C}^{N\times N}$ に対して,

$$\sigma(A) \subseteq S_R \cap S_C$$

が成り立つ．すなわち，行列 A の固有値はすべて領域 $S_R \cap S_C$ に含まれる．

例題 10 行列 A を

$$A = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}$$

とする．この行列に対する Gershgorin の円を求め，固有値の存在領域を図示する．まず，補題 6 より，二つの円

$$R_1 = \{z \in \mathbb{C} : |z| \le 1\}, \quad R_2 = \{z \in \mathbb{C} : |z-3| \le 2\}$$

が得られ，また系 1 より，もう二つの円

$$C_1 = \{z \in \mathbb{C} : |z| \le 2\}, \quad C_2 = \{z \in \mathbb{C} : |z-3| \le 1\},$$

が得られる．したがって，定理 22 より，行列 A の固有値が存在する領域

$$S_R \cap S_C = (R_1 \cup R_2) \cap (C_1 \cup C_2)$$

が図 7.1 のように図示できる．なお，行列 A の固有値は 1 と 2 であり，確かに領域 $S_R \cap S_C$ に含まれていることがわかる．

練習問題 29 Gershgorin の定理を用いて，次の行列の固有値の存在する領域を図示せよ．

$$(1)\ A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$(2)\ A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$(3)\ A = \begin{bmatrix} 1+j & j & j \\ 0 & j & 1+j \\ 0 & 0 & -1 \end{bmatrix}$$

Gershgorin の定理は行列の要素だけから固有値の存在する領域を図示することができ，固有値がおおまかにどこに存在するかを見るためには便利であるが，例題からもわかるように，その領域は広くなりすぎる傾向がある．たとえば，冒頭で述べた微分方程式の安定性を行列 A に対する Gershgorin の定理を用いて判定することを考えよう．領域 $S_R \cap S_C$ が複素平面上の領域

$$\mathbb{C}_- := \{z \in \mathbb{C} : \mathrm{Re}\ z < 0\} \tag{7.4}$$

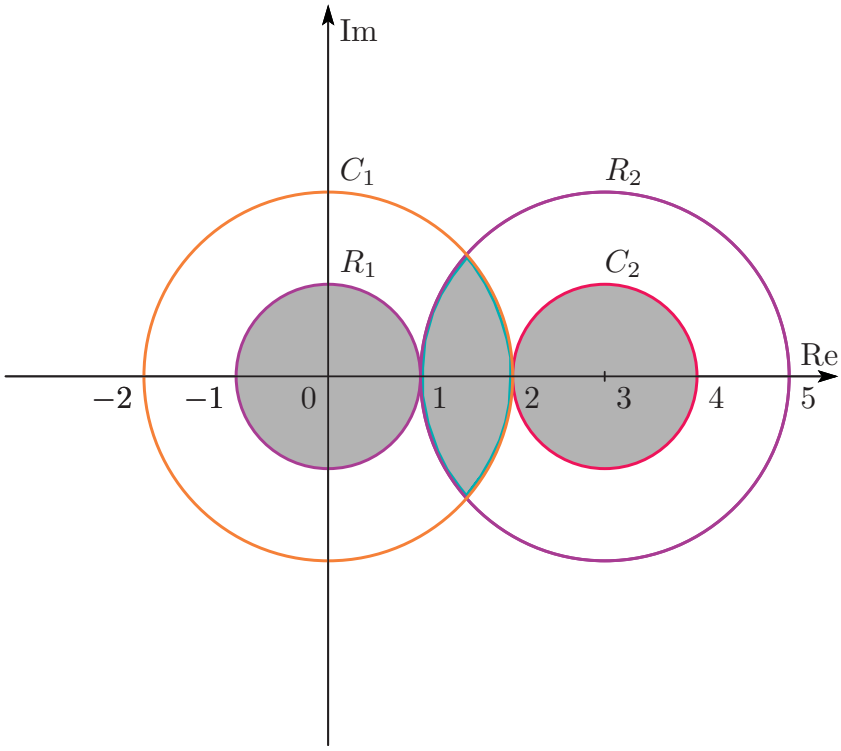


図 7.1　Gershgorin の円と固有値の存在領域 $S_R \cap S_C$

に完全に含まれていれば，微分方程式は漸近安定であると結論付けることができる．同様に，$S_R \cap S_C$ が領域

$$\mathbb{C}_+ := \{z \in \mathbb{C} : \mathrm{Re}\ z \geq 0\}$$

に完全に含まれていれば，微分方程式は安定ではないと判定できる．しかし，領域 $S_R \cap S_C$ が，例題 10 のように領域 $\mathbb{C}_-$ と $\mathbb{C}_+$ の両方にまたがる場合，微分方程式が漸近安定化どうかを判定することはできない．

このような場合は，次の Lyapunov の定理を使うことができる．

定理 23 (Lyapunov の安定定理 1) 行列 $A \in \mathbb{R}^{N\times N}$ の固有値が (7.4) で定義される領域 $\mathbb{C}_-$ に存在するための必要十分条件は，ある正定値対称行列 $P > 0$ が存在して

$$A^\top P + PA < 0 \tag{7.5}$$

が成り立つことである．

系 2 微分方程式

$$\frac{d\boldsymbol{x}(t)}{dt} = A\boldsymbol{x}(t), \quad \boldsymbol{x}(0) \in \mathbb{R}^N$$

の解が任意の初期値 $\boldsymbol{x}(0) \in \mathbb{R}^N$ に対して $\lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{0}$ となるための必要十分条件は，ある正定値対称行列 $P > 0$ が存在して，不等式 (7.5) が成り立つことである．

この定理の良い点は，不等式 (7.5) を満たす正定値対称行列 P を求めるための効率の良いアルゴリズムが使える点である．(7.5) のような行列変数 P が線形で依存する不等式は，線形行列不等式と呼ばれ，特に制御工学の分野でよく使われている．`Scilab` では，`lmisolver` や `lmitool` のルーチンで線形行列不等式を解くことができる．

また第 6 で述べた線形方程式のための反復法の収束性を調べる定理 15 の必要十分条件，すなわち行列 $M \in \mathbb{R}^{N\times N}$ に対して $\rho(M) < 1$ が成り立つかどうかを判定するのにも，同様の定理が成り立つ．

定理 24 (Lyapunov の安定定理 2) 行列 $M \in \mathbb{R}^{N\times N}$ の固有値が領域

$$\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$$

に存在するための必要十分条件は，ある正定値対称行列 $Q > 0$ が存在して

$$M^\top QM - Q < 0$$

が成り立つことである．

この定理の不等式も線形行列不等式であり，判定は計算機を使えば容易に行える．

7.2　固有値の数値計算

本節では，行列 $A \in \mathbb{C}^{N\times N}$ の固有値を全て求める計算法を調べる．行列の固有値は，固有方程式の根を求めることで得られるが，N が 5 以上の場合，一般に厳密解を求めることは不可能である．したがって，固有値を求めるためには数値計算を必要とする．固有値の近似計算には様々な手法があるが， 本節で述べる QR 分解法が最もポピュラーである．まず，行列の QR **分解** (QR factorization) について述べる．

定理 25 (行列の QR 分解) 行列 $A \in \mathbb{C}^{N\times N}$ を正則とする．このとき，次のような分解が一意的に可能である．

$$A = QR$$

ここで Q はユニタリ行列（すなわち，$Q^*Q = QQ^* = I$ が成り立つ行列），R は対角成分が正である右上三角行列である．

証明

行列 A を次のように列ベクトルに分解する．

$$A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_N \end{bmatrix}, \quad \boldsymbol{a}_i \in \mathbb{C}^N, \quad i = 1, 2, \dots, N.$$

A は正則だから $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ は一次独立となる．この N 本の一次独立な列ベクトルを使って，N 本の列ベクトル $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N\}$ を次の手順で作る．すなわち

$$\boldsymbol{u}_1 := \boldsymbol{a}_1, \quad \boldsymbol{u}_n := \boldsymbol{a}_n - \sum_{i=1}^{n-1} (\boldsymbol{a}_n, \boldsymbol{q}_i)\boldsymbol{q}_i, \quad n = 2, \ldots, N$$

とおき，

$$\boldsymbol{q}_i := \frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}, \quad i = 1, 2, \ldots, N$$

とする[*2]．ただし，$(\boldsymbol{a}, \boldsymbol{q})$ はベクトル $\boldsymbol{a}$ と $\boldsymbol{q}$ の標準内積 $\boldsymbol{q}^*\boldsymbol{a}$ を表し[*3]，$\|\boldsymbol{u}\| = \sqrt{(\boldsymbol{u}, \boldsymbol{u})}$ とする．このとき，N 本の一次独立なベクトル $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N\}$ は $\mathbb{C}^N$ の正規直交基底となる．すなわち $i, j = 1, 2, \ldots, N$ に対して

$$(\boldsymbol{q}_i, \boldsymbol{q}_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{7.6}$$

が成り立つ．これを用いて，

$$Q = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_N \end{bmatrix}$$

とおくと，(7.6) より Q はユニタリ行列となる．次に右上三角行列 $R = [r_{ij}]$ を

$$r_{ij} = \begin{cases} (\boldsymbol{a}_j, \boldsymbol{q}_i), & i < j \\ \left\| \boldsymbol{a}_j - \sum_{k=1}^{j-1} (\boldsymbol{a}_j, \boldsymbol{q}_k)\boldsymbol{q}_k \right\|, & i = j \\ 0, & i > j \end{cases}$$

によって作る．このとき，

$$\boldsymbol{u}_j = \boldsymbol{a}_j - \sum_{i=1}^{j-1} (\boldsymbol{a}_j, \boldsymbol{q}_i)\boldsymbol{q}_i = \boldsymbol{a}_j - \sum_{i=1}^{j-1} r_{ij}\boldsymbol{q}_i \tag{7.7}$$

と書ける．また $r_{jj} = \|\boldsymbol{u}_j\|$ であり，$\boldsymbol{q}_j = \boldsymbol{u}_j / \|\boldsymbol{u}_j\|$ より，

$$\boldsymbol{u}_j = \|\boldsymbol{u}_j\|\boldsymbol{q}_j = r_{jj}\boldsymbol{q}_j$$

が成り立つことがわかる．これと (7.7) より，

$$\boldsymbol{a}_j = \boldsymbol{u}_j + \sum_{i=1}^{j-1} r_{ij}\boldsymbol{q}_i = r_{jj}\boldsymbol{q}_j + \sum_{i=1}^{j-1} r_{ij}\boldsymbol{q}_j = \sum_{i=1}^{j} r_{ij}\boldsymbol{q}_j$$

[*2] これは Gram-Schmidt の直交化である．付録の A.2.10 を参照せよ

[*3] ベクトル $\boldsymbol{q} \in \mathbb{C}^N$ に対して，$\boldsymbol{q}^*$ は複素共役転置を表す．

が成り立つ．ここで，$i > j$ で $r_{ij} = 0$ であるから，

$$\boldsymbol{a}_j = \sum_{i=1}^{N} r_{ij}\boldsymbol{q}_j, \quad j = 1, 2, \ldots, N$$

が成り立つことがわかる．これより，

$$\boldsymbol{a}_j = \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \ldots & \boldsymbol{q}_N\end{bmatrix} \begin{bmatrix} r_{1j} \\ r_{2j} \\ \vdots \\ r_{Nj} \end{bmatrix}$$

と書ける．したがって，

$$\begin{aligned} A &= \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \ldots & \boldsymbol{a}_N\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_N\end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_N\end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ 0 & r_{22} & \cdots & r_{2N} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & r_{NN} \end{bmatrix} \\ &= QR \end{aligned}$$

となることがわかる．

次に一意性を証明する．まず正則行列 A が次の 2 通りに QR 分解されたとする．

$$A = Q_1 R_1 = Q_2 R_2. \tag{7.8}$$

ここで，Q_1 と Q_2 はユニタリ行列，R_1 と R_2 は対角成分がすべて正の右上三角行列である．行列 A は正則であるので，R_1 も正則となる．(7.8) の両辺に右から R_1^{-1}，左から $Q_2^{-1} = Q_2^*$ を掛けると，

$$Q_2^* Q_1 = R_2 R_1^{-1} \tag{7.9}$$

が得られる．ここで，

$$(Q_2^* Q_1)^* (Q_2^* Q_1) = Q_1^* Q_2 Q_2^* Q_1 = Q_1^* Q_1 = I$$

より，$Q_2^* Q_1$ はユニタリ行列となる．また，R_1 は上三角行列より，R_1^{-1} も上三角行列となり，R_2 も上三角行列だから，$R_2 R_1^{-1}$ は上三角行列となる．さらに R_1 および R_2 の対角成分は正であるので，$R_2 R_1^{-1}$ の対角成分は正となる．(7.9) の左辺はユニタリ行

列，右辺は対角成分がすべて正の右上三角行列であるから，$Q_2^* Q_1 = R_2 R_1^{-1}$ は対角ユニタリ行列でなければならない．さらに，この行列の対角成分は正であるので，結局，$Q_2^* Q_1 = R_2 R_1^{-1} = I$ となる．すなわち，

$$Q_1 = (Q_2^*)^{-1} = Q_2, \quad R_1 = R_2$$

が成り立つ． 【証明終】

さて，行列 $A \in \mathbb{C}^{N\times N}$ を QR 分解したとする．すなわち，

$$A = QR$$

とする．ここで，左から $Q^* = Q^{-1}$，右から Q を掛けると

$$Q^{-1}AQ = Q^{-1}(QR)Q = RQ$$

が得られる．すなわち行列 Q による A の相似変換は QR 分解して掛け算を入れ替えることで得られることがわかる．これをふまえて，QR **分解法** (QR factorization method) のアルゴリズムは以下のように与えられる．

QR 分解法のアルゴリズム

1. $A_1 := A$
2. $n = 1, 2, \ldots$ に対して，

$$\begin{aligned} A_n &:= Q_n R_n \quad \text{（QR 分解）} \\ A_{n+1} &:= R_n Q_n \quad \text{（掛け算の入れ替え）} \end{aligned}$$

この QR 分解法により行列 A の固有値の近似値が得られる．

定理 26 (固有値の QR 分解法) 行列 $A \in \mathbb{C}^{N\times N}$ を正則行列とし，その固有値 $\lambda_1, \lambda_2, \ldots, \lambda_N$ は全て相異なり，

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_N| > 0$$

を満足しているとする．このとき $A_1 = A$ からはじめる QR 分解法のアルゴリズムによって作られる行列 A_n は，$n \to \infty$ で次の行列に収束する．

$$A_\infty = \begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \lambda_N \end{bmatrix}.$$

証明

[redacted] 【証明終】

この定理より，QR 分解法を繰り返し，行列 A_n の左下の成分がすべてほぼ 0 になれば，その対角成分は行列 A の固有値の近似値となることがわかる．

なお，行列 A が正則でない場合は，適当な実数 c を選んで $A+cI$ を正則行列にしてその固有値の近似値を QR 分解法により求め，その近似値から c を引いた値を求めれば，それらが行列 A の固有値の近似値となる．

練習問題 30 QR 分解法により次の行列の固有値の近似値を求めよ．

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

7.3 行列の特異値とその数値計算法

7.4 Scilab プログラム

- Gershgorin circle
- Lyapunov 不等式
- QR 分解
- SVD

7.5 さらに勉強するために

Gershgorin の定理に関しては，[22, 31] を参考にした．Lyapunov の定理や線形行列不等式については[redacted] を参照せよ．QR 分解法については，[46, 22] を参考にした．また，行列の固有値を求める数値計算としては，QR 分解法のほかにさまざまな方法がある．これらに関しては，[3] が詳しい．また，行列の特異値を求める数値計算については，[40] を参照せよ．

第 8 章

補間多項式と数値積分

この章では，N 組の離散データ $\{(x_n, y_n)\}_{n=1}^{N}$ が与えられたとき，そのデータ点を通る関数を求める方法を調べる．このように N 個の点 (x_n, y_n), $n = 1, 2, \ldots, N$ を必ず通る関数 $y = f(x)$ を求めることを**補間** (interpolation) と呼ぶ．一方，N 個の点 (x_n, y_n), $n = 1, 2, \ldots, N$ のなるべく近くを通る（必ずしもそれらの点を通る必要はない）関数を求めることを**回帰** (regression) と呼び，次章で述べる．

8.1 関数の補間

次のデータが与えられているとしよう．

x	0	1	2	3
y	1	3	3	5

すなわち，

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 2, \quad x_4 = 3$$
$$y_1 = 1, \quad y_2 = 3, \quad y_3 = 3, \quad y_4 = 5$$

である．これらの点を必ず通る関数 $y = f(x)$ を求めたい．ここでは，関数 $f(x)$ は x の多項式とする．異なる 2 点を通る曲線は直線として決まり，異なる 3 点を通る曲線は（存在すれば） 2 次以下の多項式として決まる．一般に N 点が与えられれば，それらを全て通る曲線は（存在すれば） $N - 1$ 次以下の多項式で与えられる．そこで，上のデータを補間する多項式を 3 次多項式とする．すなわち，

$$f(x) = a + bx + cx^2 + dx^3$$

とおく．これに上のデータを代入すれば，

$$\begin{aligned} 1 &= a \\ 3 &= a + b + c + d \\ 3 &= a + 2b + 4c + 8d \\ 5 &= a + 3b + 9c + 27d \end{aligned}$$

となり，係数 a, b, c, d に関する線形連立方程式となる．この連立方程式を解くと，

$$a = 1, \quad b = \frac{13}{3}, \quad c = -3, \quad d = \frac{2}{3}$$

が得られる．したがって求める 3 次多項式関数は

$$y = 1 + \frac{13}{3}x - 3x^2 + \frac{2}{3}x^3$$

となる．

一般に N 個のデータ $\{(x_n, y_n)\}_{n=1}^{N}$ が与えられたとき，これらの点を全て通る $N-1$ 次多項式関数を**補間多項式** (interpolating polynomial) と呼ぶ．一般の補間多項式を求めてみよう．補間多項式を

$$f(x) = a_1 + a_2 x + a_3 x^2 + \cdots + a_N x^{N-1}$$

とおく．全ての点 (x_n, y_n), $n = 1, 2, \ldots, N$ で $f(x_n) = y_n$ が成り立てば良いので，上式にデータを代入すると次の線形連立方程式が得られる．

$$\begin{aligned} y_1 &= a_1 + a_2 x_1 + a_3 x_1^2 + \cdots + a_N x_1^{N-1} \\ y_2 &= a_1 + a_2 x_2 + a_3 x_2^2 + \cdots + a_N x_2^{N-1} \\ &\vdots \\ y_N &= a_1 + a_2 x_N + a_3 x_N^2 + \cdots + a_N x_N^{N-1} \end{aligned} \tag{8.1}$$

ここで，次の行列とベクトルを定義する．

$$M := \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{N-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \ldots & x_N^{N-1} \end{bmatrix}, \quad \boldsymbol{y} := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{a} := \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

これらの行列とベクトルを用いれば，上の線形連立方程式は $\boldsymbol{y} = M\boldsymbol{a}$ と書くことができる．この行列 M を Vandermonde **行列** (Vandermonde matrix) と呼ぶ．Vandermonde 行列には次の性質がある（証明は [10, 11] などを参照せよ）．

補題 7 Vandermonde 行列 M の行列式は以下で与えられる.

$$\det M = \prod_{\substack{i,j=1 \\ i>j}}^{N} (x_i - x_j)$$

この補題より，次の補題が得られる.

補題 8 線形連立方程式 (8.1) の解が存在して唯一つであるための必要十分条件は，$x_1, x_2, \ldots, x_N$ が互いに異なることである.

練習問題 31 補題 7 を用いて，補題 8 を証明せよ.

以上より，x のデータ $x_1, x_2, \ldots, x_N$ が互いに異なっていれば，補間多項式は線形連立方程式 (8.1) を解くことにより求まる.

数値計算においては上記の方法を用いるのがよいが，理論的な考察では次に示す Lagrange **補間公式** (Lagrange interpolation formula) を用いることが多い[*1].

定理 27 (Lagrange 補間公式) N 組のデータ $\{x_n, y_n\}_{n=1}^{N}$ が与えられ，x_1, x_2, …, x_N は互いに異なるとする．このとき，これらの点を全て通る $N-1$ 次多項式関数は以下で与えられる.

$$y = y_1 L_1(x) + y_2 L_2(x) + \cdots + y_N L_N(x). \tag{8.2}$$

ここで，$L_n(x)$ $(n = 1, 2, \ldots, N)$ は次式で与えられる.

$$L_n(x) := \prod_{\substack{k=1 \\ k\neq n}}^{N} \frac{(x - x_k)}{(x_n - x_k)}, \quad n = 1, 2, \ldots, N.$$

(8.2) の多項式を Lagrange **補間多項式** (Lagrange interpolation polynomial) と呼ぶ.

例題 11 ここでは 定理 27 の Lagrange 補間公式を用いて，3 次までの Lagrange 補間多項式を導出する.

[*1] 例えば次節の数値積分の公式の導出などに Lagrange 補間公式が用いられる.

1 次多項式　2 組のデータ $\{(x_1, y_1), (x_2, y_2)\}$ が与えられたとき，この 2 点を通る 1 次関数は Lagrange 補間公式より次式で与えられる．

$$y = y_1 \frac{x - x_2}{x_1 - x_2} + y_2 \frac{x - x_1}{x_2 - x_1}.$$

2 次多項式　3 組のデータ $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ が与えられたとき，この 3 点を通る 2 次関数は Lagrange 補間公式より次式で与えられる．

$$y = y_1 \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}.$$

3 次多項式　4 組のデータ $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$ が与えられたとき，この 4 点を通る 3 次関数は Lagrange 補間公式より次式で与えられる．

$$\begin{aligned} y = {} & y_1 \frac{(x - x_2)(x - x_3)(x - x_4)}{(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)} + y_2 \frac{(x - x_1)(x - x_3)(x - x_4)}{(x_2 - x_1)(x_2 - x_3)(x_2 - x_4)} \\ & + y_3 \frac{(x - x_1)(x - x_2)(x - x_4)}{(x_3 - x_1)(x_3 - x_2)(x_3 - x_4)} + y_4 \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)}. \end{aligned}$$

練習問題 32 次のデータに対する補間多項式を求めよ．

1.

x	0	1	2
y	1	2	3

2.

x	0	1	2	3
y	1	2	3	5

3.

x	0	1	2	3	4
y	1	2	3	5	9

8.2　スプライン補間

8.2.1　スプライン

8.2.2　カーネル法

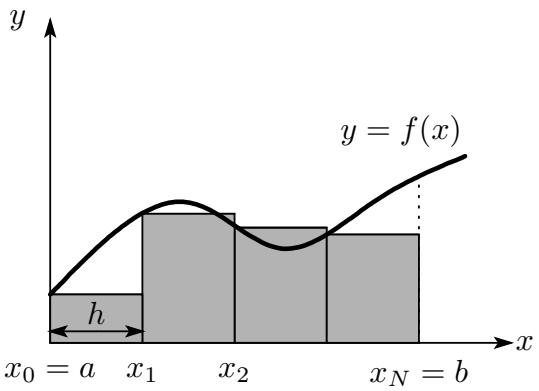


図 8.1 矩形公式

8.3 数値積分

ここでは，前節の補間多項式の応用として，数値積分を考える．区間 $[a,b]$ 上の可積分関数 f が与えられたとき，

$$I = \int_a^b f(x)dx \tag{8.3}$$

を求めたい．しかし，一般に，関数の積分は厳密に求めることは困難である．そこで，関数 f を前節で述べた多項式関数で近似して，f の積分値を近似することを考える．多項式関数はその積分が容易であるので，この近似値は容易に求まる．

8.3.1 矩形公式

一番簡単な近似は，関数 f を 0 次多項式，すなわち区分的に定数の関数で近似する方法である．これを**矩形公式** (rectangle formula) と呼ぶ．図 8.1 のように区間 $[a,b]$ を N 等分し，分割の幅を

$$h = \frac{b-a}{N}$$

とおく．区間 $[a,b]$ 上の各分点を

$$x_n = a + n \cdot \frac{b-a}{N} = a + nh, \quad n = 0, 1, 2, \ldots, N$$

とすると，(8.3) の積分の近似値は図 8.1 の区分的定数関数の積分（矩形の面積の和）で与えられる．これより，矩形公式による積分の近似値は

$$I_0 = \sum_{n=0}^{N-1} f(x_n)h = h \sum_{n=0}^{N-1} f(x_n)$$

と計算できる．

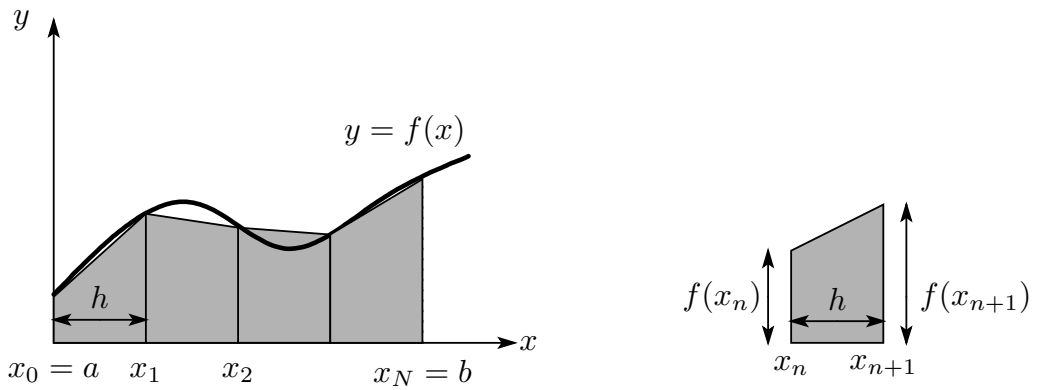


図 8.2　台形公式

8.3.2　台形公式

矩形公式のように区分的定数関数で f を近似するのではなく，区分的 1 次関数で近似する方法を**台形公式** (trapezoidal formula) と呼ぶ．図 8.2 のように区間 $[a,b]$ を N 等分し，矩形公式のときと同じように，$[a,b]$ 上の各分点を

$$x_n = a + n \cdot \frac{b-a}{N} = a + nh, \quad n = 0,1,2,\ldots,N, \quad h = (b-a)/N$$

とおく．すると，(8.3) の積分の近似値は図 8.2 の区分的 1 次関数の積分（台形の面積の和）で与えられる．各台形の面積を $I_1^{(n)}$ とおくと，

$$I_1^{(n)} = \frac{f(x_n) + f(x_{n+1})}{2} h, \quad n = 0,1,2,\ldots,N-1$$

となるので，台形公式による積分値の近似値を I_1 とおくと，

$$I_1 = \sum_{n=0}^{N-1} I_1^{(n)} = \frac{h}{2} \sum_{n=0}^{N-1} \{f(x_n) + f(x_{n+1})\}$$

が得られる．

8.3.3　Simpson の公式

関数 f を区分的 2 次関数で近似して積分の近似値を求める公式を Simpson **の公式** (Simpson formula) と呼ぶ 矩形公式や台形公式と同じように区間 $[a,b]$ を N 等分し，その分点を $x_0(=a), x_1, x_2, \ldots, x_N(=b)$ とおく．各 x_n と x_{n+1} の中点 $M_n = (x_n + x_{n+1})/2$ をとり，3 点

$$(x_n, f(x_n)), \quad (M_n, f(M_n)), \quad (x_{n+1}, f(x_{n+1}))$$

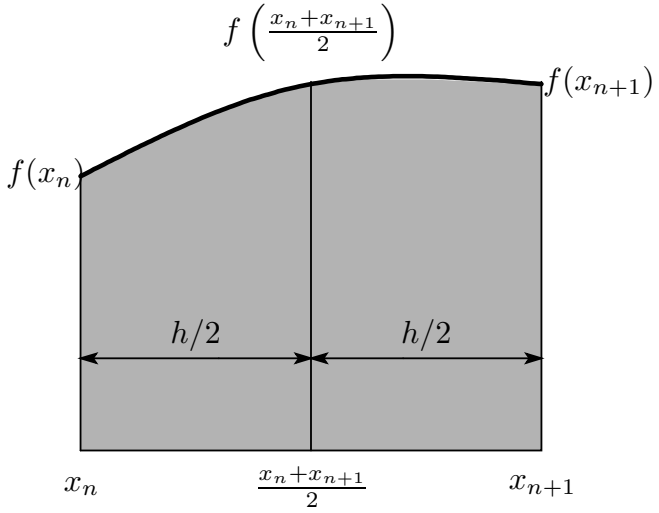


図 8.3 Shimpson の公式

を通る 2 次関数で区間 $[x_n, x_{n+1}]$ 上の f を近似する（図 8.3 を参照）．上の 3 点を通る 2 次多項式を Lagrange 補間公式で求め，区間 $[x_n, x_{n+1}]$ で積分すれば，その区間における f の積分の近似値 $I_2^{(n)}$ は以下で与えられる．

$$\begin{aligned} I_2^{(n)} &= \frac{h}{6}\{f(x_n) + 4f(M_n) + f(x_{n+1})\} \\ &= \frac{h}{6}\left\{f(x_n) + 4f\left(\frac{x_n + x_{n+1}}{2}\right) + f(x_{n+1})\right\} \\ &= \frac{h}{6}\left\{f(a+nh) + 4f\left(a + nh + \frac{h}{2}\right) + f(a+nh+h)\right\} \\ n &= 0, 1, 2, \ldots, N-1 \end{aligned}$$

したがって，Simpson の公式による積分 (8.3) の近似値 I_2 は以下で与えられる．

$$\begin{aligned} I_2 &= \sum_{n=0}^{N-1} I_2^{(n)} \\ &= \frac{h}{6}\sum_{n=0}^{N-1}\left\{f(a+nh) + 4f\left(a + nh + \frac{h}{2}\right) + f(a+nh+h)\right\} \end{aligned}$$

練習問題 33 次の各積分の近似値を Shimpson の公式を用いて求めよ．

1. $\displaystyle\int_0^1 \frac{1}{\sqrt{1+x^2}}dx$
2. $\displaystyle\int_0^{\pi/2} \sqrt{1 - \frac{1}{4}\sin^2 x}\; dx$
3. $\displaystyle\int_0^1 e^{-x^2}dx$
4. $\displaystyle\int_0^{\pi} \frac{\sin x}{x}dx$

ただし，問題 4 では，次の性質を用いよ．

$$\lim_{x \to 0} \frac{\sin x}{x} = 1.$$

8.4　さらに勉強するために

Lagrange 補間と数値積分については，[46, 37, 22] などが参考になる．Lagrange 補間以外の補間法としては，スプライン補間が工学ではよく使われる．スプライン補間に関しては，[37, 4, 34] などを参照せよ．多変数の数値積分にも補間型の公式があるが，変数の数に対して計算量が指数的に増大するので，変数が非常に多い場合は使えない．このような積分の計算には，**モンテカルロ法** (Monte Carlo method) がよく用いられる．モンテカルロ法に関しては，[40, 22] などを参照せよ．

第 9 章

最小 2 乗法と正則化法

N 組のデータ $\{(x_n, y_n)\}_{n=1}^N$ が与えられたとき，このデータから連続関数を求める手法として，前章では多項式補間の手法を述べた．これは，データに誤差が含まれていない場合は有効な手法であるが，もしデータに誤差が含まれている場合，その誤差に多項式が影響されるのであまり有効な手法ではない．また，データ数 N が非常に大きい場合，多項式の次数も非常に大きくなる．

このような状況では，補間よりも**近似** (approximation) または**回帰** (regression) と呼ばれる手法を用いるほうが良い．近似または回帰とは，データ $\{(x_n, y_n)\}_{n=1}^N$ のなるべく近くを通る簡単な関数を求める手法である．ここでは，必ずしもデータ点の上を関数が通る必要はない．簡単な関数としては，例えば，1 次関数（直線）や 2 次関数などがある．データと関数との近さを測るには，ノルムを導入する必要があるが，2 乗ノルムを用いた**最小 2 乗法** (least squares) は計算が簡単であり，非常に良く用いられる．

本章では，この最小 2 乗法についてまず述べ，次に最小 2 乗法における問題点（オーバーフィッティング）とその解決法の一つである正則化法について述べる．

9.1 最小 2 乗近似

最小 2 乗近似問題は以下のように定式化される．

最小 2 乗近似問題

N 組のデータ $\{(x_n, y_n)\}_{n=1}^N$ が与えられたとき，次の **2 乗誤差** (squared error) E

を最小にする M 次多項式 $f_M(x)$ を求めよ[*1].

$$E = \sum_{n=1}^{N} |y_n - f_M(x_n)|^2. \tag{9.1}$$

この問題を解くために，M 次多項式 $f_M(x)$ を

$$f_M(x) = c_0 + c_1 x + \cdots + c_M x^M = \sum_{m=0}^{M} c_m x^m$$

とおく．すなわち，2 乗誤差 E を最小化する c_m $(m = 0, 1, 2, \ldots, M)$ を求める．この多項式を (9.1) に代入すると

$$\begin{aligned} E &= \sum_{n=1}^{N} |y_n - f_M(x_n)|^2 \\ &= \sum_{n=1}^{N} \left(y_n - \sum_{m=0}^{M} c_m x_n^m \right)^2 \\ &= \sum_{n=1}^{N} y_n^2 - 2 \sum_{n=1}^{N} y_n \sum_{m=0}^{M} c_m x_n^m + \sum_{n=1}^{N} \left(\sum_{m=0}^{M} c_m x_n^m \right)^2 \end{aligned}$$

が得られる．誤差 E を最小化する c_m $(m = 0, 1, 2, \ldots, M)$ は，

$$\frac{\partial E}{\partial c_m} = 0, \quad m = 0, 1, 2, \ldots, M \tag{9.2}$$

を満たす必要がある．ここで，各 m に対して

$$\frac{\partial E}{\partial c_m} = -2 \sum_{n=1}^{N} y_n x_n^m + 2 \sum_{n=1}^{N} \sum_{k=0}^{M} c_m x_n^{m+k}$$

が成り立つので，(9.2) より，

$$\sum_{n=1}^{N} y_n x_n^m = \sum_{n=1}^{N} \sum_{k=0}^{M} c_m x_n^{m+k}, \quad m = 0, 1, 2, \ldots, M$$

となる．これは c_m $(m = 0, 1, 2, \ldots, M)$ についての線形連立方程式であるので，解くのは容易である．すなわち，

$$\sum_{n=1}^{N} y_n x_n^m = \beta_m, \quad \sum_{n=1}^{N} x_n^{m+k} = \alpha_{mk}, \quad m, k = 0, 1, 2, \ldots, M$$

[*1] ここでは，次数 M はあらかじめ与えられているものとする．

とおくと,

$$\beta_m = \sum_{k=0}^{M} c_k \alpha_{mk}, \quad m = 0, 1, 2, \ldots, M$$

すなわち,

$$\begin{bmatrix} \alpha_{00} & \alpha_{01} & \cdots & \alpha_{0M} \\ \alpha_{10} & \alpha_{11} & \cdots & \alpha_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{M0} & \alpha_{M1} & \cdots & \alpha_{MM} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_M \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_M \end{bmatrix} \tag{9.3}$$

が得られる．これを**正規方程式** (normal equation) と呼ぶ．すなわち，データ $\{x_n, y_n\}$ が与えられたとき，その最小 2 乗近似多項式は正規方程式 (9.3) を解くことにより得られることがわかる．

9.1.1 線形近似

近似多項式の次数が $M = 1$ の場合，すなわち与えられたデータに対して，直線で近似する場合を考える．これを**線形近似** (linear approximation) または**線形回帰** (linear regression) という．

まず α_{mk} および β_m $(m, k = 0, 1)$ を求める．

$$\alpha_{00} = \sum_{n=1}^{N} x_n^{0+0} = \sum_{n=1}^{N} 1 = N, \qquad \alpha_{01} = \sum_{n=1}^{N} x_n^{0+1} = \sum_{n=1}^{N} x_n,$$
$$\alpha_{10} = \sum_{n=1}^{N} x_n^{1+0} = \sum_{n=1}^{N} x_n = \alpha_{01}, \qquad \alpha_{11} = \sum_{n=1}^{N} x_n^{1+1} = \sum_{n=1}^{N} x_n^2,$$
$$\beta_0 = \sum_{n=1}^{N} y_n x_n^0 = \sum_{n=1}^{N} y_n, \qquad \beta_1 = \sum_{n=1}^{N} y_n x_n^1 = \sum_{n=1}^{N} y_n x_n.$$

また，正規方程式 (9.3) は以下で与えられる．

$$\begin{bmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

これを解くと,

$$\begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \frac{1}{\alpha_{00}\alpha_{11} - \alpha_{01}\alpha_{10}} \begin{bmatrix} \alpha_{11} & -\alpha_{01} \\ -\alpha_{10} & \alpha_{00} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \frac{1}{\alpha_{00}\alpha_{11} - \alpha_{01}^2} \begin{bmatrix} \alpha_{11}\beta_0 - \alpha_{01}\beta_1 \\ -\alpha_{01}\beta_0 + \alpha_{00}\beta_1 \end{bmatrix}$$

となる．各項を具体的に計算すると，

$$\alpha_{00}\alpha_{11}-\alpha_{01}^2 = N\sum_{n=1}^{N}x_n^2-\left(\sum_{n=1}^{N}x_n\right)^2 = N^2\left\{\frac{1}{N}\sum_{n=1}^{N}x_n^2-\left(\frac{1}{N}\sum_{n=1}^{N}x_n\right)^2\right\}$$

$$\begin{aligned}\alpha_{11}\beta_0-\alpha_{01}\beta_1 &= \left(\sum_{n=1}^{N}x_n^2\right)\left(\sum_{n=1}^{N}y_n\right)-\left(\sum_{n=1}^{N}x_n\right)\left(\sum_{n=1}^{N}y_nx_n\right)\\ &= N^2\left\{\left(\frac{1}{N}\sum_{n=1}^{N}x_n^2\right)\left(\frac{1}{N}\sum_{n=1}^{N}y_n\right)-\left(\frac{1}{N}\sum_{n=1}^{N}x_n\right)\left(\frac{1}{N}\sum_{n=1}^{N}y_nx_n\right)\right\}\end{aligned}$$

$$\begin{aligned}-\alpha_{01}\beta_0+\alpha_{00}\beta_1 &= -\sum_{n=1}^{N}x_n\sum_{n=1}^{N}y_n+(N)\sum_{n=1}^{N}y_nx_n\\ &= (N)^2\left\{\frac{1}{N}\sum_{n=1}^{N}y_nx_n-\left(\frac{1}{N}\sum_{n=1}^{N}x_n\right)\left(\frac{1}{N}\sum_{n=1}^{N}y_n\right)\right\}\end{aligned}$$

となる．ここで，

$$m_x := \frac{1}{N}\sum_{n=1}^{N}x_n,\quad m_y := \frac{1}{N}\sum_{n=1}^{N}y_n,\quad m_{xx} := \frac{1}{N}\sum_{n=1}^{N}x_n^2,\quad m_{xy} := \frac{1}{N}\sum_{n=1}^{N}x_ny_n$$

とおくと，近似多項式の係数 c_0 と c_1 は以下のように与えられる．

$$c_0 = \frac{m_{xx}m_y-m_xm_{xy}}{m_{xx}-m_x^2}, \tag{9.4}$$

$$c_1 = \frac{m_{xy}-m_xm_y}{m_{xx}-m_x^2}. \tag{9.5}$$

統計学の用語では，$m_{xy}-m_xm_y$ はデータ $\{x_n,y_n\}_{n=1}^N$ の**共分散** (covariance)，$m_{xx}-m_x^2$ はデータ $\{x_n\}_{n=1}^N$ の**分散** (variance) と呼ばれる量である．すなわち，近似多項式（直線）の傾き c_1 は共分散と分散の比で与えられることが (9.5) からわかる．また (9.4) より，

$$c_0 = m_y+\frac{m_ym_x-m_{xy}}{m_{xx}-m_x^2}\cdot m_x = m_y-c_1m_x$$

となることから，直線 $y=c_1x+c_0$ はデータの平均値 (m_x,m_y) を通ることがわかる．

以上より，データ $\{x_n,y_n\}_{n=1}^N$ が与えられたときに最小 2 乗近似直線を求める手順は以下のようになる．

1. 傾き c_1 を共分散と分散の比 (9.5) により求める．
2. 直線の式 $y=c_0+c_1x$ に $x=m_x$, $y=m_y$ を代入し，c_0 を求める．

練習問題 34　1. 次のデータに対する最小 2 乗近似直線 $y=ax+b$ を求めよ．

x	0	1	2	3	4
y	1	2	3	3	4

2. 次のデータの空欄を推定せよ．

x	0	1	2	3	4	5	6
y	1	2	3	3	4		

3. 次のデータについて 2 乗誤差 E を最小にする関数 $f(x) = be^{ax}$ を求めよ（係数 a と b を求めよ）．

x	1	2	3	4
y	7	11	17	27

（ヒント：データ $\{x_n, \log y_n\}$ に対して，最小 2 乗直線近似を用いる）

9.1.2　多項式近似

ここでは，一般の M 次多項式による近似を考える．

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^M \\ 1 & x_2 & x_2^2 & \dots & x_2^M \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^M \end{bmatrix}$$

とおく．この行列 Φ を**計画行列** (design matrix) と呼ぶ．また，データ $\{y_1, y_2, \dots, y_N\}$ に対して，

$$\boldsymbol{y} := \begin{bmatrix} y_1 & y_2 & \dots & y_N \end{bmatrix}^\top$$

とおく．すると，(9.3) の正規方程式は簡単な計算により

$$(\Phi^\top \Phi)\boldsymbol{c} = \Phi^\top \boldsymbol{y} \tag{9.6}$$

と書ける．ただし，

$$\boldsymbol{c} = \begin{bmatrix} c_0 & c_1 & \dots & c_M \end{bmatrix}^\top$$

である．

練習問題 35 正規方程式 (9.3) が (9.6) と書けることを示せ．■■■■■■■■■■■■■■■■■■■■■■■■ lecture13.tex を見よ．

行列 $(\Phi^\top\Phi)$ が正則であると仮定すると，正規方程式の解は次式で与えられる．

$$\boldsymbol{c} = (\Phi^\top\Phi)^{-1}\Phi^\top\boldsymbol{y}. \tag{9.7}$$

ここで，$(\Phi^\top\Phi)^{-1}\Phi^\top$ は行列 Φ の Moore-Penrose **擬似逆行列** (Moore-Penrose pseudo-inverse matrix) である．すなわち，最小 2 乗法による M 次の多項式近似は計画行列 Φ の Moore-Penrose 擬似逆行列を求めることに他ならない．

9.2　正則化法

ここでは，前節で求めた M 次の多項式近似において，データに誤差が含まれる場合を考える．データ $\{x_n\}$ を $\{x_0, x_1, \ldots, x_{10}\} = \{0, 1, \ldots, 10\}$ とし，データ $\{y_n\}$ を

$$y_n = \sin x_n + \varepsilon$$

で生成する．ただし，ε は平均 0，分散 0.2 の正規分布に従う確率変数とする[*2]．このデータの組 $\{(x_n, y_n)\}_{n=0}^{10}$ を最小 2 乗法により $M = 10$ 次多項式で近似してみよう．コンピュータを使って計算した多項式を図 9.1 に示す．求めた多項式は雑音を含んだデータ点 $\{x_n, y_n\}_{n=0}^{10}$ のほぼ真上を通っているが，データ点の間で振動していることがわかる．この現象を**オーバーフィッティング** (overfitting) という．この現象の原因を調べるために，多項式の係数 $\boldsymbol{c}$ を見てみよう．

```
c  =
  - 0.4469088
  - 40.784588
    115.72864
  - 127.11682
    74.264944
  - 25.909959
    5.6595434
  - 0.7802941
    0.0659165
  - 0.0031132
    0.0000629
```

[*2] 正規分布や確率変数などについては [2] を参照せよ．おおざっぱにいって，ε は 0 を中心として，大きさが 0.2 以下である確率が約 68%，0.4 以上である確率が約 5% であるような雑音である．

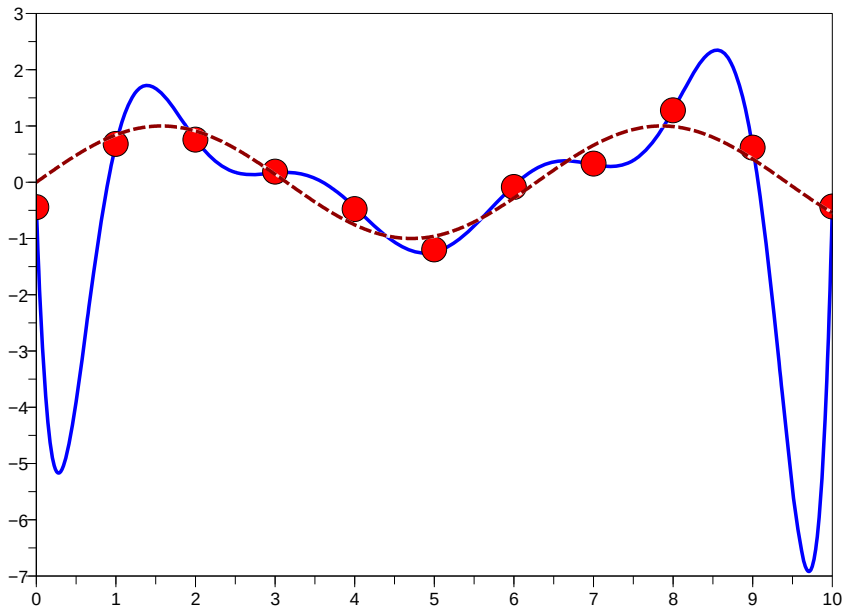


図 9.1　最小 2 乗法を用いた 10 次多項式による近似（実線）．点線は元の関数 $y = \sin x$，丸印はデータ点である．

係数の中に非常に大きな値があることがわかるだろう．これがオーバーフィッティングの原因（の一つ）である．そこで，係数 $\boldsymbol{c}$ の大きさを制限するために，最小 2 乗法の 2 乗誤差 (9.1) のかわりに次の目的関数を最小化することを考える．

$$E_\lambda = \sum_{n=1}^{N} |y_n - f_M(x_n)|^2 + \lambda \Omega(f_M). \tag{9.8}$$

ただし，

$$\Omega(f_M) := \sum_{m=0}^{M} c_m^2 = \boldsymbol{c}^\top \boldsymbol{c} \tag{9.9}$$

であり，これを**正則化項** (regularization term) と呼ぶ．また (9.8) を最小化する $\boldsymbol{c}$ を求める方法を**正則化最小 2 乗法** (regularized least squares method) または，**正則化法** (regularization) と呼ぶ．特に (9.9) のように係数 $\boldsymbol{c}$ のユークリッドノルムを正則化項に持つ正則化法を Ridge **回帰** (Ridge regression) とも呼ぶ．目的関数 (9.8) を係数 $\boldsymbol{c}$ によって表すと次式が得られる．

$$E_\lambda = \sum_{n=1}^{N} |y_n - \boldsymbol{c}^\top \boldsymbol{\phi}(x_n)|^2 + \lambda \boldsymbol{c}^\top \boldsymbol{c}. \tag{9.10}$$

ここで，

$$\boldsymbol{\phi}(x_n) = \begin{bmatrix} 1 & x_n & x_n^2 & \dots & x_n^M \end{bmatrix}^\top$$

である．目的関数 (9.10) を $\boldsymbol{c}$ で微分し 0 とおくことにより，次の正規方程式が得られる．

$$\left(\lambda I + \Phi^\top \Phi\right) \boldsymbol{c} = \Phi^\top \boldsymbol{y}.$$

これより，行列 $\lambda I + \Phi^\top \Phi$ が正則とすると，この正規方程式の解，すなわち (9.10) を最小化する係数 $\boldsymbol{c}$ は次で与えられることがわかる．

$$\boldsymbol{c} = \left(\lambda I + \Phi^\top \Phi\right)^{-1} \Phi^\top \boldsymbol{y}.$$

ここで，パラメータ λ は自由に決めることができるので[*3]，もし $\Phi^\top \Phi$ が正則でなくても，適当に λ を与えることにより $(\lambda I + \Phi^\top \Phi)$ を正則にすることができる．これも正則化法の利点である．

図 9.2 に $\lambda = 1$ の場合の正則化法の計算結果を示す．図 9.1 と比べデータ点の間の振動がなくなっていることがわかる．この正則化法により求まった係数 $\boldsymbol{c}$ を見てみよう．

```
c  =
    0.1723135
    0.2255873
    0.1675314
    0.0659603
  - 0.0371057
  - 0.0388592
    0.0215341
  - 0.0042282
    0.0003999
  - 0.0000182
    0.0000003
```

最小 2 乗法と比べて係数の値がどれも小さく，正則化項の効果があらわれていることがわかる．

9.3　Representer 定理，カーネル法，サポートベクター回帰

[*3] この λ はパラメータ $\boldsymbol{c}$ を決めるためのパラメータであり，Bayes **理論** (Bayes theory) では**超パラメータ** (hyperparameter) と呼ばれるものである．Bayes 理論を用いれば，超パラメータもデータから推定することが可能である．詳しくは [2] を参照せよ．

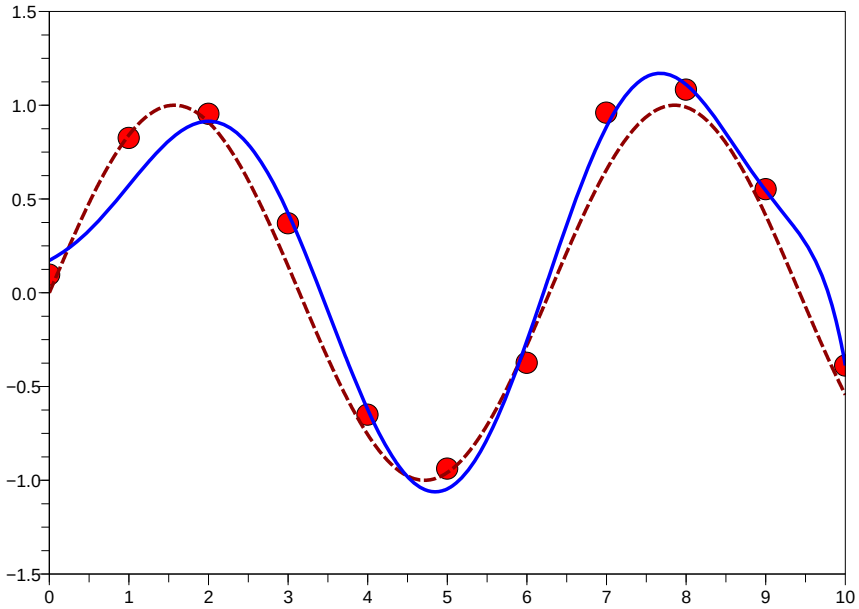


図 9.2　正則化法を用いた 10 次多項式による近似（実線）．点線は元の関数 $y = \sin x$，丸印はデータ点である．

9.4　赤池情報量基準とモデル選択

9.5　L^1 正則化と圧縮センシング

9.6　さらに勉強するために

最小 2 乗法については，[16, 2] などを参考にした．本章で述べた手法は多項式，すなわち x^m $(m = 0, 1, 2, \ldots, M)$ の線形結合で表される関数で近似する手法であるが，一般の関数 $\{\phi_m(x)\}$ の線形結合としても，やり方は同じである．例えば，$\{\phi_m(x)\}$ として直交多項式を使う方法が [46] にある．また，正則化法については [2, 24] などが参考になる．本章では，多項式の次数 M は与えられたものとして議論したが，データから次数 M を

決定する手法として**赤池情報量基準** (Akaike Information Criterion, AIC) がある．これについては，[2] などを参照せよ．

付録 A

本書で使う数学

A.1 連続関数の性質

本書でひんぱんに使用する連続関数の重要な定理を 3 つ述べる．証明は [23, 35] などを参照のこと．

定理 28 (中間値の定理 (intermediate value theorem)) 実数値関数 f を $\mathbb{R}$ 上の有界閉区間 $[a, b]$ で連続とする．このとき，任意の $\gamma \in (f(a), f(b))$ または $\gamma \in (f(b), f(a))$ に対して，$f(c) = \gamma$ を満たす $c \in (a, b)$ が存在する．

系 3 実数値関数 f を $\mathbb{R}$ 上の有界閉区間 $[a, b]$ で連続とする．このとき，

$$f(a) \cdot f(b) < 0$$

ならば，$f(c) = 0$ となる実数 $c \in (a, b)$ が存在する．

定理 29 (平均値の定理 (mean value theorem)) 実数値関数 f を $\mathbb{R}$ 上の有界閉区間 $I = [a, b]$ で連続，かつ開区間 (a, b) で微分可能とする．このとき，

$$f(b) - f(a) = f'(c)(b - a)$$

を満たす $c \in (a, b)$ が存在する．

定理 30 (Taylor の定理 (Taylor's theorem)) 実数値関数 f を $\mathbb{R}$ 上の有界閉区間 $I = [a, b]$ で C^{n-1} 級（n は正の整数）かつ，任意の $t \in (a, b)$ に対して $f^{(n)}(t)$ が存在するとする．また，$\alpha, \beta \in [a, b]$ $(\alpha \neq \beta)$ とする．このとき，ある $x \in (\alpha, \beta)$ が存在して，

$$f(\beta) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(\beta - \alpha)^k + \frac{f^{(n)}(x)}{n!}(\beta - \alpha)^n$$

が成り立つ．

A.2　ベクトル空間

集合 X が**ベクトル空間** (vector space) であるとは，任意の元 $x, y \in X$ およびスカラー $\alpha \in \mathbb{K}$ ($= \mathbb{R}$ または $\mathbb{C}$) に対して，**和** $x + y \in X$ と **スカラー倍** $\alpha x \in X$ が定義され，以下の規則が成り立つ場合である．

1. 任意の $x, y, z \in X$ に対して $(x + y) + z = x + (y + z)$ (**結合法則**).
2. 任意の $x, y \in X$ に対して $x + y = y + x$ (**交換法則**)
3. X の中に 0 で表される元（**ゼロ元**という）が一意に存在し，X の任意の元 x に対し $x + 0 = x$ が成り立つ．
4. 任意の $x \in X$ に対し，$x + x' = 0$ となる X の元 x' がただ 1 つ存在する．これを x の**逆元**といい，$-x$ で表す．
5. 任意の $x \in X$ に対して $1x = x$.
6. 任意の $x \in X$ と任意の $a, b \in \mathbb{K}$ に対して $a(bx) = (ab)x$.
7. 任意の $x \in X$ と任意の $a, b \in \mathbb{K}$ に対して $(a + b)x = ax + bx$ (**分配法則**).
8. 任意の $x, y \in X$ と任意の $a \in \mathbb{K}$ に対して $a(x + y) = ax + ay$ (**分配法則**).

例えば，$\mathbb{R}^n$，$\mathbb{C}^n$，$C[a, b]$ (区間 $[a, b]$ 上で定義された連続関数の集合) 等はベクトル空間である．

A.2.1　ベクトルのノルム

ベクトル空間 X で定義され，実数値をとる関数 $\|\cdot\| : X \to \mathbb{R}$ が X の**ノルム** (norm) であるとは，次の条件 1〜3 が成り立つことである．

1. 任意の $x \in X$ に対して $\|x\| \geq 0$．とくに等号については $\|x\| = 0 \Leftrightarrow x = 0$.
2. 任意の $x \in X$ と任意の $a \in \mathbb{K}$ に対して $\|ax\| = |a|\|x\|$.
3. 任意の $x, y \in X$ に対して $\|x + y\| \leq \|x\| + \|y\|$.

条件 3 の不等式を，**三角不等式** (triangle inequality) と呼ぶ．上のようにノルムが定義されたベクトル空間を**ノルム空間** (normed space) という．

ノルムの連続性

ベクトル $\mathbb{R}^n$ のノルム

ベクトル空間として $X = \mathbb{R}^n$ を考える．$\mathbb{R}^n$ の元

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

に対して，次のノルムを定義する．

1. $\|x\|_1 := \sum_{i=1}^{n} |x_i|$
2. $\|x\|_2 := \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}$
3. $\|x\|_\infty := \max_{1 \le i \le n} |x_i|$

上の 3 つのノルムはそれぞれノルムの定義の条件 1〜3 を満たす（証明せよ）．

特に，ノルム $\|x\|_2$ は Euclid **ノルム** (Euclidean norm) と呼ばれることが多い．

なお，同じ $\mathbb{R}^n$ でもノルムが異なれば別の空間となることに注意．これらの空間を区別するために $(\mathbb{R}^n, \|\cdot\|_1)$, $(\mathbb{R}^n, \|\cdot\|_2)$, $(\mathbb{R}^n, \|\cdot\|_\infty)$ 等と表す場合もある．

A.2.2 行列 $\mathbb{R}^{n\times n}$ のノルム

$n \times n$ の行列に対してもノルムが定義される．

$$\|A\| := \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$$

このように定義されたノルムを行列の**自然なノルム** (natural norm)，あるいはベクトルのノルムから**誘導されたノルム** (induced norm) という．

補題 9 次の関係が成り立つ．

1. $\|Ax\| \le \|A\|\|x\|$
2. $\|AB\| \le \|A\|\|B\|$

ここで $\|Ax\|$ や $\|x\|$ は前節で定義したベクトルのノルムであり，ベクトルのノルムの定義によって各種のノルムが存在する．

1. $\|A\|_1 := \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|_1}{\|x\|_1}$

2. $\|A\|_2 := \sup_{x\in\mathbb{R}^n, x\neq 0} \dfrac{\|Ax\|_2}{\|x\|_2}$
3. $\|A\|_\infty := \sup_{x\in\mathbb{R}^n, x\neq 0} \dfrac{\|Ax\|_\infty}{\|x\|_\infty}$

補題 10 次の関係が成り立つ.

1. $\|A\|_1 = \max_{1\le k\le n} \sum_{i=1}^{n} |a_{ik}|$
2. $\|A\|_2 = \sqrt{\max_{1\le i\le n} |\lambda_i(A^\top A)|}$
   (ただし $\lambda_i(A^\top A)$ は行列 $A^\top A$ の i 番目の固有値を表す)
3. $\|A\|_\infty = \max_{1\le i\le n} \sum_{k=1}^{n} |a_{ik}|$

A.2.3 完備なノルム空間

ノルム空間 X の元の列 $\{x_k\}_{k=0}^{\infty}$ を考える．任意の $\varepsilon > 0$ に対して，ある自然数 N が存在して，任意の $m, n \ge N$ に対して

$$\|x_m - x_n\| < \varepsilon$$

となるとき，ベクトル列 $\{x_k\}$ を **Cauchy 列** (Cauchy sequence) という．空間 X における任意の Cauchy 列の極限が空間 X の元となるとき，X は**完備** (complete) であるという．完備なノルム空間を **Banach 空間** (Banach space) と呼ぶ.

A.2.4 行列の固有値と固有ベクトル

行列 $A \in \mathbb{R}^{n\times n}$ の**固有値** (eigenvalue) は**特性方程式** (characteristic equation)

$$\begin{aligned}\psi(\lambda) &= \det(\lambda I - A)\\ &= \det \begin{bmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{bmatrix}\end{aligned}$$

の根で与えられる．特性方程式 $\psi(\lambda)$ は λ の n 次多項式

$$\psi(\lambda) = \lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \cdots + (-1)^n p_n$$

となる．ここで，とくに

$$p_1 =^\top A = \sum_{i=1}^{n} a_{ii}, \quad p_n = \det A$$

である．

定理 31 (スペクトル写像定理 (spectral mapping theorem)) $\lambda_1, \ldots, \lambda_n$ を A の固有値とし，$f(z)$ を $z = \lambda_i, i = 1, \ldots, n$ で解析的な複素関数とする．このとき $f(A)$ の固有値は，$f(\lambda_1), \ldots, f(\lambda_n)$ で与えられる．

系 4 λ を A の固有値とすると，λ^{-1} は A^{-1} の固有値である．

定理 32 (Cayley-Hamilton の定理 (Cayley-Hamilton theorem)) $\psi(A) = 0$

λ_i を A の固有値とする．この λ_i に対応して

$$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$$

を満足するベクトル $\boldsymbol{v}_i$ が存在する．この $\boldsymbol{v}_i$ を λ_i に属する**固有ベクトル** (eigenvector) という．

A.2.5 スペクトル半径

A の固有値 λ_i，固有ベクトル $\boldsymbol{v}_i$ に関して

$$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$$

より，両辺のノルムをとれば

$$\|A\boldsymbol{v}_i\| = \|\lambda_i \boldsymbol{v}_i\| = |\lambda_i| \|\boldsymbol{v}_i\|$$

であるから

$$\frac{\|A\boldsymbol{v}_i\|}{\|\boldsymbol{v}_i\|} = |\lambda_i|$$

となる．したがって任意の i について

$$\|A\| = \sup_{\boldsymbol{v} \neq 0} \frac{\|A\boldsymbol{v}\|}{\|\boldsymbol{v}\|} \geq |\lambda_i|$$

が成り立つ．これより，

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

とおくと，

$$\|A\| \geq \rho(A)$$

が成り立つ．この $\rho(A)$ を行列 A の**スペクトル半径** (spectral radius) という．

定理 33 $\|A\|$ を行列 A の自然なノルムとするとき

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

定理 34 A を与えられた行列とする．任意の $\varepsilon > 0$ に対して

$$\|A\|_\alpha \leq \rho(A) + \varepsilon$$

を満たす自然なノルム $\|\cdot\|_\alpha$ が存在する．

補題 11 行列 $A \in \mathbb{R}^{n\times n}$ が $\|A\| < 1$ を満たすならば， $I - A$ は正則であり

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$$

を満たし，さらに

$$\|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}$$

が成り立つ．

A.2.6 行列の条件数

行列 $A \in \mathbb{R}^{n\times n}$ が正則のとき

$$k_\alpha(A) = \|A\|_\alpha \|A^{-1}\|_\alpha$$

を A の**条件数** (conditional number) という．とくに，A が Hermite **行列** (Hermitian matrix) のとき，すなわち $A = \bar{A}^\top$ が成り立つとき，$\alpha = 2$ の場合の条件数 $k_2(A)$ は

$$k_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|}$$

で与えられる．

A.2.7 有限次元ベクトル空間におけるノルムの同値性

定理 35 (ノルムの同値性) $\mathbb{R}^n$ の 2 つのベクトルノルム $\|\boldsymbol{x}\|_\alpha$, $\|\boldsymbol{x}\|_\beta$ が与えられたとき，すべての n 次元ベクトル $\boldsymbol{x}$ に対して

$$m\|\boldsymbol{x}\|_\alpha \leq \|\boldsymbol{x}\|_\beta \leq M\|\boldsymbol{x}\|_\alpha$$

を満足する正の数 m, M が存在する．

系 5 与えられた n 次元ベクトルの列 $\{\boldsymbol{x}[n]\}_{n=0}^{\infty}$ に対して

$$\lim_{n\to\infty}\|\boldsymbol{x}[n]-\boldsymbol{x}^*\|_\alpha=0 \Leftrightarrow \lim_{n\to\infty}\|\boldsymbol{x}[n]-\boldsymbol{x}^*\|_\beta=0$$

が成り立つ.

上の 定理および系より，n 次元ベクトルの列 $\{\boldsymbol{x}[n]\}_{n=0}^{\infty}$ があるノルムに関してベクトル $\boldsymbol{x}^*$ に収束すれば，他のすべてのノルムに関して $\{\boldsymbol{x}[n]\}_{n=0}^{\infty}$ は同じベクトル $\boldsymbol{x}^*$ に収束する．この意味で，有限次元ベクトル空間のノルムは互いに**同値** (equivalent) であるという.

A.2.8　対角優位行列

行列 $A\in\mathbb{R}^{n\times n}$ が

$$|a_{ii}|>\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|,\quad i=1,2,\ldots,n$$

を満足するとき，行列 A を**対角優位行列** (diagonally dominant matrix) という.

A.2.9　正定値対称行列

行列 $A\in\mathbb{R}^{n\times n}$ を対称行列とする．$\boldsymbol{x}\neq\boldsymbol{0}$ である任意のベクトル $\boldsymbol{x}\in\mathbb{R}^n$ に対して

$$\boldsymbol{x}'A\boldsymbol{x}=\sum_{i,j=1}^{n}a_{ij}x_ix_j>0$$

が成立するとき，行列 A を**正定値行列** (positive definite matrix) といい，$A>0$ と表す.

定理 36 対称行列 A が正定値であるための必要十分条件は，A の固有値が全て正であることである.

定理 37 (Sylvester の判定法) 対称行列 A が正定値であるための必要十分条件は，A の**主座小行列式** (leading principal minor) が全て正であることである．すなわち,

$$a_{11}>0,\quad \det\left(\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{bmatrix}\right)>0,\ldots,\quad \det(A)>0$$

が成り立つことである.

定理 38 (Schur 補元) 対称行列 $A\in\mathbb{R}^{n\times n}$ を $n=n_1+n_2$ として

$$A=\begin{bmatrix}A_{11}&A_{12}\\A_{12}^\top&A_{22}\end{bmatrix},\quad A_{11}\in\mathbb{R}^{n_1\times n_1},A_{12}\in\mathbb{R}^{n_1\times n_2},\quad A_{22}\in\mathbb{R}^{n_2\times n_2}$$

と分解する．このとき，以下の 3 つの条件は互いに等価である.

1. $A > 0$
2. $A_{11} > 0$ かつ $A_{22} - A_{12}^{\top} A_{11}^{-1} A_{12} > 0$
3. $A_{22} > 0$ かつ $A_{11} - A_{12} A_{22}^{-1} A_{21}^{\top} > 0$

なお，定理 38 に出てくる行列 $A_{22} - A_{12}^{\top} A_{11}^{-1} A_{12}$ および $A_{11} - A_{12} A_{22}^{-1} A_{21}^{\top}$ は，行列 A の Schur **補元** (Schur complement) と呼ばれる．

A.2.10 Gram-Schmidt の直交化法

A.2.11 もっと勉強するために

連続関数の性質に関しては，微分積分の標準的な教科書 [23, 35] のほか，[12] が参考になる．また，ベクトル空間やノルム Banach 空間などについては [50, 15, 16, 18] などを参照せよ．行列に関する性質は，[11, 10] が詳しい．

参考文献

[1] A. Bhaya and E. Kaszkurewicz. *Control Perspectives on Numerical Algorithms and Matrix Problems*. SIAM, 2006.

[2] C. M. Bishop. パターン認識と機械学習. Springer, 2007.

[3] F. C. Chatelin. 行列の固有値. Springer, 1993. 伊理（訳）.

[4] C. K. Chui. ウェーブレット入門. 東京電機大学出版局, 1993. 桜井, 新井（訳）.

[5] S. D. Conte and C. de Boor. *Elementary Numerical Analysis*. McGraw-Hill, 3rd edition edition, 1981.

[6] J. J. D'Azzo and C. H. Houpis. *Linear Control System Analysis and Design*. McGraw-Hill, 1995.

[7] C. A. Desoer and E. S. Kuh. *Basic Circuit Theory*. McGraw-Hill, 1969.

[8] W. Feller. 確率論とその応用 I, II. 紀伊國屋書店, 1960, 1970. 河田（監訳）.

[9] D. Goldberg. What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys*, Vol. 23, pp. 5–48, 1991.

[10] D. A. Harville. 統計のための行列代数. シュプリンガージャパン, 2007. 伊理（監訳）.

[11] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge, 1985.

[12] J. Jost. ポストモダン解析学. Springer, 2000. 小谷（訳）.

[13] W. Kahan. *IEEE Standard 754 for Binary Floating-Point Arithmetic*. Lecture Notes on the Status of IEEE 754, 1996.

[14] K. Kashima and Y. Yamamoto. System theory for numerical analysis. *Automatica*, Vol. 43, pp. 1156–1164, 2007.

[15] T. Kato. 行列の摂動. Springer, 1999. 丸山（訳）.

[16] J. P. Keener. キーナー応用数学. 日本評論社, 2007. 坂元（訳）.

[17] B. Kernighan and R. Pike. プログラミング作法. ASCII, 2000. 福崎（訳）.

[18] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Dover, 1975.

[19] W. Lichten. 計測データと誤差解析の入門. ピアソン・エデュケーション, 2004. 村上（訳）.

[20] J. M. Ortega. *Numerical Analysis: A Second Course*. SIAM, 1990.

[21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Numerical Recipes in C （日本語版）. 技術評論社, 1993.

[22] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Springer, 2000.

[23] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1964.

[24] B. Schölkopf and A. J. Smola. *Learning with Kernels*. The MIT Press, 2002. http://www.learning-with-kernels.org/.

[25] R. Schreier and G. C. Temes. ΔΣ アナログ/デジタル変換器入門. 丸善, 2007. 和保, 安田（監訳）.

[26] J. L. Shearer, A. T. Murphy, and H. H. Richardson. *Introduction to System Dynamics*. Addison-Wesley, 1971.

[27] IEEE Computer Society. *IEEE Standard for Floating-Point Arithmetic*. IEEE Std 754$^{\mathrm{TM}}$, 1985, 2008 (revised version).

[28] 荒木. 古典制御理論. 培風館, 2000.

[29] 映像情報メディア学会（編）. *MPEG*. オーム社, 1996.

[30] 大石. 非線形解析入門. コロナ社, 1997.

[31] 太田. システム制御のための数学 (1)—線形代数編—. コロナ社, 2000.

[32] 大野. Scilab 入門—フリーソフトで始める数値シミュレーション. CQ 出版, 2009.

[33] 川田. Scilab で学ぶわかりやすい数値計算法. 森北出版株式会社, 2008.

[34] 桜井（監）菅野, 吉村, 高山（著）. C によるスプライン関数. 東京電機大学出版局, 1993.

[35] 杉浦. 解析入門 I, II. 東京大学出版会, 1980 (I), 1985 (II).

[36] 杉江, 藤田. フィードバック制御入門. コロナ社, 1999.

[37] 杉原, 室田. 数値計算法の数理. 岩波書店, 1994.

[38] 高橋. 非線形・凸解析学入門. 横浜図書, 2005.

[39] 二宮（編）. 数値計算のつぼ. 共立出版, 2004.

[40] 二宮（編）. 数値計算のわざ. 共立出版, 2006.

[41] 橋本, 石井. Scilab/Scicos で学ぶシミュレーションの基礎. オーム社, 2008.

[42] 福井ほか. 新数値計算. 共立出版, 1999.

[43] 増田久弥（編）. 応用解析ハンドブック. Springer, 2010.

[44] 森口, 宇田川, 一松. 岩波数学公式 I, II, III. 岩波書店, 1956 (I), 1957 (II), 1960 (III).

[45] 森, 杉原, 室田. 線形計算, 岩波講座応用数学 12. 岩波書店, 1994.

[46] 森. 数値解析. 共立出版, 2002 (第 2 版).

[47] 矢部, 八巻. 非線形計画法. 朝倉書店, 1999.

[48] 山口. はじめての DSP 活用大全. CQ 出版社, 2006.
[49] 山本. 数値解析入門 [改訂版]. サイエンス社, 2003.
[50] 山本. システムと制御の数学. 朝倉書店, 1998.
[51] 和達, 十河. キーポイント確率・統計. 岩波書店, 1993.

索引