

TOOL
ツール活用シリーズ

卒論や
レポート作成
にも役立つ!

波形解析のための数値計算ソフト

Scilab

入門

見本

信号のスペクトラム、ノイズ分析から特徴抽出まで

大川善邦 著

Yoshikuni Okawa

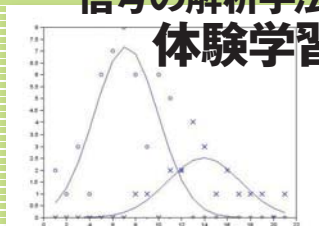


DVD-ROM付き

Scilab 5.4.0

試しながら学べるプロジェクト・ファイルを収録

信号の解析手法を
体験学習!



●ノイズ解析 ●ウェーブレット解析 ●フーリエ解析
●パラメトリック法 ●ヒューリスティック・アルゴリズム

CQ出版社

まえがき

例えば理工系の大学生は、大学生活も終わりに近づくと、卒業論文の締め切りが迫ってきます。

卒業論文は、文章によって構成するものもありますが、多くはデータを扱います。データを解析して、その結果に基づいて結論を述べます。

データを解析する過程を、一般にデータ処理と言います。

私は、長いあいだ大学の教授を務め、多くの学生の卒業論文を見てきました。

学生が提出する論文はさまざまで、経験不足のために残念ながら合格できないものもありましたが、なかにはよく書けている論文もありました。少し磨きをかければ、そのまま学会の論文として通用する、そういった論文もありました。

このような論文の質の違いは、どこから生まれるのでしょうか。

それは、学生の「データを読む」能力の違いです。

学生は、卒論を書き上げるために、実験データを採取して、グラフ化して、論文に貼り付けます。ここまでは、誰でもできます。

もっとも重要なポイントは、データのなかに何を見るか、数の並びから何を引き出すか、ここが大事なポイントです。

いま仮に、赤ちゃんが泣いていたとしましょう。

赤ちゃんが泣くの中には、それなりの理由があります。経験を積んだ母親ならば、瞬間的に赤ちゃんが泣いている理由を読み取り、その原因を取り除きます。ほどなく赤ちゃんはニコニコ顔に戻るでしょう。

「赤ちゃんが泣いている」ということは、誰でもわかることです。しかし、何が原因で泣いているのかを読み取ることができなければ、母親としては失格です。

論文において重要なのは、データを集めることではありません。データは単なる素材です。

素材を調理して結論を絞り出す、ここが最も重要なポイントです。

* * *

本書は、理工系の大学卒業論文やレポートなどを作成する場合に、実験データを数学的テクニックを使って波形の特徴を抽出するコツをまとめました。ツールとして、オープンソースのScilabという計算ソフトウェアを使います。インターネット上で無料で公開されているソフトウェアです。

皆さんが、卒業論文やレポートなどをまとめる際に活用してください。

なお、本書は、Scilabを使って数学的処理を行い、収集したデータの波形から特徴抽出をする手法を解説しています。多くのページは、Scilabを使って数学的処理をする部分に費やされており、一冊にまとめるというページの制限から、Scilabのインストールの詳細やScilabの持つ機能すべてを解説したものではありません。

2013年4月 筆者

第1章

イントロダクション

1.1 — はじめに

具体的な作業に入る前に、波形解析に関するイントロダクションを述べます。

1.2 — 卒業論文やレポート

例えば、大学を卒業する際に、学生は卒業論文を書きます。

卒業論文は、大学を卒業するための必修条件だからです。

つまり、これがなければ、大学卒業の資格は得られません。

理工系学部の場合、卒業論文のなかには、文章だけで構成するものもありますが、これは一部の少数派です。多くは、データを集め、それを分析して結論を導きます。

レポートも同様です。集めたデータを分析して、結論を出します。

例えば、宇宙のかなたから飛んでくる電波を解析して、星群の爆発を確認する、出生率とGDPの関係を数式化する、がんの遺伝性を明らかにする……、などです。

本書では、データを収集後、そのデータからどんなことが見いだせるのかを数学的に検証していきます。ツールとして、オープンソースのScilabという計算ソフトウェアを使います。インターネット上で無料で公開されているソフトウェアです。

1.3 — データを読む

インターネットの時代において、データを集めることはずいぶん容易になりました。データは、あらゆるところに溢れています。

しかし、もっとも大事なことは、データを集めることではなく、データを解析して、データのなかに潜んでいる事実を引っ張り出すことです。

これを、

データを読む

と言います。

論文の価値はデータの量ではありません。データから何を引き出すか、そこが大切なポイントです。

1.4——波形解析

データは、時間とともに変化する数値の並びです。

時間とともに変化するデータを、ここで波形(wave)と呼びます。

いま仮に、大学の健康診断において、学生と教職員の心電図を採取したとします。

膨大な量の波形が集まります。

この波形を短時間で処理するには、どのようにしたらよいでしょうか。

専門の医師に、波形の検査を委託しますか？

処理しなくてはならないデータの量は膨大です。

受診者の多くは健康体です。普通に日常生活を過ごしています。

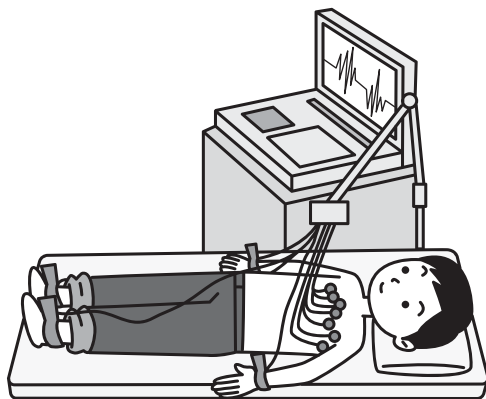


図1.1
心電図の計測

心電図に異常がある人は、ごくわずかです。ゼロの可能性も十分にあります。

医師は、忙しい身です。

おそらく、ある程度の見逃しが起きるでしょう。医師がヤブだと言っているのではありません。要求する作業に無理があるからです。

1.5——コンピュータの利用

人の能力は素晴らしいものだけでも、それには限界があります。

人に、先のような限界を超える作業を要求しても、事実上、正確な作業は行われません。

こういう場面に、コンピュータが登場します。

例えば、検診において採取したチャートを、まずコンピュータにかけます。

20歳前後の若い学生のデータです。

ほとんどは、コンピュータの診断をクリアします。

しかし、なかには問題のある波形があります。

このときは、チャートからその部分を切り出して、医師に診断を依頼します。そうすることによって、医師への負担は劇的に減少します。

1.6——波形解析のツール

波形をコンピュータにかけて、問題の場所を見つけるプログラムを総称して、

波形解析のツール

と呼びます。

波形解析のツールは、例えば地震予知など、多くの分野において使用されています。

ここでは、大学の卒業論文を書く際に使うことを前提にしているので、

波形解析のツールを自分で操作する

とします。

ツールを探して、それを使いこなして論文を書く、ここが本書のメインテーマです。

第3章

ノイズ解析

3.1 —はじめに

計測データは、環境に起因するノイズを含みます。データの解析を始める前に、データからノイズを除去して、信号波形を抽出します。

3.2 —ノイズとは

計測したデータには、ほとんどの場合、ノイズ（あるいは、雑音）を含みます。例えば、宇宙の彼方からやってくる電波は、太陽の黒点活動の影響を受けます。

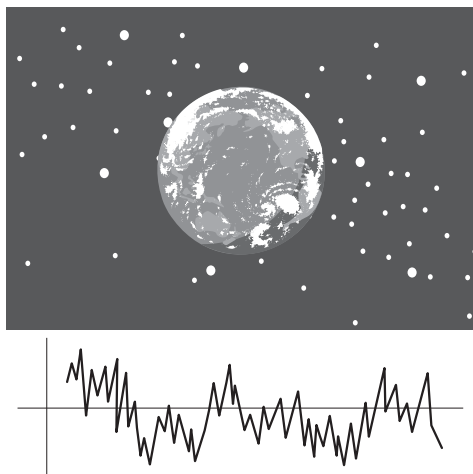


図3.1
宇宙の彼方からやっ
てくる電波



図3.2 米を洗ってゴミを取る

$$\boxed{?} + \boxed{?} = \boxed{10}$$

図3.3 2数の加算

計測機のアンプは、電子の熱運動によるノイズ、あるいは温度変動のドリフトなどを含みます。

つまり、一般に計測データは、信号とノイズを合成した、

$$\text{データ} = \text{信号} + \text{ノイズ}$$

という構成だと考えられます。

データを解析して卒業論文に貼り付ける前に、まず、このノイズを取り除きます。

これは例えば、米を炊飯器にセットする前に米を水洗いして、小さなゴミなどを取り除くと同じです。

ではここで、ノイズを含む信号から、ノイズだけを除去する方法について考えます。

まず、皆さんに簡単な質問をします。

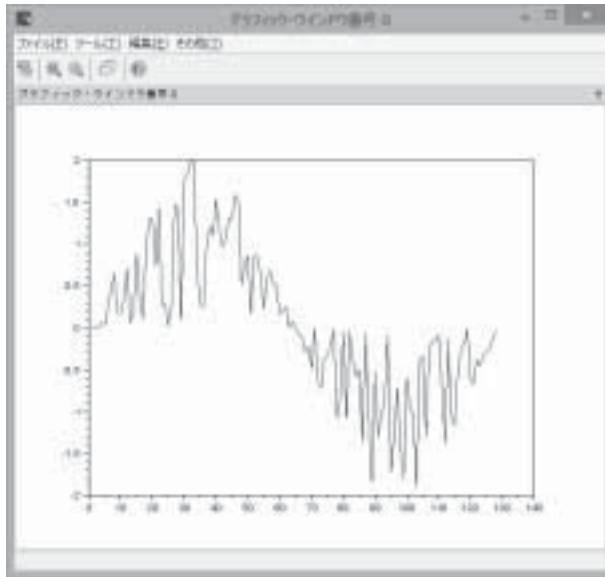
二つの数 x と y があります。

x と y を加算すると10になります。

x と y はいくつですか。

この質問に答えることはできません。

$x = 1$, $y = 9$, これはOK, ところが, $x = 2$ で, $y = 8$ でもOKです。問題を満足する答えは無限に存在します。



画面3.1
計測したデータの
分布

代数学の立場から言うと、

x と y に、解が存在するためには、他の条件が必要

となります。

例えば、 x は y の2倍、という条件を加えると、 $x = \frac{20}{3}$ 、 $y = \frac{10}{3}$ が解となります。これ以外の答えはありません。

計測したデータからノイズを除去して信号を抽出するためには、データあるいはノイズに関する情報が必要です。これを仮説 (hypothesis) と言います。

仮説は、データを採取する環境に依存します。一般論として記述することはできません。実験を行っている状況に応じて仮説を立てます。

例として、画面3.1に示すデータを得たとします。

データの傾向 (trend) は、しばらく増加して、その後に減少してマイナスに転じます。

データは、統計的なばらつき (statistical dispersion) を含みます。

ゆっくり変化するトレンドを「信号」とみれば、ばらつきは「ノイズ」です。

逆に、ゆっくり変化するトレンドを「ノイズ」とみれば、ばらつきは「信号」です。どち

第4章

フーリエ解析

4.1 —はじめに

フーリエ解析 (Fourier analysis) によって、波形を周波数に分離するアルゴリズムを述べます。

4.2 —最小2乗法

家を建てるときに、まず、コンクリートを流し込んで基礎を作ります。

フーリエ解析の基礎は、最小2乗法です。

最小2乗法に関しては、第3章において述べました。

ここでは、最小2乗法を深く掘り下げて、その上にフーリエ解析というビルを建築します。

最初に、小さなサンプルを使って、最小2乗法のアウトラインを示します。

いま、時刻 x_j における変量 y_j を測定します。測定は、 n 回実施しました。

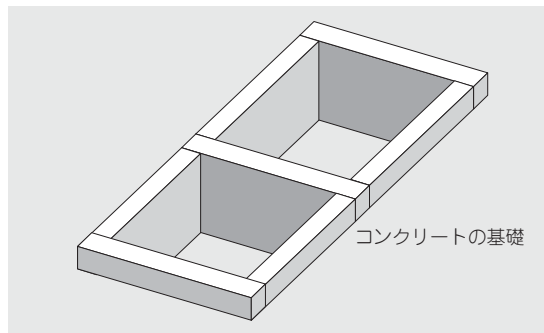


図4.1
基礎を作る

時刻と変量のペアを,

$$(x_j, y_j) \quad (j=1,2,\dots,n) \quad \dots\dots\dots (4.1)$$

と書き, 以下, 観測データと呼びます.

最小2乗法は, 観測データに対して, 特定の関数を当てはめます.

3.5節 (p.78)において, 計測データに対して, 関数,

$$y=ax^2+bx+c \quad \dots\dots\dots (4.2)$$

を当てはめました.

この関数をここでも取り上げます.

(4.1)式に, 最小2乗法を適用して, 未定係数 a, b, c を計算すると, (3.5)式 (p.78) になりました.

(3.5)式を再記すると,

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum x^4 & \sum x^3 & \sum x^2 \\ \sum x^3 & \sum x^2 & \sum x \\ \sum x^2 & \sum x & n \end{bmatrix}^{-1} \begin{bmatrix} \sum yx^2 \\ \sum yx \\ \sum y \end{bmatrix} \quad \dots\dots\dots (3.5)$$

となります.

まず, 右辺の第1項の行列に注目します.

この行列を抜き出すと,

$$\begin{bmatrix} \sum x^4 & \sum x^3 & \sum x^2 \\ \sum x^3 & \sum x^2 & \sum x \\ \sum x^2 & \sum x & n \end{bmatrix}$$

となります.

この行列を書き換えると,

$$\begin{bmatrix} \sum x^2x^2 & \sum x^2x^1 & \sum x^2x^0 \\ \sum x^1x^1 & \sum x^1x^1 & \sum x^1x^0 \\ \sum x^0x^0 & \sum x^1x^0 & \sum x^0x^0 \end{bmatrix} \dots\dots\dots (4.3)$$

となります.

対比を明確にするために,

$$\begin{aligned} 1 &= x_0 \\ n &= \sum 1 \times 1 = \sum x^0x^0 \end{aligned}$$

としました.

ここで,

$$\begin{aligned} f_1 &= x^2 \\ f_2 &= x \quad \dots\dots\dots (4.4) \\ f_3 &= 1 \end{aligned}$$

と置くと, (4.1) 式は,

$$y = af_1 + bf_2 + cf_3 \quad \dots\dots\dots (4.5)$$

となります.

(4.4) 式を使って, (4.3) 式の係数行列を書き換えると,

$$\begin{bmatrix} \sum f_1f_1 & \sum f_1f_2 & \sum f_1f_3 \\ \sum f_2f_1 & \sum f_2f_2 & \sum f_2f_3 \\ \sum f_3f_1 & \sum f_3f_2 & \sum f_3f_3 \end{bmatrix} \dots\dots\dots (4.6)$$

第5章

ウェーブレット解析

5.1 —はじめに

観測データに対して、ウェーブレット解析 (Wavelet Analysis) を行います。

5.2 —ウェーブレットとは

いまから100年以上も前、1909年ハンガリー生まれのHaarは、一つの論文を書き上げます。この論文が、現在ウェーブレットの始まりと位置付けられています。

しかし、当時は誰一人(おそらく本人も含めて)、この論文の価値を認める人はいませんでした。

それからおよそ80年が経過して、ウェーブレット解析の研究が爆発的にスタートします。

ウェーブレットの研究がスタートすると、これまでは大洋の離れ小島のような存在だったフーリエ解析が、実は、

線形変換に属する一つのツール

であることが明らかになります。

しかし、仲間が増えたからと言って、

フーリエ解析の価値は低下する

などと主張するものではありません。

コロンブスは、大西洋を横断してサンサルバドル島を発見します。

しかし、その背後に、巨大なアメリカ大陸があったのです。

フーリエ変換は、前章において述べたように、時間区間いっばいにsinあるいはcosの波を張ります。

このために、フーリエ変換は、時間空間 (time space) から離脱して周波数 (frequency) の空間へ移行します。

これに対してウェーブレットは、時間空間の波形を扱います。

時間と周波数に関連して、議論を進めます。

ここが、ウェーブレットの重要なポイントです。

これまでも述べてきましたが、フーリエ変換とウェーブレット変換は、ともに線形変換 (linear transform) に属します。

数学の記号を使って説明します。

いま、変数 w と y の間に、例えば、

$$w = 2y$$

という式が成立すると、

w は、 y の線形変換

と言います。 w は、 y の定数倍になっているからです。

ところが、

$$w = 2y + 8$$

となると、 w は、 y の線形変換ではありません。

定数8が加算されているからです。

一般化します。

変数は n 個で、 y_1, y_2, \dots, y_n があります。

この n 個の変数に対して、 n 個の定数を使って、

$$w = a_1 y_1 + a_2 y_2 + \dots + a_n y_n \quad \dots\dots\dots (5.1)$$

とするならば、 w は y_1, y_2, \dots, y_n の線形変換と言います。

ここで、 a は定数 (例えば、12.3, ..., など) でなければいけません。これは、絶対の条件です。

(5.1) 式の w が n 個あると、

$$\begin{aligned} w_1 &= a_{11}y_1 + a_{12}y_2 + \dots + a_{1n}y_n \\ w_2 &= a_{21}y_1 + a_{22}y_2 + \dots + a_{2n}y_n \quad \dots\dots\dots (5.2) \\ &\dots \\ w_n &= a_{n1}y_1 + a_{n2}y_2 + \dots + a_{nn}y_n \end{aligned}$$

となります。

通常、(5.2) 式は、行列を使って、

$$\begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ w_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad \dots\dots\dots (5.3)$$

と書きます。

以下、(5.3) 式の右辺の行列を**変換行列** (transform matrix) と呼びます。

変換行列が、

正規直交行列 (orthonormal matrix)

の条件を満足すると、 w と y の構造 (あるいは、相対的な位置関係) は変化しません。同じです。

w は、 y を別の角度から見たものです。

```

-> exec:
  #
  #
  #
-> exec: [1, 1, 1, 1, 1, 1, 1, 1, 1]
  #
  #
  #
-> exec: ["test.exe"]

```

画面5.35
プログラムの使用例

```

-> exec: ["C:\Users#skawa#Documents#SciLab#test.sce", -1]
警告: ファイル "C:\Users#skawa#Documents#SciLab#test.sce" は既に SciLab で開かれています。

#
  0.4829828  0.7071068  0.  0.  0.  0.  0.  0.
  0.8985183  0.  0.7071068  0.7071068  0.  0.  0.  0.
  0.2241435  0.  0.  0.  0.7071068  0.7071068  0.  0.
- 0.1294035  0.  0.  0.  0.  0.  0.7071068  0.7071068
  0.7071068 - 0.7071068  0.  0.  0.  0.  0.  0.
  0.  0.  0.7071068 - 0.7071068  0.  0.  0.  0.
  0.  0.  0.  0.  0.7071068 - 0.7071068  0.  0.
  0.  0.  0.  0.  0.  0.  0.7071068 - 0.7071068

#
- 0.1294035
- 0.2241435
  0.8985183
- 0.4829828

details:
  0.7071068
- 1.11001E
- 0.7071068
  1.11001E

approximation:
- 1.2247449
  1.4142138
  1.2247449
- 1.4142138

```

画面5.36 結果のプリントアウト

db2のレベル1の計算を行います。

参考文献 (1) のリスト4.2 (p.74~76) には、db2を任意のレベルに関して計算するプログラムが掲載されています。

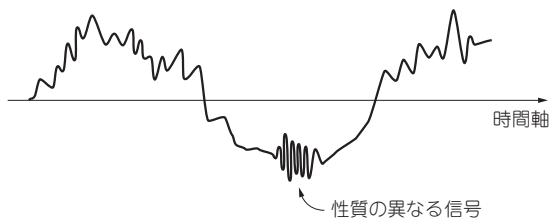


図6.5
性質の異なるデータ

コンピュータは、データをスキャンして「問題がない」と判断できれば、そのデータをすべて削除します。

問題があると思われる場所を絞り込んで、これを出力します。

専門家の負担は、大きく減ります。

時間的な負担が減れば、それだけ真剣にデータの処理に取り組みます。

問題は、どのようにしてデータを絞り込むかです。

まず、観測したデータがあります。

データの量は、膨大です。

しかし、ほとんどのデータは同じデータの繰り返しです。

データの一部に、性質の異なるデータが混入している可能性があります。

全体とは異なる部分があれば、それを抽出するプログラムを作ります。

これが、ここでの課題です。

6.4——特徴の抽出と決定

観測データから、異質な部分を抽出するアルゴリズムについて考察します。

いま仮に、観測データが100万個あるとします。

このデータに対して、第4章において述べたフーリエ解析を適用すると、100万個の複素数が算出されます。複素数は2個の実数からなるので、実数で言えば200万個の数になります。

もし、何らかの理由によって、例えば、周波数が100kHz、……、150kHzを残して、その他は捨てることにすれば、データの数は一気に複素数50個に減ります。

1,000,000 → 50

CQ出版社



波形解析のための数値計算ソフト

Scilab

入門

信号のスペクトラム、ノイズ分析から特徴抽出まで

- 卒論やレポートを作成するとき、実験や観測で採取したデータを集計し、グラフ化してレポートに貼り付けたりしますが、もっとも重要なポイントは、グラフの中に何を見つけるか、数の並びから何を引き出すか、ここが大事なポイントです。
- 本書は、実験データを数学的テクニックにより波形の特徴抽出するコツをまとめました。ツールとして、オープン・ソースとして無料で公開されているScilabという数値計算ソフトウェアを使います。
- ヒストグラム/移動平均/曲線の当てはめ/最急降下法/ニュートン・ラフソン法/ピラミッド・アルゴリズム/パラメトリック法/非パラメトリック法/ヒューリスティック・アルゴリズム

このPDFは、CQ出版社発売の「波形解析のための数値計算ソフト Scilab入門」の一部見本です。

内容・購入方法などにつきましては以下のホームページをご覧ください。

内容 <http://shop.cqpub.co.jp/hanbai/books/49/49501.htm>

購入方法 <http://www.cqpub.co.jp/order.htm>