

エクセルを使ったグラフの描き方

山形大学理学部物質生命化学科 天羽優子

2005 年 4 月 12 日 火曜日

1 はじめに

学生実験のレポートで、実験データのグラフを作るのに、エクセルを使う人が多い。ところが、エクセルのグラフは、それなりに手を入れないと、実験データを正しく表現するグラフにならない。昨年、レポートを見た限りでは、これなら手書きでグラフ用紙に描いた方がなんぼかマシというものがしばしばあった。自然科学での——物理化学や分析化学での——グラフの描き方には、実は「お作法」があるのだが、そのことを全くわかっていないとしか思えない。おそらく、情報処理の授業などで「エクセルで数値からグラフを描く」方法は習っているが、「自然科学のレポートで要求されるグラフの描き方」を学ぶ機会が無かったものと思われる。

そこで、どういうグラフが良くないのか、グラフはどう描くと良いのかについてまとめておくことにする。

2 悪いグラフの例

2.1 折れ線を使うな

図 1 は、ショ糖溶液の閃光度を測定する実験のレポートに添付されたものである。

元のグラフは、A 4 用紙横いっぱい印刷されていた（グラフの背景のムラは印刷物では無かったが、スキャナで取り込んだときにできたものである）。このグラフの問題点は、

1. グラフの背景をグレーで塗りつぶしている。自然科学の論文やレポートでは、グラフの背景に色をつけることはない。
2. 「8 %ショ糖溶液」というキャプションがグラフの上部にある。グラフのタイトルをグラフの外側の上部に書くことは通常ない。どうしても書きたければ、グラフの中に入れること。同時にグラフの下部にキャプションを付ける。
3. グラフで使った記号の説明 (legend) がグラフの外にあり、「系列 1」となっている。グラフの中に 1 種類しか記号がなければ、legend を付ける必要はない。また、一般に、legend をつける場合はグラフの中に入れる。
4. 横軸の目盛りは、通常は、グラフの一番下に入れる。
5. 折れ線グラフにしてはいけない。

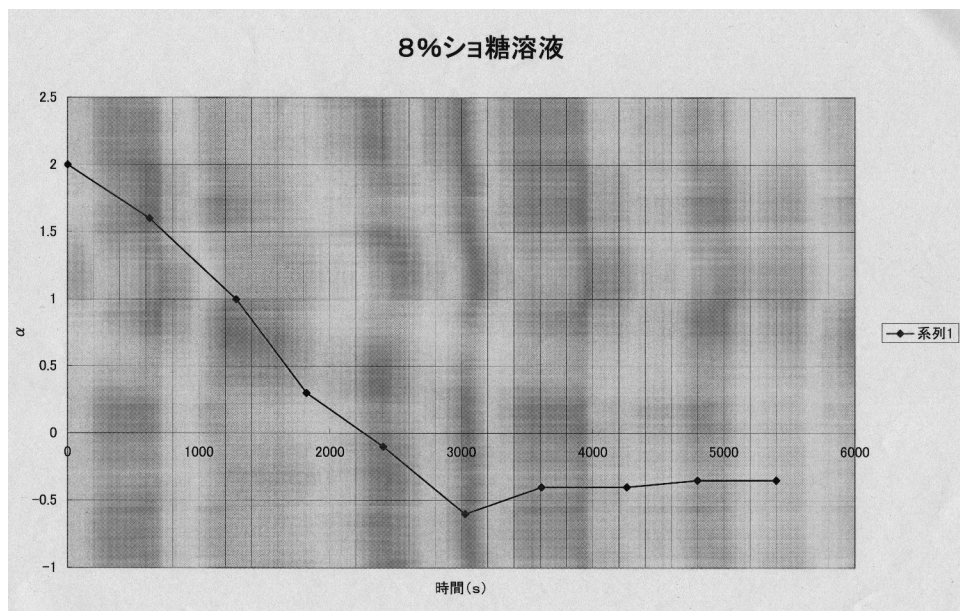


図 1: ショ糖溶液の旋光度を測定する実験の例

5 番目については、背景の説明が必要だろう。

この実験では、ショ糖を酸触媒で加水分解し、ブドウ糖と果糖ができる過程を観測している。ショ糖とブドウ糖と果糖ではそれぞれ比旋光度が異なり、旋光度が濃度に対して比例しかつ加成性が成り立つ。反応速度論から、各時間での旋光度 α_t と無限時間における α_∞ とすると、 $\ln(\alpha_t - \alpha_\infty)$ と時間の間に直線関係が成り立つ。つまり、 α_t は、測定誤差を除いて指数関数的に振る舞うことがわかっている。変化の仕方（＝関数型）があらかじめわかっているようなデータを折れ線で結ぶのは、何の意味もない。測定点だけを記載するか、測定点と共に指数関数による最小自乗フィットを行った結果を一緒に示すか、どちらかであろう。

2.2 わかりにくいフィッティング

図 2 も、ショ糖の加水分解のデータ処理例である。 $\ln(\alpha_t - \alpha_\infty)$ と時間の間に直線関係が成り立つのは直前に説明した通りである。そこで、測定データより $\ln(\alpha_t - \alpha_\infty)$ を求め、時間に対してプロットした後、各濃度のデータに対して直線でフィッティングを行い、傾きから速度定数を求めることになる。このためのグラフである。

問題点は次の通り。

1. 測定データから計算された $\ln(\alpha_t - \alpha_\infty)$ の値がマーカーで示されていない。
2. $\ln(\alpha_t - \alpha_\infty)$ の値を結ぶ折れ線だけが示されている。しかも、線が細くて見づらい。
3. 実は、レポートを見たところ、直線の傾きを求めるのに、時間 0 s のときと時間 6000 s の時の値を読んで計算で求めていた。直線でフィッティングしたのなら、こんなことをしなくても傾きが出るはずだが……？
4. 横軸がグラフの中にある・legend がグラフの外にある（これは直前のグラフの問題点と共通）。

5. レポート内でキャプションと図番号が付いていない。必ずつけるべき。

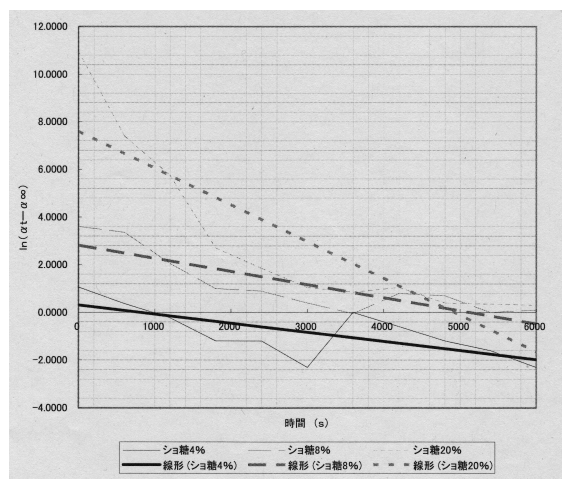


図 2: データから最小2乗法によって直線の傾きを求める処理のはずだが？

2.3 安易に曲線で結ぶな

次ぎは、反応熱の実験結果の処理例。測定点を示した後、間を曲線で結んでいる。

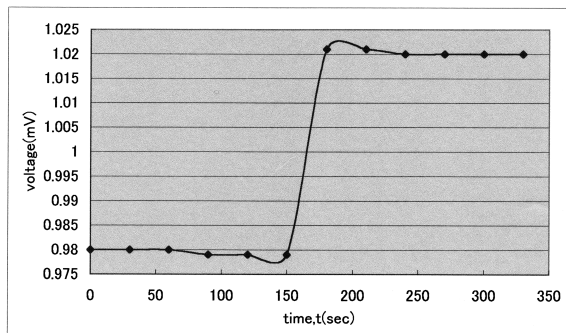


図 4 電圧を加えたとき（温度を上昇させたとき）の時間ごとの起電力変化（2回目）。

図 3: 反応熱の実験より。電圧を加えて温度を上昇させたときの時間後との起電力変化。

測定点の間はなめらかな曲線で結ばれている。時間 150 秒の直前と、データが立ち上がった直後の点と点の間で、曲線はそれぞれ下に凸、上に凸になっている。この振る舞いには全く意味がない。もっと細かくデータを取ったとして、このように、オーバーシュートが起きる理由はどこにもない。

この曲線を描く時に、おそらく、データをスプライン補間したものと思われる。スプライン補間というのは、点と点の間をそれらしい曲線で結びつつ、それぞれの区間をなめらかに（つまり微分係数が一致するように）つなぐ方法の 1 つである。微分係数を合わせるために、急激な変化の前後で、いかにもそれらしい下に凸や上に凸の曲線が出てくるといふ癖がある。しかし、点で区切られた区間ごとに高次の多項式で作った曲線をつないでいるだけなので、現実には起きている物理

化学現象には何の関係もない曲線が引かれることになる。その結果、実際の系の振る舞いとして、オーバーシュートが起きると勘違いしてしまう可能性が高い。自然現象の中には、本当にオーバーシュートする現象もあるが、この実験ではそうではないから、よけいにややこしい。このような場合は、点だけを示して、曲線で結ぶべきではない。

なお、グラフの中を塗りつぶしたり、グラフの外側に枠が描かれているのが良くないのは、前に示したグラフと同様である。

2.4 測定で得られたのは一体どこか？

今度は、曲線だけ描いてあってデータ点がわからない。

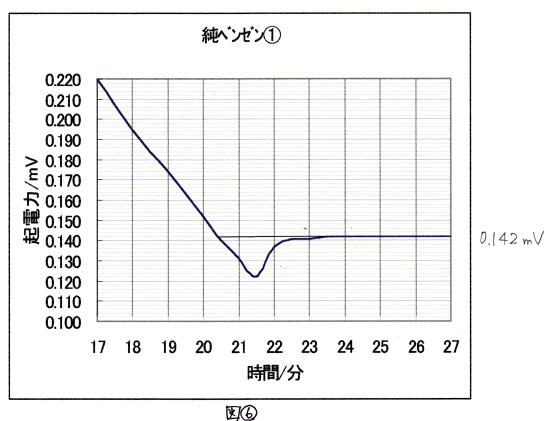


図 4: 凝固点降下のデータ処理例

この実験では、データを 30 秒おきに記録することになっている。グラフをみると、折れ線ではないので、実際の測定点を示さず、点と点の間を補間したのではないかと思われる。このため、微妙にきれいな曲線で結果が出ているが、どこが測定結果なのか、どこが造られたデータなのか、はつきりしない。やはり、測定した点を明確に示し、余計な曲線を引くべきではない。

図のタイトルが図の上についているとか、図全体に不必要な枠を付けているのは、前に示したのと同様の良くない描き方である。

2.5 横軸の描き方が悪い例

図 5 も一次反応のデータ処理例の 1 つである。この実験では、時間に対して指数関数で減少するデータが得られる。速度定数を求めるために、得られたデータの対数を片対数でプロットすると、元が指数関数であるため直線的に減少するデータが得られる。

問題は横軸の表示位置と表示方法である。時間を、hh:mm:ss 型式で書いているが、通常、時間の単位は、秒、分、時、のどれか 1 つを選んで使う。この実験であれば、秒で統一するのが適切だろう。

横軸の位置も悪い。縦軸の 0 に合わせて横軸を描いたため、横軸の目盛りがグラフの中にきてしまった上、縦軸の目盛りに一部重なっている。こんなグラフは絶対にまともな報告書や論文には出せない。時間の単位を統一し、横軸はグラフの一番下に出すべきである。

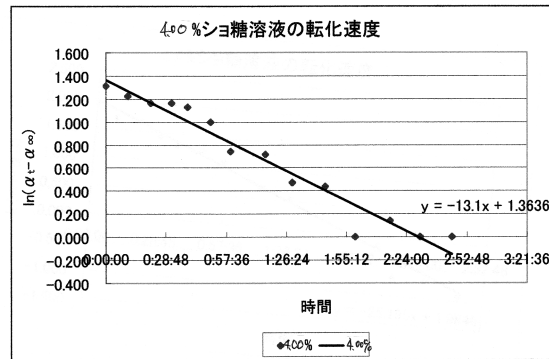


図2. 4.00% ショ糖溶液の転化速度

図 5: 一次反応のデータ処理例。フィッティングの場合

さらに、グラフの説明として、マーカーと直線の両方に、4.00%と手書きされている。この説明は、グラフの中の空いた場所に移動させ、マーカーには「測定データ (measured)」、直線には「フィッティングの結果 (fitted line)」とでも書いて、測定によって得られたものとフィッティングの結果をそれぞれ明示するべきである。

グラフ上部に印刷されたタイトルは削除し、グラフを囲む枠も削除すべきであることは言うまでもない。ただし、手書きであっても図番号とキャプションを付けているのは、正しい書き方であるといえる。

3 正しいグラフの描き方

3.1 折れ線無し、変化を反映

Excel で描いた望ましいグラフの一例を、図 6 に示した。

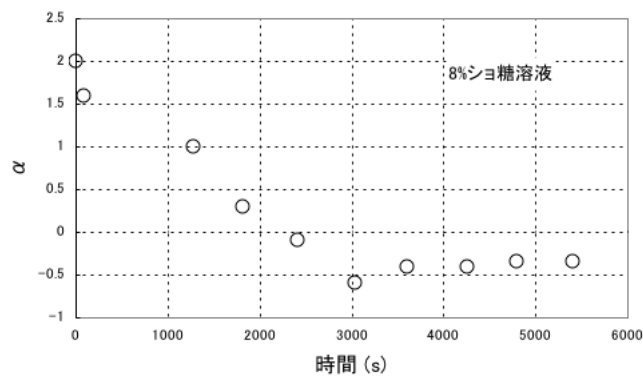


図 6: ショ糖溶液の旋光度を測定する実験の例

レポートにするときは、グラフの下に図番号とキャプションを入れること。

4 補間とフィッティングは意味が違う

測定の結果（たまには測定以外の場合でも）得られるデータはとびとびのものである。限られたデータから情報を引き出す方法の代表的なものに、補間とフィッティングの2通りのやり方がある。

補間は、点と点の間をうまくつないで、実際には測定していないところのデータを推定する方法である。得られたデータ全体を一度にあらわすのは難しいので、部分毎に多項式などをつかって、うまく曲線を引く。代表的なものに「ラグランジュ補間」「スプライン補間」がある。

フィッティングは、データを表すモデル（関数の形、たとえば直線だとか指数関数であるとか対数関数であるとか）が決まっているときに、関数を与えて、できるだけデータをうまく再現するようにパラメータを決める方法である。データが直線にのる場合は、 $y = ax + b$ という直線を仮定し、 a 、 b の値を変えて、直線がデータの最も近くを通るようにする。直線とデータのずれの2乗の和が最小になったとき、直線がデータ全体の一番近いところを通ったことになる。 a 、 b を計算で求める方法を、最小自乗法という。直線の場合、この計算は、電卓があれば、少し手間をかけるだけでできる。

指数関数の場合は、たとえば、 $y = a \exp(bx)$ といった形を仮定し、 a 、 b を変化させて、データ点に一番近いところを通るようにする。関数が曲線の時は、電卓で簡単に計算することはできないが、すでに計算方法は知られている。非線形最小自乗法の1つである、Levenberg-Marquardt 法がよく使われる。

化学反応の速度など、もともとどういう関数型で振る舞うかが分かっている場合や、使うべきモデルが決まっている場合は、最小自乗法によるフィッティングを行う。

得られたデータの振る舞いのモデルがそもそも無い場合に点と点の間の値を推定したいときは、補間を用いる。逆にすると、無意味であるし、データ処理として明らかな間違いでもある。

Excel では（多分他のソフトも）、補間やフィッティングの機能が最初から組み込まれている。どちらを使うのが妥当かは、手持ちのデータの意味をよく考えて判断すべきである。安易にデータ点の間を埋めて、勝手なデータをつくってはいけない。

数値計算の基礎的な話は、「理工系の基礎数学 8 数値計算」高橋大輔著（岩波書店）が読みやすい。本格的に使うのであれば、Numerical Recipes in C (Cambridge University Press) を勧める。市販の主なデータ処理ソフトで使われている計算方法のほとんどが網羅されている。