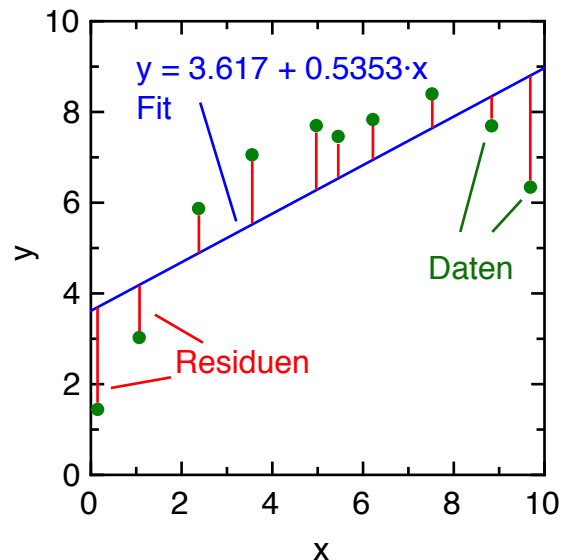


Kurvenanpassung

Verwandte Begriffe: Ausgleichsrechnung, Regressionsfunktion, Fit, (“Trendlinie” in Excel™)

Das Ziel dieser mathematischen Methode ist es, einen Satz numerischer Daten (x_i, y_i) ; $i = 1, \dots, N$ möglichst gut durch eine Funktion zu beschreiben. Zuerst muss eine Funktion mit freien Parametern gewählt werden, z.B. ein Polynom ersten Grades $y = ax + b$ mit den freien Parametern a und b . Dann berechnet der Computer die Zahlenwerte der Regressionsparameter a und b , siehe Abb. 1128.

Abbildung 1128: Fit einer Geradenfunktion $y = ax + b$ an $N = 10$ Datenpunkte
 Der Computer berechnet die Parameter a sowie b und zeichnet den Funktionsgraphen.
 Meistens wird die “Methode der kleinsten Quadrate” verwendet, um die optimalen Regressionsparameter zu finden. Bei dieser wird die Summe der Quadrate der Residuen minimiert. Ein Residuum $r_i = y_i - (ax_i + b)$ ist der Unterschied zwischen der Ordinate y_i des Datenpunktes Nummer i und dem Funktionswert $ax_i + b$ für die dazu gehörende Abszisse x_i . Die Optimierungsbedingung ist dann $r_1^2 + r_2^2 + \dots + r_N^2 = \min$.



Die meisten Tabellenkalkulationsprogramme stellen folgende Ausgleichsfunktionen zur Verfügung:

Linear	$y = ax + b$	Option: $b = 0$
Potenziell	$y = a \cdot x^b$	
Exponentiell	$y = a \cdot e^{bx}$	
Polynomisch	$y = a + bx + cx^2 + \dots$	Option: $a = 0$

Im verbreiteten Tabellenkalkulationsprogramm Excel™ werden die Parameter der Regressionsfunktion berechnet, indem man die Daten im Diagramm markiert und dann “Trendlinie einfügen” auswählt. Anschliessend werden der Funktionstyp sowie die weiteren Optionen selektiert.

Die Funktionsgleichung mit den berechneten Regressionsparametern soll man sich vom Computer im Diagramm anzeigen lassen. Die Darstellung soll so formatiert werden, dass möglichst viele Stellen der Regressionsparameter angezeigt werden.

Die Zahl der Fitparameter (a, b, \dots) einer Ausgleichsfunktion darf die Zahl der Datenpunkte (N) nicht übersteigen. Meistens wird aber eine möglichst einfache Beschreibung gesucht, i.e. eine Fitfunktion, die mit einem Minimum an Parametern die Daten korrekt wiedergibt.

Die Regressionsfunktion soll – sofern vorhanden – mit Blick auf eine Theorie gewählt werden. Wird z.B. ein proportionaler Zusammenhang erwartet, so wählt man $y = ax$ als Fitfunktion und nicht $y = ax + b$ oder gar eine Exponentialfunktion.

Die Genauigkeitsanalyse erfolgt in zwei Schritten:

- a) Mit der “Jackknife-Methode” wird die Anzahl wesentlicher Ziffern der Fitparameter bestimmt.
- b) Mit den Residuen wird die Passung zwischen Ausgleichsfunktion und Daten diskutiert.

Genauigkeit der Regressionsparameter

Der Computer berechnet die Parameter einer Ausgleichsfunktion fast beliebig genau. Da aber Messwerte Fehlerschranken aufweisen, pflanzt sich diese auf die gefitteten Parameter fort.

Die Anzahl signifikanter Stellen lässt sich abschätzen, wenn einzelne Punkte aus dem Datensatz gelöscht werden. Jene Ziffern der Fitparameter, die sich dabei nicht oder kaum ändern, sind signifikant. Dieses Vorgehen heisst “Jackknife-Methode”: Analyse mit einer Teilmenge des Datensatzes.

Beispiel: Tabelle 68 enthält Strom- und Spannungsmesswerte von einem (nominell) $330\ \Omega$ Widerstand. Eine Messreihe wurde mit Gleichstrom aufgenommen, die andere mit 50 Hz-Wechselstrom. In Abb. 1129 entsprechen sich die Messpunkte. Die zwei gezeichneten Ausgleichsfunktionen (Proportionalitäten, Nullpunktsgerechten) liegen im Rahmen der Zeichengenauigkeit übereinander. Der Fitparameter ist gerade der Widerstand ($y = ax \Leftrightarrow U = RI$). Der Computer liefert unterschiedliche Zahlen für die Regressionsparameter der zwei Messreihen. Ist der Unterschied bedeutsam oder nur “digitaler Mist”?

U_{AC} (V)	I_{AC} (mA)	U_{DC} (V)	I_{DC} (mA)
0.228	0.70	0.047	0.16
0.593	1.80	0.585	1.79
1.021	3.11	1.26	3.85
1.555	4.73	1.752	5.32
2.004	6.09	2.119	6.45
2.603	7.93	2.752	8.38
3.116	9.60	3.388	10.3
3.636	11.06	4.181	12.71
4.302	13.11	4.955	15.08
4.80	14.60	5.659	17.23
5.30	16.14	6.048	18.43
5.64	17.18	6.306	19.22
6.281	19.15	0.267	0.82
± 0.01	± 0.02	± 0.01	± 0.02

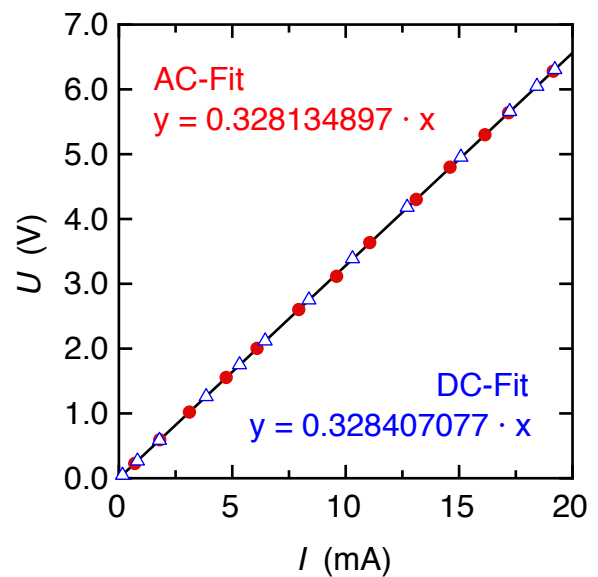


Tabelle 68: Zwei Strom-Spannung Messreihen an einem Widerstand, gemessen mit zwei Digitalmultimetern. Eine Messung erfolgte mit Wechselspannung (AC, Effektivwerte), die andere mit Gleichstrom (DC). Das Multimeter zeigte $329.6\ \Omega$ für das Widerstandselement an. 22. März 2017, Lie.

Abbildung 1129: Strom-Spannung-Kennlinien eines $330\ \Omega$ Widerstands gemessen mit Wechsel- und Gleichstrom. Die Graphen der zwei Ausgleichsfunktionen ($x \hat{=} I$ in Milliampere, $y \hat{=} U$ in Volt) liegen im Rahmen der Zeichengenauigkeit übereinander.

Werden einzelne Datenpunkte von der Regressionsrechnung ausgeschlossen, so erhält man für Wechselstrom die Fitparameter 0.3281 3490, 0.3281 3561, 0.3283 4398, 0.3281 7619 und 0.3280 9345. Es sind höchstens vier Ziffern signifikant, denn die vierte schwankt um drei Einheiten. Als vernünftigen Wert könnte man 0.3281 angeben. Für Gleichstrom erhält man z.B. 0.3284 0708, 0.3284 0761, 0.3284 9620, 0.3283 7123 und 0.3284 9513. Hier müsste man 0.3284 runden. Da die letzte Ziffer nicht sicher ist, stimmen die Werte im Rahmen der Messgenauigkeit überein. Aus 0.3283 folgt ein Widerstandswert von $328.3\ \Omega$.

Eine vornehmere Methode variiert die Daten im Rahmen ihrer Fehlerschranken und beobachtet dabei die Schwankungen der Regressionsparameter (Fehlersimulation).

Residuen

Das Residuum (lat. Rest, Plural Residuen) ist das was übrig bleibt, wenn von einem Messwert ein berechneter Wert subtrahiert wird. Die Messwerte sind bei uns meist Datenpunkte (x_i, y_i) ; $i = 1, \dots, N$. Die berechneten Werte sind oft Ausgleichsfunktionen $f(x)$ oder theoretische Zusammenhänge. Die Residuen sind dann $r_i = y_i - f(x_i)$; $i = 1, \dots, N$. Je kleiner die Residuen sind, desto besser ist die Theorie oder der Fit.

Um die Güte der Kurvenanpassung oder der Theorie zu beurteilen, werden die Residuen separat nach dem eigentlichen Diagramm gezeichnet. Die Abszisse wird gleich skaliert, aber die Ordinate wird gespreizt. Auf diese Weise können auch kleine Fehlerbalken vernünftig gezeichnet werden (in Excel™ "Fehlerindikatoren" genannt). In Abbildung 1130 sind die Residuen gut verteilt, in Abbildung 1131 ungünstig.

Abbildung 1130: Gute Verteilung der Residuen
Die Residuen sind gleichmässig und zufällig um Null herum verteilt. Alle Residuen (oder zumindest die meisten) sind innerhalb der Fehlerschranken Null, d.h. alle Fehlerbalken kreuzen die Nulllinie.

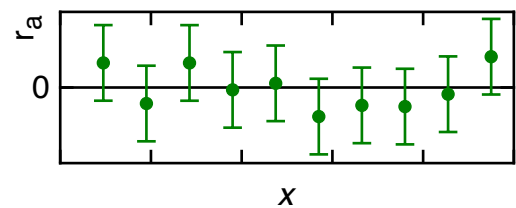
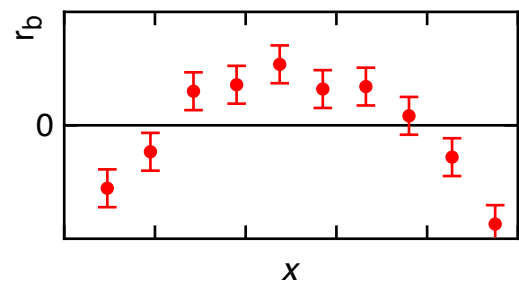


Abbildung 1131: Ungünstige Verteilung der Residuen
Zu viele Residuen unterscheiden sich im Rahmen der Fehlerschranken von Null. Die Verteilung der Residuen um die Nulllinie ist nicht zufällig, sondern folgt einem systematischen Trend.



Eine Theorie oder Ausgleichsrechnung, die Residuen wie in Abbildung 1131 zeigt, müsste überdacht werden. Vielleicht enthält die Messung einen systematischen Fehler, der Theorie fehlt ein wichtiges Element oder es ist eine ungünstige Ausgleichsfunktion gewählt worden.

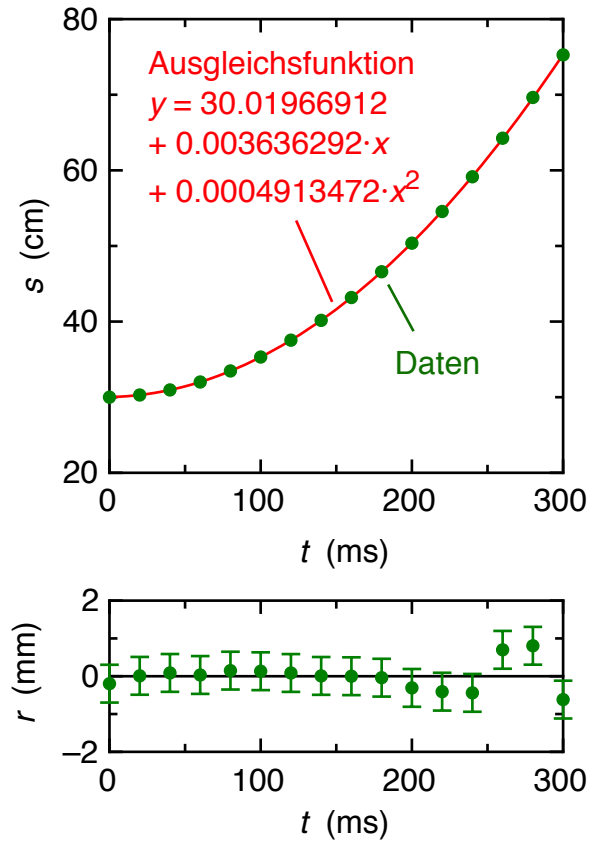
Berusster Stab (Muster Ausgleichsrechnung)

Ein vertikaler, berusster Metallstab wurde frei fallen gelassen. Der Stab fiel an einer gleichmässig rotierenden Scheibe vorbei. Ein mitdrehender Nylonfaden hinterliess in 20 ms-Abständen Spuren im Russ. Diese Spuren konnten durch eine Messlupe auf einem Massstab mit 0.01 cm Auflösung verortet werden. (1993, Lie.)

t (ms)	s (cm)
0	30.00
20	30.29
40	30.96
60	32.01
80	33.47
100	35.31
120	37.54
140	40.16
160	43.18
180	46.59
200	50.37
220	54.56
240	59.15
260	64.25
280	69.64
300	75.27

Tabelle 69: (links) Position s versus Zeit t für den freien Fall eines Stabes. Die Fehlerschranken der Messungen betragen 1 ms und 0.05 cm. Der Start war schwierig auszumessen: $t = 0$ entspricht nicht der anfänglichen Ruhelage.

Abbildung 1132: rechts oben: Position s während des freien Falls als Funktion der Zeit t nach Tabelle 69 mit Ausgleichsfunktion ($x \hat{=} t$ in Millisekunden, $y \hat{=} s$ in Zentimetern) rechts unten: Die meisten Residuen r des Fits stimmen im Rahmen der Fehlerschranken mit Null überein. Zwei Messungen scheinen Ausreisser zu sein.



Eine quadratische Regressionsfunktion wurde gewählt, weil eine solche für den freien Fall erwartet wird. Diese Wahl passt im Rahmen der Fehlerschranken zu den Daten, siehe Abbildung 1132. In der üblichen Schreibweise lauten die Regressionsfunktion und die Regressionsparameter

$$s = s_0 + v_0 t + \frac{1}{2} g t^2$$

$$s_0 = 30.02 \text{ cm}$$

$$v_0 = 0.0036 \text{ cm/ms} = 3.6 \text{ cm/s}$$

$$\frac{1}{2} g = 491 \text{ cm/s}^2 \Rightarrow g = 9.82 \text{ m/s}^2$$

Wird das letzte Datum aus Tabelle 69 gelöscht, so ändern sich der konstante Term in der vierten, der lineare Term in der zweiten und der quadratische Term in der dritten wesentlichen Ziffer. Entsprechend wurde auf diese Stelle gerundet. Im Rahmen der Fehlerschranke (einige Einheiten in der dritten wesentlichen Ziffer) stimmt die Fallbeschleunigung aus dem Fit mit dem Literaturwert in Zürich (9.806 m/s^2) überein.