

# 10. Großübung

## Nichtlineare Ausgleichsrechnung

### Allgemeines Problem:

Wir betrachten ein nichtlineares System

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

mit  $(m > n)$ , d.h. das System ist überbestimmt und

$$F_i(x) = g(t_i; x) - g_i \stackrel{!}{=} 0 \quad i = 1, \dots, m.$$

Die  $i$ -te Zeile  $F_i$  repräsentiert hierbei zumeist eine Messung aus einer Messreihe von  $m$  Messungen.

$g_i$  ist dabei der Messwert und  $g(t_i, x)$  das Ergebnis eines Modells an der Stelle  $t_i$  für noch *unbekannte* Modellparameter, die in dem Parametervektor  $x \in \mathbb{R}^n$  zusammengefasst sind.

### Beispiel 1 (Teil 1):

Wir betrachten eine einfache skalare Funktion. Das Modell sei  $g(t; \lambda) = e^{-\lambda \cdot t}$ , d.h. der Parametervektor ist eindimensional ( $n = 1$ ) und  $x = \lambda$ . Es wurden zwei Messungen ( $m = 2$ ) durchgeführt. An der Stelle  $t_1 = 1$  erhält man den Messwert  $g_1 = 0.8$ , an der Stelle  $t_2 = 2$  erhält man  $g_2 = 0.5$ . Das fasst man (in Hinblick auf komplexere Probleme) gerne in einer Tabelle zusammen:

$t_i$	1	2
$g_i$	0.8	0.5

In Abbildung 1 sieht man für verschiedene Werte für  $\lambda$  die Modellfunktion. Man beobachtet, dass es kein  $\lambda$  gibt, sodass beide Werte exakt getroffen werden können. Dies kann in der Praxis aus verschiedenen Gründen auftreten. Zum Beispiel könnten Messfehler vorliegen, die das Abweichen von dem Modell erklären oder aber das Modell ist nur eine Annäherung an die Realität und Abweichungen liegen im Rahmen der Ungenauigkeit des Modells vor.

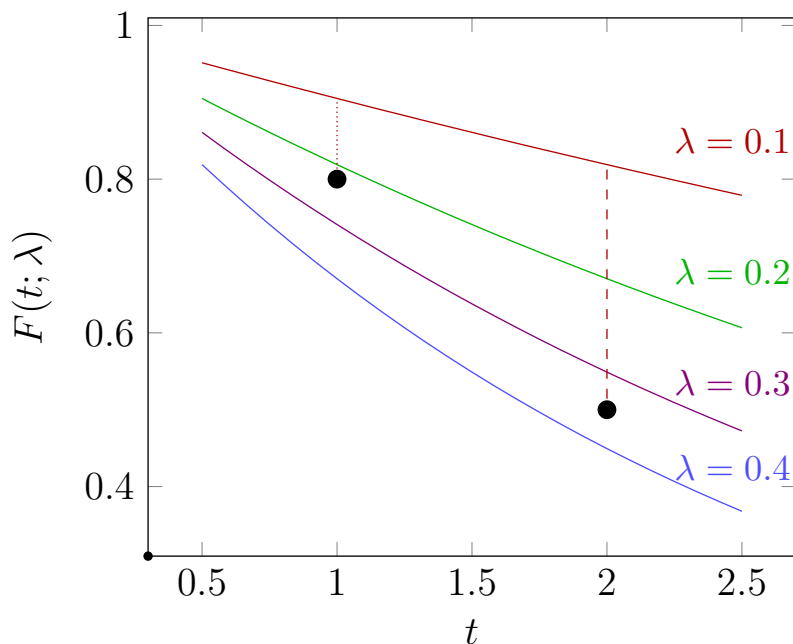


Abbildung 1: Messpunkte zu Beispiel 1 und Beispiele für die Modellfunktion für verschiedene  $\lambda$ -Werte.

### Formulierung des (nichtlinearen) Ausgleichsproblems:

Minimiere den Fehler, d.h. bestimme die Parameter  $x = (x_1, \dots, x_n)^T$ , so dass der Fehler minimal wird, d.h.:

$$\|F(x)\| \rightarrow \min_{x \in \mathbb{R}^n}$$

Das Problem ist leichter zu behandeln, wenn man die 2-Norm betrachtet, dann gilt

$$\sum_{i=1}^m (g(t_i; x_1, \dots, x_n) - g_i)^2 \rightarrow \min_{x \in \mathbb{R}^n}.$$

Den gesuchten Parameter bezeichnen wir mit  $x^*$  und es gilt

$$\|F(x^*)\|_2 = \min_{x \in \mathbb{R}^n} \|F(x)\|_2 \quad (1)$$

### Beispiel 1 (Teil 2):

Für jede Wahl von  $\lambda$  können wir einen Abstand auswerten. In Abbildung 1 sind die Abstände zu den Messpunkten für  $\lambda = 0.1$  (gestrichelt) angedeutet.

Die zugehörige Funktion  $F$  in diesem Beispiel ist

$$F(\lambda) = \begin{pmatrix} g(t_1; \lambda) - g_1 \\ g(t_2; \lambda) - g_2 \end{pmatrix} = \begin{pmatrix} e^{-\lambda \cdot 1} - 0.8 \\ e^{-\lambda \cdot 2} - 0.5 \end{pmatrix}$$

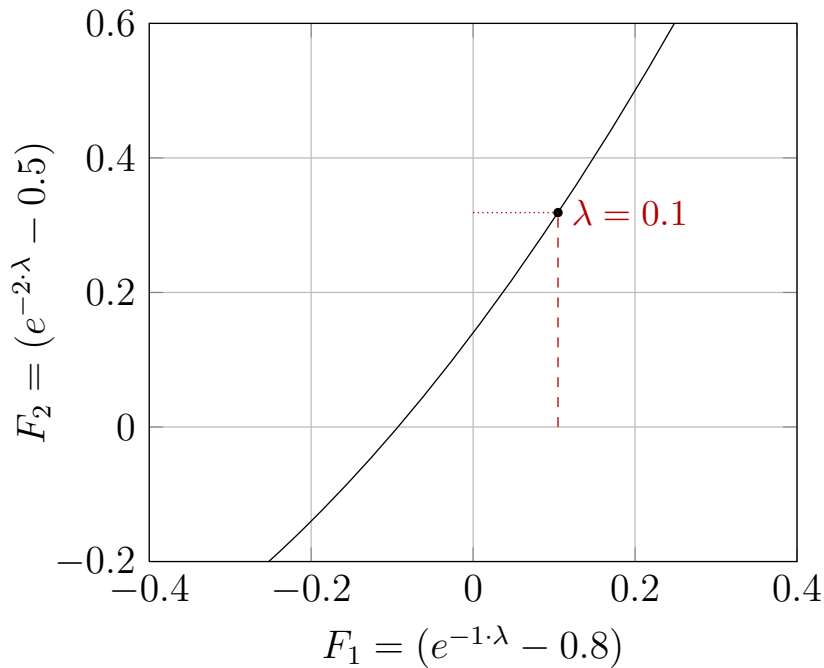


Abbildung 2: Funktionsverlauf von  $F$  in Abhängigkeit von  $\lambda$ . Die x-Koordinate beschreibt die Abweichung in der ersten Messung, die y-Koordinate die Abweichung in der zweiten Messung.

## Grundidee zur Lösung des nichtlinearen Ausgleichsproblems:

Wie löst man (1) ?

Ähnlich wie beim Newton-Verfahren für Gleichungssystem linearisiert man  $F(x)$  an einer gegebenen Stelle (aktueller Iterationswert). Taylorreihe:

$$F(x^k + \Delta x^k) \doteq F(x^k) + F'(x^k) \cdot \Delta x^k$$

Die linearisierte Funktion wird jetzt genommen, um ein *lineares* Ausgleichsproblem zu lösen:

$$\Rightarrow \|F(x^k) + F'(x^k)\Delta x^k\|_2 \rightarrow \min_{\Delta x^k \in \mathbb{R}^n}$$

In der Form eines linearen Ausgleichsproblem haben wir zumeist geschrieben

$$\Rightarrow \|Ax - b\|_2 \rightarrow \min_{x \in \mathbb{R}^n},$$

d.h. wir haben jetzt  $A \leftrightarrow F'(x^k)$ ,  $b \leftrightarrow -F(x^k)$  und  $x \leftrightarrow \Delta x^k$ . Löst man dieses lineare Ausgleichsproblem erhält man  $\Delta x^k$  und damit auch  $x^{k+1} = x^k + \Delta x^k$ . Dieses Verfahren ist das Gauss-Newton-Verfahren. Man beachte, dass die Jacobi-Matrix, die folgende Form hat ( $F'_i(x)$  ist die i-te Zeile der Jacobi-Matrix):

$$F'_i(x) = \left( \frac{\partial F_i}{\partial x_1}, \dots, \frac{\partial F_i}{\partial x_n} \right)$$

### Beispiel 1 (Teil 3):

Wir nehmen an, dass  $\lambda^0 = 0.3$  ein guter Startwert ist. Nun linearisieren wir  $F$  um  $\lambda^0$ :

$$\begin{aligned} \tilde{F}^1(\Delta\lambda) &= \begin{pmatrix} e^{-\lambda^0 \cdot 1} - 0.8 \\ e^{-\lambda^0 \cdot 2} - 0.5 \end{pmatrix} + \begin{pmatrix} -1 \cdot e^{-\lambda^0 \cdot 1} \\ -2 \cdot e^{-\lambda^0 \cdot 2} \end{pmatrix} \cdot \Delta\lambda \\ &= \underbrace{\begin{pmatrix} -0.05918 \\ 0.04881 \end{pmatrix}}_{=-b} + \underbrace{\begin{pmatrix} -0.7408 \\ -1.098 \end{pmatrix}}_{=A} \cdot \Delta\lambda \end{aligned}$$

$\tilde{F}^1(\Delta\lambda)$  beschreibt gerade eine Gerade, die der Tangente an dem  $F(\lambda^0)$  entspricht. Siehe dazu auch Abbildung 3. Auf dieser Geraden suchen wir den Punkt mit dem kleinsten Abstand zur Null. Rechnerisch lösen wir jetzt also das lineare Ausgleichsproblem (geschrieben mit der erweiterten Matrix  $(A|b)$ ):

$$\begin{pmatrix} -0.7408 & | & 0.05918 \\ -1.098 & | & -0.04881 \end{pmatrix},$$

Givens:  $r = 1.3245$ ,  $c = -0.5593$ ,  
 $s = -0.8290$ .

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} -0.7408 & | & 0.05918 \\ -1.098 & | & -0.04881 \end{pmatrix} = \begin{pmatrix} r & | & 0.007364 \\ 0 & | & 0.07636 \end{pmatrix}$$

Also ist  $\Delta\lambda^1 = 0.005560$  und damit  $\lambda^1 = 0.3056$ . Betrachten wir die Residuen:

$$\|F(\lambda^0)\|_2 = 0.07671, \quad \|F(\lambda^1)\|_2 = 0.07637$$

Wir sehen, dass das Residuum leicht abgenommen hat. Wir sehen aber auch, dass das Residuum nicht gegen Null geht. Das ist auch nicht zu erwarten,

da wir auch für die exakte Lösung ein Residuum größer Null haben werden. Wichtig ist auch anzumerken, dass das Residuum des linearisierten Problems (kann man nach der QR-Zerlegung einfach ablesen) nicht immer mit dem des nichtlinearen Problems übereinstimmen muss! Aber: Je näher man an das Minimum herankommt, um so stärker nähern sie sich an.

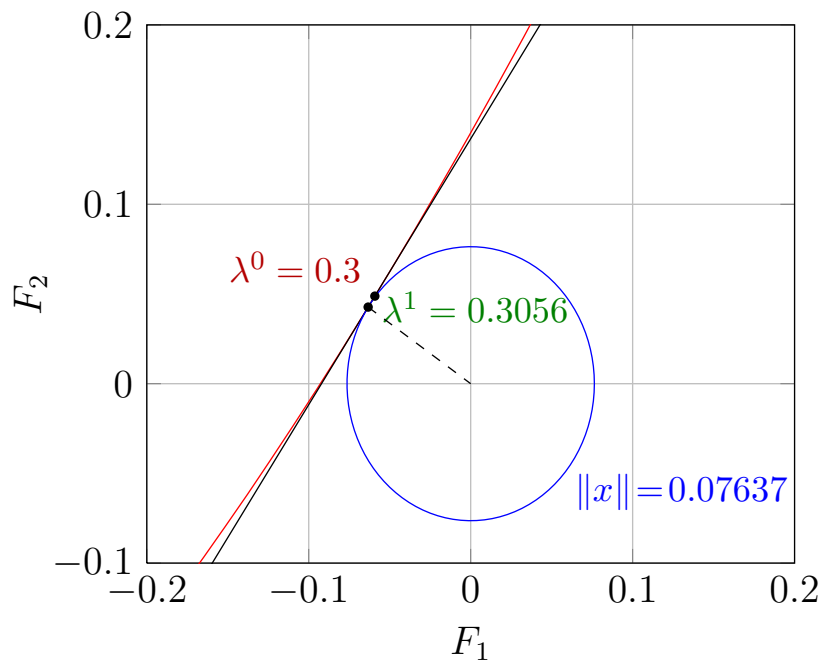


Abbildung 3: Linearisiertes Problem (Gerade) durch  $\lambda^0 = 0.3$  und Punkt mit dem minimalen Abstand auf dieser Geraden.

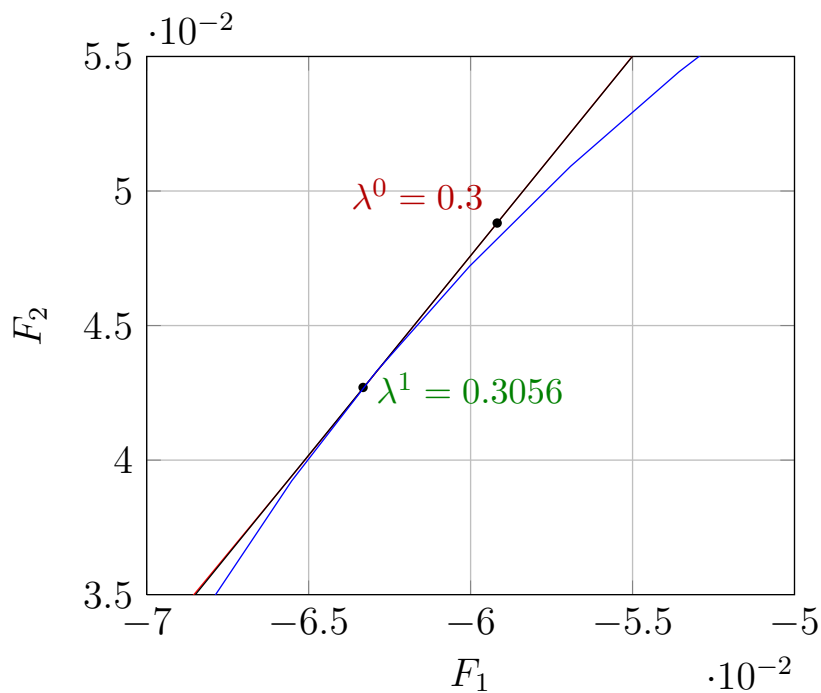


Abbildung 4: Zoom-In von Abbildung 3

## Beispiel 2:

Wir betrachten nun das Modell

$$g(t) = A \cdot (1 - e^{-\lambda t})$$

das von den Parametern  $A$  und  $\lambda$  abhängt. Diese fassen wir zusammen zu  $x = (A, \lambda)^T \in \mathbb{R}^2$ . Gemessen wurden vier Messwerte, die in der folgenden Tabelle angegeben sind.

$t_i$	0.1	1	3	5
$g_i$	1	2	3	4

Gesucht ist  $x = (A, \lambda)^T \in \mathbb{R}^2$ , sodass die Kurve  $g(t)$  möglichst nah an den Messwerten  $g_i$  liegt. In Abbildung 6 sind die Messpunkte und Beispielmmodelle geplottet.

### Formulierung der Aufgabe:

Unsere Aufgabe ist nun: Finde  $x^* = (A^*, \lambda^*)^T \in \mathbb{R}^2$  mit

$$\|F(x^*)\|_2 = \min_{x=(A,\lambda)^T \in \mathbb{R}^2} \|F(x)\|_2$$

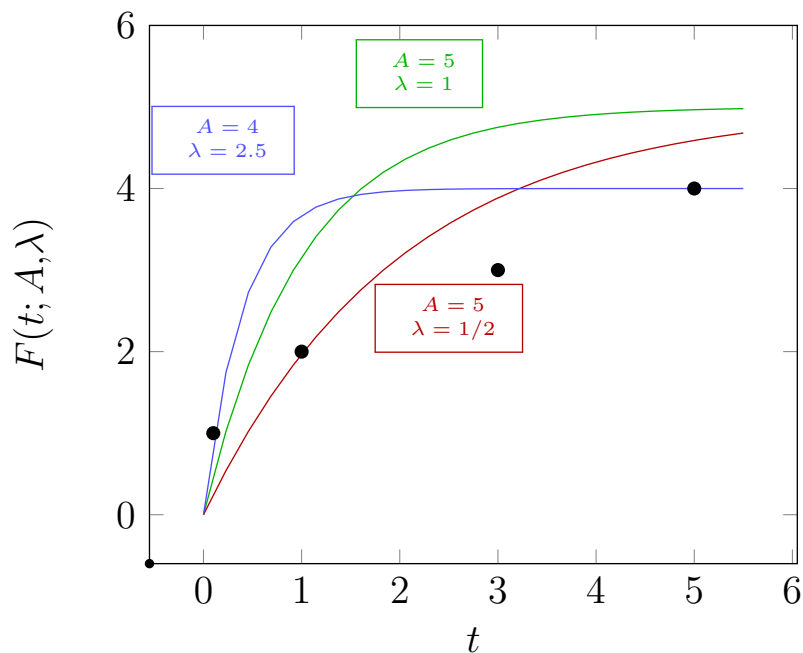


Abbildung 5: Messpunkte zu Beispiel 2 und Beispiele für die Modellfunktion für verschiedene  $A$  und  $\lambda$ -Werte.

wobei

$$F(x) = F(A, \lambda) = \begin{pmatrix} A(1 - e^{-\lambda \cdot 0.1}) - 1 \\ A(1 - e^{-\lambda \cdot 1}) - 2 \\ A(1 - e^{-\lambda \cdot 3}) - 3 \\ A(1 - e^{-\lambda \cdot 5}) - 4 \end{pmatrix},$$

d.h.

$$F_i(x) = F_i(A, \lambda) = A(1 - e^{-\lambda \cdot t_i}) - g_i$$

### Aufstellen der Jacobi-Matrix:

Bei Aufstellen der Jacobi-Matrix kann man entweder

- die Darstellung von  $F(x)$  nehmen und ableiten
- oder  $F_i(x)$  einmal ableiten (in Abhängigkeit von  $i$ ) und einsetzen für alle  $i$ .

Typischerweise ist die zweite Variante schneller und weniger fehleranfällig:

$$F'_i(x) = \begin{pmatrix} \frac{\partial g}{\partial A} & \frac{\partial g}{\partial \lambda} \end{pmatrix}_{i=1, \dots, 4} = (1 - e^{-\lambda t_i} \quad t_i A e^{-\lambda t_i})_{i=1, \dots, 4}$$

## 0. Iterationsschritt (Startwert):

Startwert  $x^0 = (4, 2.5)^T = (A^0, \lambda^0)$  (aus Vorüberlegungen / Plot)

### Residuum auswerten:

$$\|F(A^0, \lambda^0)\|_2 = 1.950$$

## 1. Iterationsschritt:

Einsetzen in Linearisierung um nichtlineares Ausgleichsproblem in lineares Ausgleichsproblem zu verwandeln, d.h. suche  $\Delta x^{0,*} \in \mathbb{R}^2$  mit

$$\|F(x^0) + F'(x^0)\Delta x^{0,*}\|_2 = \min_{\Delta x^0 \in \mathbb{R}^2} \|F(x^0) + F'(x^0)\Delta x^0\|_2 \quad (2)$$

wobei  $\Delta x^0 = (\Delta A^0, \Delta \lambda^0)^T$ . Damit bekommen wir ein lineares Ausgleichsproblem (wieder mit erweiterter Matrix notiert):

$$(F'(x^0) | -F(x^0)) = (A|b)$$

Setzen wir nun die Werte  $x^0$  ein, erhalten wir:

$$\left( \begin{array}{cc|c} 0.2212 & 0.3115 & 0.1152 \\ 0.9179 & 0.3283 & -1.672 \\ 0.9994 & 0.006637 & -0.9978 \\ 1.0 & 7.453 \cdot 10^{-5} & 1.491 \cdot 10^{-5} \end{array} \right)$$

**Lineares Ausgleichsproblem (2) wie üblich lösen:** Das Ausgleichsproblem kann prinzipiell mit Hilfe der Normalgleichung gelöst werden. Zu bevorzugen ist aus Stabilitätsgründen aber immer eine QR-Zerlegung. Wir benutzen hier die Householder-Transformation.

1. Zunächst erste Spalte  $y = a^1$ :

(a) Bestimme Spiegelvektor  $v = y + \text{sign}(y_1)\|y\|e^1$ :

$$\|y\| = 1.700, \quad \text{sign}(y_1) = 1 \Rightarrow v = \begin{pmatrix} 1.921 \\ 0.9179 \\ 0.9994 \\ 1.0 \end{pmatrix}$$



(b) Bestimme  $\beta$ :

$$\beta = \frac{2}{v^T v} = 0.3061$$

(c) Skalarprodukte (Zeilenvektor  $h$ ) berechnen:

$$h = v^T \cdot (A|b) = (3.266 \quad 0.9066 \quad -2.310)$$

(d) Korrekturmatrix  $r$  berechnen:

$$r = \beta v h = \left( \begin{array}{cc|c} 1.921 & 0.5333 & -1.359 \\ 0.9179 & 0.2548 & -0.6492 \\ 0.9994 & 0.2774 & -0.7069 \\ 1.0 & 0.2776 & -0.7073 \end{array} \right)$$

(e) Korrektur anwenden:

$$(\hat{A}|\hat{b}) = (A|b) - r = \left( \begin{array}{cc|c} -1.700 & -0.2218 & 1.474 \\ 0 & 0.07357 & -1.022 \\ 0 & -0.2708 & -0.2909 \\ 0 & -0.2775 & 0.7073 \end{array} \right)$$

(f) Definiere Untermatrix

$$(\tilde{A}|\tilde{b}) = \left( \begin{array}{cc|c} 0.07357 & -1.022 & \\ -0.2708 & -0.2909 & \\ -0.2775 & 0.7073 & \end{array} \right)$$

2. Zweite Spalte  $y = \tilde{a}^1$ :

(a) Bestimme Spiegelvektor  $v = y + \text{sign}(y_1)\|y\|e^1$ :

$$\|y\| = 0.3946, \quad \text{sign}(y_1) = 1 \Rightarrow v = \left( \begin{array}{c} 0.4682 \\ -0.2708 \\ -0.2775 \end{array} \right)$$

(b) Bestimme  $\beta$ :

$$\beta = \frac{2}{v^T v} = 5.413$$

(c) Skalarprodukte (Zeilenvektor  $h$ ) berechnen:

$$h = v^T \cdot (A|b) = (0.1848 \quad -0.5962)$$

(d) Korrekturmatrix  $r$  berechnen:

$$r = \beta v h = \left( \begin{array}{cc|c} 0.4682 & & -1.511 \\ -0.2708 & & 0.8737 \\ -0.2775 & & 0.8954 \end{array} \right)$$

(e) Korrektur anwenden:

$$(\bar{A}|\bar{b}) = (\tilde{A}|\tilde{b}) - r = \left( \begin{array}{cc|c} -0.3946 & & 0.4884 \\ 0 & & -1.165 \\ 0 & & -0.1881 \end{array} \right)$$

3. Transformiertes System:

$$\left( \begin{array}{cc|c} -1.700 & -0.2218 & 1.474 \\ 0 & -0.3946 & 0.4884 \\ \hline 0 & 0 & -1.165 \\ 0 & 0 & -0.1881 \end{array} \right)$$

4. Löse Dreieckssystem:

$$\Rightarrow \begin{pmatrix} \Delta A^1 \\ \Delta \lambda^1 \end{pmatrix} = \begin{pmatrix} -0.7057 \\ -1.238 \end{pmatrix}$$

5. Wende Update an:

$$\begin{pmatrix} A^1 \\ \lambda^1 \end{pmatrix} = \begin{pmatrix} A^0 \\ \lambda^0 \end{pmatrix} + \begin{pmatrix} \Delta A^1 \\ \Delta \lambda^1 \end{pmatrix} = \begin{pmatrix} 3.294 \\ 1.262 \end{pmatrix}$$

**Residuum auswerten:**

$$\|F(A^1, \lambda^1)\|_2 = 1.028$$

Das linearisierte Problem hat das Residuum vom nichtlinearen Problem reduziert.

**Bemerkung:**

Das Residuum vom linearisierten Problem ist ca. 1.18 also unterschiedlich vom Residuum des nichtlinearen Problems.

**2. Iterationsschritt:**

Setzen wir nun die Werte  $x^1$  ein, erhalten wir:

$$\left( \begin{array}{cc|c} 0.1186 & 0.2904 & 0.6093 \\ 0.7170 & 0.9322 & -0.3622 \\ 0.9773 & 0.2239 & -0.2197 \\ 0.9982 & 2.988 \cdot 10^{-2} & 0.7116 \end{array} \right)$$

**Lineares Ausgleichsproblem (2) wie üblich lösen:**

(...)

$$\dots \Rightarrow \begin{pmatrix} \Delta A^2 \\ \Delta \lambda^2 \end{pmatrix} = \begin{pmatrix} 0.3085 \\ -0.4801 \end{pmatrix}$$

(...) Wende Update an:

$$\begin{pmatrix} A^2 \\ \lambda^2 \end{pmatrix} = \begin{pmatrix} A^1 \\ \lambda^1 \end{pmatrix} + \begin{pmatrix} \Delta A^2 \\ \Delta \lambda^2 \end{pmatrix} = \begin{pmatrix} 3.603 \\ 0.7823 \end{pmatrix}$$

**Residuum auswerten:**

$$\|F(A^2, \lambda^2)\|_2 = 0.9056$$

Das linearisierte Problem hat das Residuum vom nichtlinearen Problem reduziert.

**Lösung:**

Weitere Iterationen konvergieren gegen die exakte Lösung des Problems. Die Lösung des Problems ist  $A = 3.8605284$  und  $\lambda = 0.69519100$ .

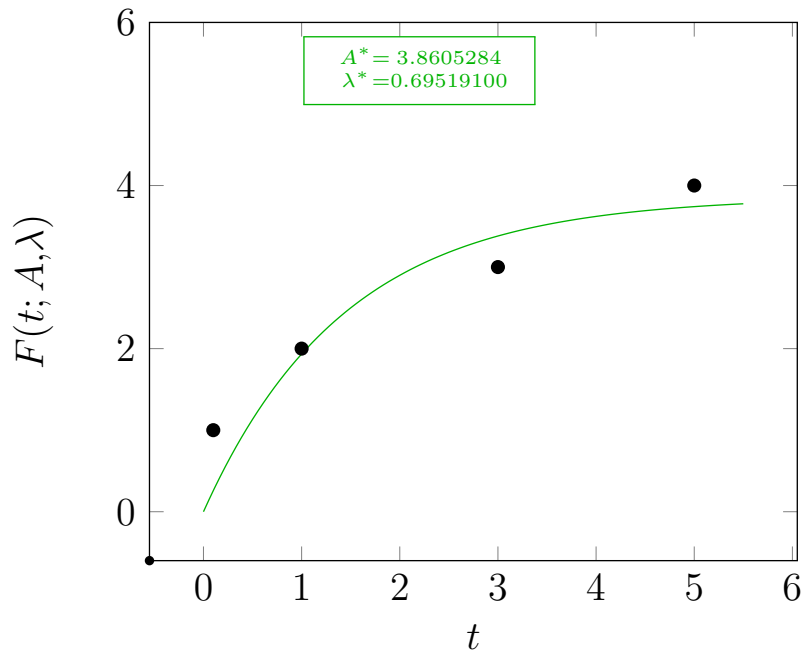


Abbildung 6: Messpunkte zu Beispiel 2 und Modellfunktion für beste  $A$  und  $\lambda$ -Werte.

## Eigenschaften Gauss-Newton

- Das Gauß-Newton-Verfahren ist eine Fixpunktiteration Folie 6.6
- Das Gauß-Newton-Verfahren ist ein modifiziertes Newton-Verfahren: Definiere  $\phi(x) = \frac{1}{2}\|F(x)\|_2^2 = \frac{1}{2}F(x)^T F(x)$  und  $g(x) = \nabla\phi(x)$ . Um  $\phi(x)$  zu minimieren suchen wir die Punkte an denen  $g(x) = \nabla\phi(x) = 0$  gilt. Jetzt betrachten wir das Newton-Verfahren für  $g(x)$ :

$$\begin{aligned}
 x^{k+1} &= x^k - [g'(x^k)]^{-1}g(x^k) \\
 &= x^k - [\phi''(x^k)]^{-1}\nabla\phi(x^k) \\
 &= x^k - [F'(x^k)^T F'(x^k) + \sum_{i=1}^m F_i(x^k)F_i''(x^k)]^{-1}F'(x^k)^T F(x^k) \\
 &\not\approx x^k - [F'(x^k)^T F'(x^k)]^{-1}F'(x^k)^T F(x^k) \\
 &\uparrow \text{keine Näherung erster Ordnung (außer } F_i(x^*) = 0\text{)}.
 \end{aligned}$$

Beim Gauß-Newton-Verfahren wird der Term  $+ \sum_{i=1}^m F_i(x^k)F_i''(x^k)$  vernachlässigt, weswegen die Konvergenzordnung i.A. nur  $p = 1$  beträgt.

- lokale Maxima und Sattelpunkte sind abstoßend (positiv!).
- lokale Minima können auch abstoßend sein (negativ!).
- Fälle in denen  $F'(x)$  nicht vollen Rang hat sind schwieriger zu behandeln.
- Im Allgemeinen nur lokale Konvergenz, "Überschießen" (zu große Schritte) ist möglich.

## Eigenschaften Levenberg-Marquardt-Verfahren

Folien 6.15 + 6.17

- Erweiterung Matrix  $\rightarrow$  stets voller Rang  $\rightarrow$  linearisiertes Problem stets eindeutig lösbar
- Konvergenzgeschwindigkeit wird gegenüber Gauß-Newton **nicht** verbessert, Konvergenzordnung ist auch  $p = 1$
- Die Stabilität bzw. die Robustheit (gegenüber Startwerten) ist viel besser
- Regularisierungsparameter muss sinnvoll gewählt werden (Strategie): Kleines  $\mu$  (wenig Regularisierung) bedeutet schnellere Konvergenz (nah am Gauss-Newton-Verfahren), aber weniger Robustheit. Großes  $\mu$  bedeutet größere Robustheit, aber langsamere Konvergenz.