Beispiel 6.1. Gedämpfte Schwingung:

$$u'' + \frac{b}{m}u' + \frac{D}{m}u = 0,$$

Lösungen haben die Form:

$$u(t) = u_0 e^{-\delta t} \sin(\omega_d t + \varphi_0).$$

Modell einer gedämpften Schwingung

$$y(t; x_1, x_2, x_3, x_4) = x_1 e^{-x_2 t} \sin(x_3 t + x_4) \ ,$$

mit Parametern $x_1, \dots, x_4$. Methode der kleinsten Fehlerquadrate:

$$\sum_{i=1}^{10} \left(x_1 e^{-x_2 t_i} \sin(x_3 t_i + x_4) - b_i \right)^2 = \|F(x_1, x_2, x_3, x_4)\|_2^2,$$

mit $F : \mathbb{R}^4 \to \mathbb{R}^{10}$ definiert durch

$$F_i(x) = F_i(x_1, x_2, x_3, x_4) = x_1 e^{-x_2 t_i} \sin(x_3 t_i + x_4) - b_i \ , \quad i = 1, \dots, 10.$$

$\triangle$

Definiert man allgemein die Abbildung

$$F : \mathbb{R}^n \to \mathbb{R}^m \ , \quad F_i(x) := y(t_i; x) - b_i, \quad i = 1, \ldots, m \ ,$$

kann das *nichtlineare Ausgleichsproblem* wie folgt formuliert werden:

Bestimme $x^* \in \mathbb{R}^n$, so daß

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

oder, äquivalent,

$$\phi(x^*) = \min_{x \in \mathbb{R}^n} \phi(x) \ ,$$

wobei $\phi : \mathbb{R}^n \to \mathbb{R}$, $\phi(x) := \frac{1}{2}\|F(x)\|_2^2 = \frac{1}{2}F(x)^T F(x)$.

Die Funktion

$$\phi(x) = \frac{1}{2}\|F(x)\|_2^2 = \frac{1}{2}F(x)^T F(x)$$

hat in einem Punkt x^* ein *lokales Minimum* genau dann, wenn die folgenden zwei Bedingungen erfüllt sind:

$$\nabla\phi(x^*) = 0 \quad \text{(d.h., } x^* \text{ ist kritischer Punkt von } \phi\text{),}$$

$$\phi''(x^*) \in \mathbb{R}^{n\times n} \quad \text{ist symmetrisch positiv definit.}$$

Es läßt sich durch Nachrechnen bestätigen, daß

$$\nabla\phi(x) = F'(x)^T F(x),$$

$$\phi''(x) = F'(x)^T F'(x) + \sum_{i=1}^{m} F_i(x) F_i''(x).$$

Taylorentwicklung:

$$F(x) = F(x^k) + F'(x^k)(x - x^k) + \mathcal{O}(\|x - x^k\|_2^2).$$

Abbruch nach dem linearen Term $\longrightarrow$ *lineares* Ausgleichsproblem:

> Finde $s^k \in \mathbb{R}^n$ mit minimaler 2-Norm, so daß
>
> $$\|F'(x^k)s^k + F(x^k)\|_2 = \min_{s \in \mathbb{R}^n} \|F'(x^k)s + F(x^k)\|_2$$

Setze

$$x^{k+1} := x^k + s^k.$$

Der Erfolg dieser Strategie hängt von der Wahl des Startwertes ab.

Bemerkung: "mit minimaler 2-Norm" kann man weglassen, wenn

$$\text{Rang}(F'(x)) = n$$

gilt.

Insgesamt erhält man folgendes Verfahren:

Algorithmus 6.3 (Gauß-Newton). Wähle Startwert x^0.

Für $k = 0, 1, 2, \ldots$:

- Berechne $F(x^k)$, $F'(x^k)$.
- Löse das lineare Ausgleichsproblem (6.7).
- Setze $x^{k+1} = x^k + s^k$.

Als Abbruchkriterium für diese Methode wird häufig

$$\|F'(x^k)^T F(x^k)\|_2 \leq \varepsilon$$

benutzt, wobei ε eine vorgegebene Toleranz ist. Der zugrunde liegende Gedanke hierbei ist, daß in einem kritischen Punkt x von ϕ die Ableitung $\nabla\phi(x) = F'(x)^T F(x)$ Null sein muß.

Sei x^* ein kritischer Punkt von ϕ, der in einer Umgebung U eindeutig ist. Annahme:

$$\text{Rang}(F'(x)) = n \quad \text{für alle} \quad x \in U \ .$$

Für $x^k \in U$ hat das lineare Ausgleichsproblem (6.7) wegen Satz 4.5 die eindeutige Lösung

$$s^k = -[F'(x^k)^T F'(x^k)]^{-1} F'(x^k)^T F(x^k)$$

Deshalb gilt für die Gauß-Newton-Iteration:

$$\begin{aligned} x^{k+1} &= x^k - [F'(x^k)^T F'(x^k)]^{-1} F'(x^k)^T F(x^k) \\ &= x^k - [F'(x^k)^T F'(x^k)]^{-1} \nabla\phi(x^k) \\ &= \Phi(x^k) \ , \end{aligned}$$

mit

$$\Phi(x) := x - [F'(x)^T F'(x)]^{-1} \nabla\phi(x) \ .$$

Es gilt: $\quad x = \Phi(x) \ \Leftrightarrow \ \nabla\phi(x) = 0 \ \Leftrightarrow x = x^* \ .$

Die Gauß-Newton-Methode ist also eine **Fixpunktiteration**.

$$F(x) := \begin{pmatrix} a + r\cos x \\ r\sin x \end{pmatrix} \ , \qquad \text{mit } a > r > 0, \quad x \in [0, 2\pi] \ .$$

Ausgleichsproblem: $\min_x \|F(x)\|_2$.

$$\begin{aligned} \|F(x)\|_2 &= \sqrt{a^2 + 2ar\cos x + r^2}, \\ F'(x) &= r\begin{pmatrix} -\sin x \\ \cos x \end{pmatrix}, \qquad F'(x)^T F'(x) = r^2, \\ \nabla\phi(x) &= -ra\sin x \ . \end{aligned}$$

Es gibt zwei kritische Punkte von ϕ:

$$\begin{aligned} x^* &= 0 \quad \text{(lokales Maximum)} \ , \\ x^* &= \pi \quad \text{(lokales Minimum)} \ . \end{aligned}$$

Die Iterationsfunktion zu F ist

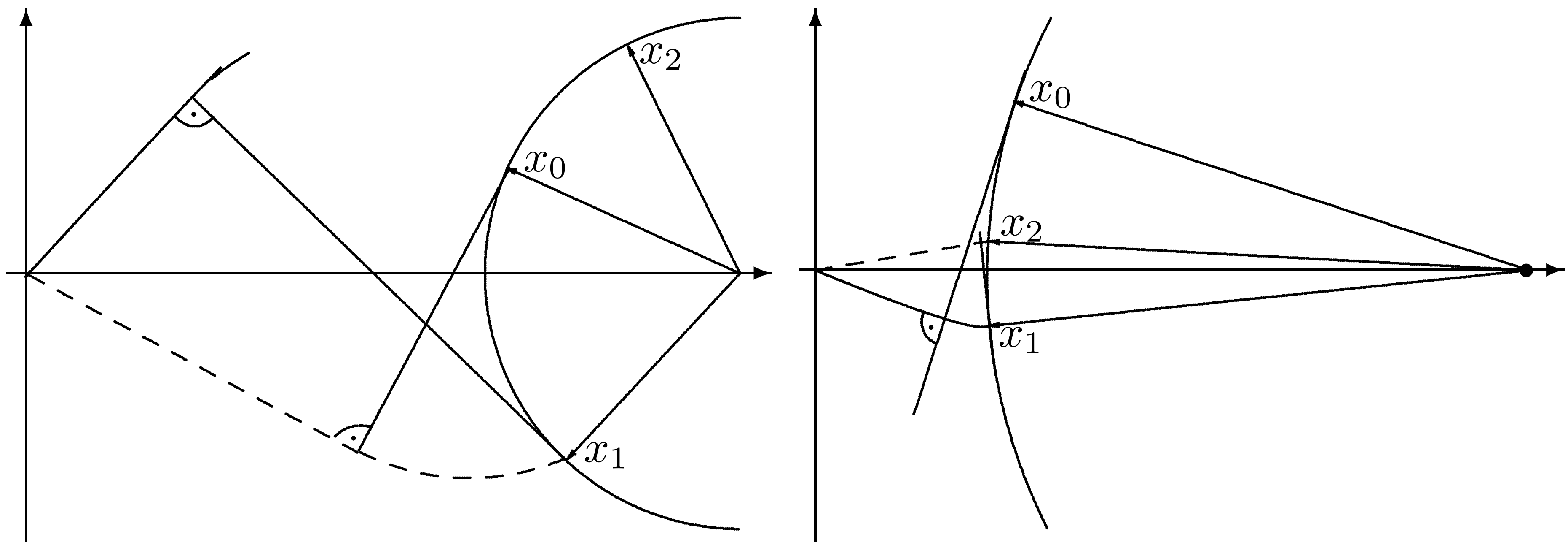
$$\Phi(x) = x + \frac{a}{r}\sin x \ .$$

In den kritischen Punkten $x^* = 0,\ x^* = \pi$ gilt

$$|\Phi'(x^*)| = |1 + \frac{a}{r} \cos x^*| \ .$$

Für $x^* = 0$ (lokales Maximum) gilt $|\Phi'(x^*)| = \frac{a+r}{r} > 1$.

Für $x^* = \pi$ (lokales Minimum) gilt $|\Phi'(x^*)| = \frac{a-r}{r} = \frac{a}{r} - 1$.


x_0
x_1
x_2
x_0
x_1
x_2

Die Gauß-Newton-Methode hat in diesem Beispiel folgende Eigenschaften:

1. Das lokale Maximum ist abstoßend.

2. Die Methode ist *linear* konvergent in einer Umgebung des lokalen Minimums (wenn $a < 2r$), oder

3. das lokale Minimum ist auch abstoßend (wenn $a > 2r$).

$\triangle$

Man kann zeigen, daß ähnliche Eigenschaften in einem allgemeinen Rahmen gültig sind.

Folgerung 6.6.

Für die Gauß-Newton-Iterationsfunktion Φ gilt

$$\|\Phi'(x^*)\|_A = \rho(K)\,\|F(x^*)\|_2\ ,$$
$$\|\Phi'(x^*)\| \geq \rho(K)\,\|F(x^*)\|_2 \quad \text{für jede Operatornorm} \quad \|\cdot\|.$$

(Mit K: Matrix aus (6.12)). Hieraus kann man folgendes schließen:

- Im Normalfall ist $F(x^*) \neq 0$, $K \neq 0$ und deshalb $\Phi'(x^*) \neq 0$.

  > Falls die Gauß-Newton-Methode konvergiert, ist die Konvergenz im allgemeinen nicht schneller als linear.

- Wenn der kritische Punkt x^* ein *lokales Maximum oder ein Sattelpunkt* ist, gilt $\rho(K)\,\|F(x^*)\|_2 \geq 1$ und deshalb $\|\Phi'(x^*)\| \geq 1$ für jede Operatornorm $\|\cdot\|$. Das Verfahren bewahrt uns also davor, einen „falschen" kritischen Punkt zu finden.

  > Solche kritischen Punkte sind für das Gauß-Newton-Verfahren also abstoßend, was günstig ist, weil ein (lokales) Minimum gesucht wird.

- Die Größe $\rho(K)\,\|F(x^*)\|_2$ ist entscheidend für die lokale Konvergenz des Gauß-Newton-Verfahrens.

  Für ein lokales Minimum x^* der Funktion ϕ ist die lokale Konvergenz des Gauß-Newton-Verfahrens gesichert, falls das Residuum $\|F(x^*)\|_2$ und die Größe $\rho(K)$ hinreichend klein sind, so daß die Bedingung $\rho(K)\,\|F(x^*)\|_2 < 1$ erfüllt ist.

- Sei x^* ein lokales Minimum von ϕ, wofür $\rho(K)\,\|F(x^*)\|_2 > 1$ gilt. Dann ist $\|\Phi'(x^*)\| > 1$ für jede Operatornorm $\|\cdot\|$. Deshalb:

  Ein lokales Minimum von ϕ *kann* für die Gauß-Newton-Methode abstoßend sein.

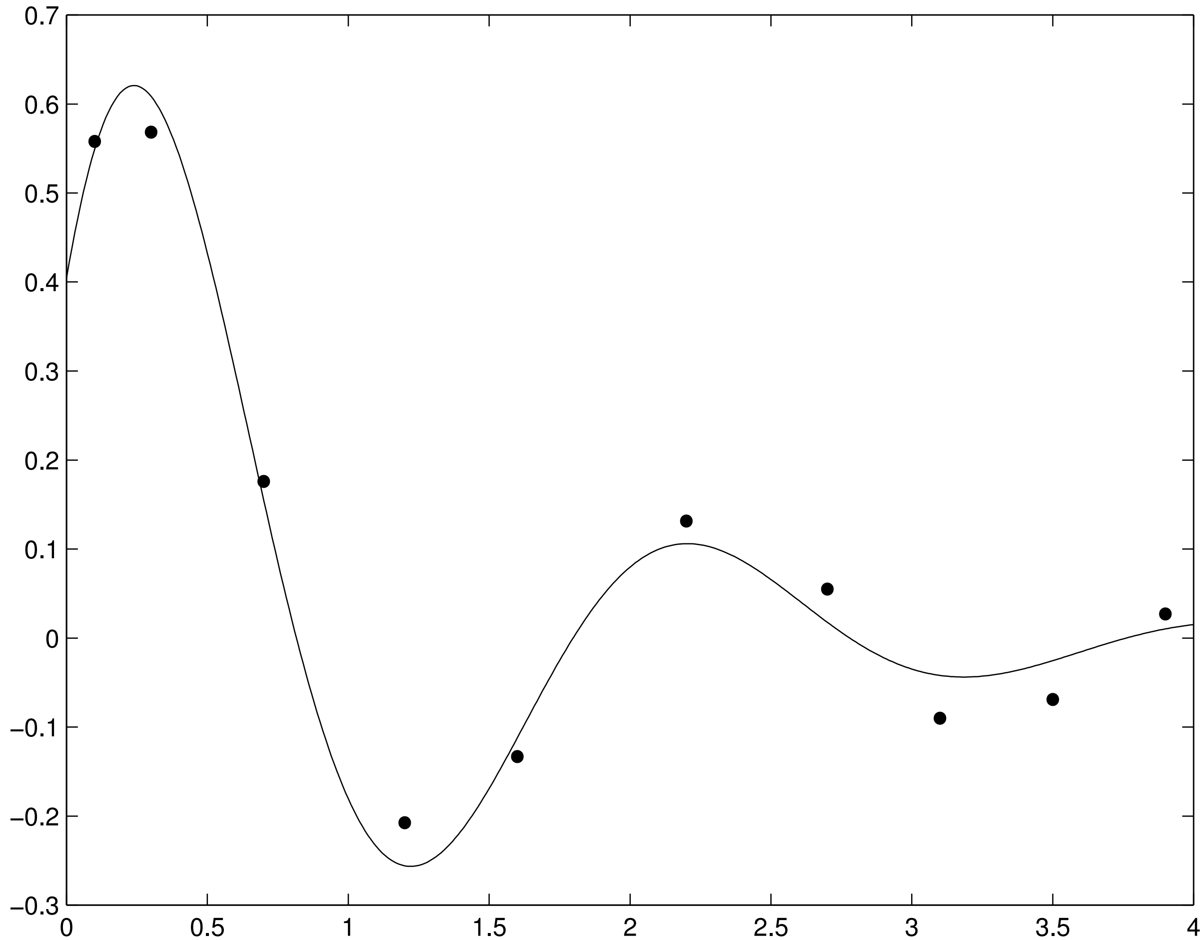
Beispiel 6.7.

Die Gauß-Newton-Methode angewandt auf das Problem in Beispiel 6.1.

k	$\|F(x^k)\|_2$	$\|\nabla\phi(x^k)\|_2$	$\|\nabla\phi(x^k)\|_2/\|\nabla\phi(x^{k-1})\|_2$
0	0.35035332090089	1.45e−01	-
1	0.34106434131008	1.33e−01	0.91
2	0.22208131421995	4.88e−02	0.37
3	0.16802866234936	1.02e−01	2.08
4	0.09190056278958	1.80e−01	0.18
5	0.08902339976144	1.18e−03	0.07
6	0.08895515308450	3.81e−04	0.32
7	0.08894991006370	1.15e−04	0.30
8	0.08894937563528	4.07e−05	0.35
9	0.08894931422207	1.38e−05	0.34
10	0.08894930687791	4.85e−06	0.35
11	0.08894930599062	1.68e−06	0.35
12	0.08894930588306	5.87e−07	0.35

In der letzten Spalte dieser Tabelle sieht man das lineare Konvergenzverhalten der Gauß-Newton-Methode.

Die berechneten Parameterwerte $x^* = x^{12}$ liefern eine entsprechende Lösung $y(t; x^*) = x_1^* e^{-x_2^* t} \sin(x_3^* t + x_4^*)$:



Finde $s^k \in \mathbb{R}^n$, so daß

$$\|F'(x^k)s^k + F(x^k)\|_2^2 + \mu^2 \|s^k\|_2^2 = \min \ ,$$

wobei $\mu > 0$ ein zu wählender Parameter ist.
Neue Annäherung:

$$x^{k+1} = x^k + s^k$$

Äquivalente Formulierung:

$$\boxed{\begin{array}{l}\text{Finde } s^k \in \mathbb{R}^n \text{, so daß} \\ \quad \left\| \begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix} s^k + \begin{pmatrix} F(x^k) \\ \emptyset \end{pmatrix} \right\|_2 = \min \ . \end{array}}$$

Großer Vorteil: die Matrix $\begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix}$ hat *immer vollen Rang*.

Für die Korrektur s^k gilt:

$$\|s^k\|_2 \leq \frac{\|F(x^k)\|_2}{\mu} \ .$$

Also kann man durch eine geeignete Wahl von μ eine „zu große" Korrektur s^k vermeiden. Mit anderen Worten, der Parameter μ kann eine **Dämpfung** der Korrektur bewirken.

Auch das Levenberg-Marquardt-Verfahren kann man als Fixpunktiteration formulieren, wobei

$$\begin{aligned}\Phi_\mu(x) &= x - [F'(x)^T F'(x) + \mu^2 I]^{-1} F'(x)^T F(x) \\ &= x - [F'(x)^T F'(x) + \mu^2 I]^{-1} \nabla\phi(x) \ .\end{aligned}$$

Algorithmus 6.10 (Levenberg-Marquardt)

Wähle Startwert x^0 und Anfangswert für den Parameter μ.

Für $k = 0, 1, 2, \ldots$:

1. Berechne $F(x^k)$, $F'(x^k)$.

2. Löse das lineare Ausgleichsproblem

$$\| \begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix} s^k + \begin{pmatrix} F(x^k) \\ \emptyset \end{pmatrix} \|_2 = \min \quad .$$

3. Teste, ob die Korrektur s^k akzeptabel ist. Wenn nein, dann wird μ angepaßt und Schritt 2 wiederholt.

Wenn ja, dann:

4. Setze $x^{k+1} = x^k + s^k$.