

Numerisches Rechnen

Nichtlineare Ausgleichsrechnung

M. Grepl

J. Berger & R. O'Connor

Institut für Geometrie und Praktische Mathematik
RWTH Aachen

Wintersemester 2015/16

Vorlesungsinhalt

4. Nichtlineare Gleichungssysteme, iterative Lösungsverfahren

geg.: $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$;

ges.: $x^* \in \mathbb{R}^n$, so dass $f(x^*) = 0$

5. Nichtlineare Ausgleichsrechnung

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

ges.: $x^* \in \mathbb{R}^n$, so dass $x^* = \arg \min_{x \in \mathbb{R}^n} \|F(x)\|_2$

6. Interpolation

geg.: Stützstellen und zugehörige Daten

$(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))$;

ges.: Polynom $P_n \in \Pi_n$, so dass

$$P_n(x_j) = f(x_j), \quad j = 0, \dots, n$$

Übersicht

Themen: Dahmen & Reusken Kap. 6.1-6.3

- ▶ Das nichtlineare Ausgleichsproblem
- ▶ Gauß-Newton-Verfahren
- ▶ Levenberg-Marquardt-Verfahren

Übersicht

Themen: Dahmen & Reusken Kap. 6.1-6.3

- ▶ Das nichtlineare Ausgleichsproblem
- ▶ Gauß-Newton-Verfahren
- ▶ Levenberg-Marquardt-Verfahren

Was Sie mitnehmen sollten:

- ▶ Unterschied lineares vs. nichtlineares Ausgleichsproblem
- ▶ Lösungsverfahren: vom linearen zum nichtlinearen Fall
- ▶ Unterschied Newton- vs. Gauß-Newton-Verfahren

Zur Erinnerung: Lineares Ausgleichsproblem

Polynomial Data-Fitting

Gegeben:

- ▶ m Messungen an den Punkten y_1, y_2, \dots, y_m mit zugehörigen Daten z_1, z_2, \dots, z_m .
- ▶ Polynom $(n - 1)$ -ter Ordnung (wobei $n < m$)

$$f(y) = c_0 + c_1 y + \dots + c_{n-1} y^{n-1}$$

Gesucht:

- ▶ Koeffizienten c_0, c_1, \dots, c_{n-1} , so dass die Summe der Fehlerquadrate

$$\sum_{i=1}^m (f(y_i) - z_i)^2$$

minimal wird.

Zur Erinnerung: Lineares Ausgleichsproblem

In Matrix-Vektor Notation

$$\|A x^* - b\|_2 = \min_{x \in \mathbb{R}^n} \|A x - b\|_2$$

mit

$$A = \begin{bmatrix} 1 & y_1 & \cdots & y_1^{n-1} \\ 1 & y_2 & \cdots & y_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & y_m & \cdots & y_m^{n-1} \end{bmatrix}, \quad x = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}, \quad b = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$

Wesentliche Eigenschaft

Die unbekannten Koeffizienten/Parameter tauchen linear auf

Zur Erinnerung: Lineares Ausgleichsproblem

In Matrix-Vektor Notation

$$\|A x^* - b\|_2 = \min_{x \in \mathbb{R}^n} \|A x - b\|_2$$

mit

$$A = \begin{bmatrix} 1 & y_1 & \cdots & y_1^{n-1} \\ 1 & y_2 & \cdots & y_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & y_m & \cdots & y_m^{n-1} \end{bmatrix}, \quad x = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}, \quad b = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$

Wesentliche Eigenschaft

Die unbekannten Koeffizienten/Parameter tauchen linear auf
bzw.

es lassen sich entsprechende Parameter definieren/identifizieren.

Beim nichtlinearen Ausgleichsproblem ist dies nicht mehr möglich...

Beispiel 6.1

Differentialgleichung einer gedämpften Schwingung:

$$m u'' + b u' + D u = 0,$$

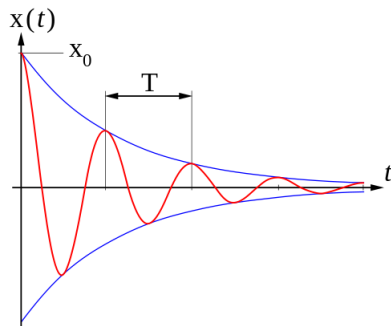
mit Masse m , Dämpfungskonstante b und Federkonstante D .

Lösungen haben die Form:

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

wobei:

u_0	...	Anfangswert
φ_0	...	Nullphasenwinkel
δ	...	Abklingkonstante
ω_d	...	ged. Eigenkreisfrequenz



Quelle: wikipedia

Beispiel 6.1

Gegeben:

- ▶ 10 Messungen an den Punkten t_1, t_2, \dots, t_{10} mit zugehörigen Daten b_1, b_2, \dots, b_{10} .
- ▶ 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 .

Beispiel 6.1

Gegeben:

- ▶ 10 Messungen an den Punkten t_1, t_2, \dots, t_{10} mit zugehörigen Daten b_1, b_2, \dots, b_{10} .
- ▶ 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 .

Gesucht:

- ▶ Parameter x_1, \dots, x_4 , so dass die Summe der Fehlerquadrate

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

minimal wird.

Beispiel 6.1

Gegeben:

- ▶ 10 Messungen an den Punkten t_1, t_2, \dots, t_{10} mit zugehörigen Daten b_1, b_2, \dots, b_{10} .
- ▶ 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 .

Gesucht:

- ▶ Parameter x_1, \dots, x_4 , so dass die Summe der Fehlerquadrate

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

minimal wird. Hierbei ist $F: \mathbb{R}^4 \rightarrow \mathbb{R}^{10}$ definiert durch

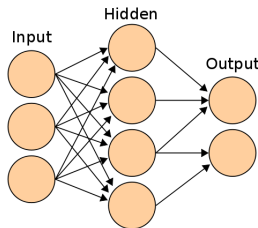
$$\begin{aligned} F_i(x) &= F_i(x_1, x_2, x_3, x_4) \\ &= x_1 e^{-x_2 t_i} \sin(x_3 t_i + x_4) - b_i, \quad i = 1, \dots, 10. \end{aligned}$$

Neuronale Netze

- ▶ Mehrlagiges Perzeptron
 - ▶ n_k Neuronen in der k -ten Schicht
 - ▶ Aktivierungsfunktion eines Neurons: skalare Abbildung $\phi : \mathbb{R} \rightarrow \mathbb{R}$ (single input-single output map)
- ▶ Ausgang des j -ten Neurons der $(k + 1)$ -ten Schicht

$$x_{k+1}^j = \phi \left(u_k^{0j} + \sum_{s=1}^{n_k} x_k^s u_k^{sj} \right), \quad j = 1, \dots, n_{k+1},$$

wobei u_k^{sj} die zu bestimmenden Gewichte sind.



Neuronale Netze

- ▶ Mehrlagiges Perzeptron mit N Schichten: parameterabhängige Abbildung h , die den Eingangsvektor x_0 in den Ausgangsvektor $x_N = h(u, x_0)$ abbildet, wobei

$$u = \{u_k^{sj} \mid k = 0, \dots, N-1, s = 0, \dots, n_k, j = 1, \dots, n_{k+1}\}$$

Neuronale Netze

- ▶ Mehrlagiges Perzeptron mit N Schichten: parameterabhängige Abbildung h , die den Eingangsvektor x_0 in den Ausgangsvektor $x_N = h(u, x_0)$ abbildet, wobei

$$u = \{u_k^{sj} \mid k = 0, \dots, N-1, s = 0, \dots, n_k, j = 1, \dots, n_{k+1}\}$$

- ▶ Gegeben seien m Eingabe-/Ausgabedaten $(y_1, z_1), \dots, (y_m, z_m)$ eines physikalischen Systems.

Neuronale Netze

- ▶ Mehrlagiges Perzeptron mit N Schichten: parameterabhängige Abbildung h , die den Eingangsvektor x_0 in den Ausgangsvektor $x_N = h(u, x_0)$ abbildet, wobei

$$u = \{u_k^{sj} \mid k = 0, \dots, N-1, s = 0, \dots, n_k, j = 1, \dots, n_{k+1}\}$$

- ▶ Gegeben seien m Eingabe-/Ausgabedaten $(y_1, z_1), \dots, (y_m, z_m)$ eines physikalischen Systems.
- ▶ **Lernverfahren:** Minimiere Ausgabefehler

$$\min_u \frac{1}{2} \sum_{i=1}^m \|z_i - h(u, y_i)\|^2$$

Neuronale Netze

- ▶ Mehrlagiges Perzeptron mit N Schichten: parameterabhängige Abbildung h , die den Eingangsvektor x_0 in den Ausgangsvektor $x_N = h(u, x_0)$ abbildet, wobei

$$u = \{u_k^{sj} \mid k = 0, \dots, N-1, s = 0, \dots, n_k, j = 1, \dots, n_{k+1}\}$$

- ▶ Gegeben seien m Eingabe-/Ausgabedaten $(y_1, z_1), \dots, (y_m, z_m)$ eines physikalischen Systems.

- ▶ **Lernverfahren:** Minimiere Ausgabefehler

$$\min_u \frac{1}{2} \sum_{i=1}^m \|z_i - h(u, y_i)\|^2$$

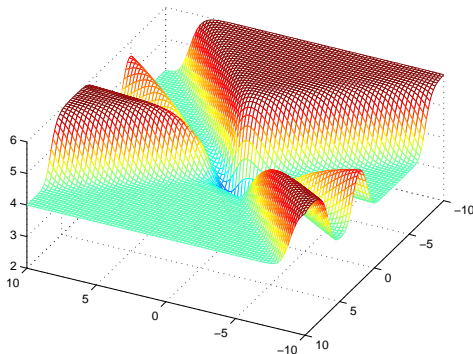
- ▶ Typische Aktivierungsfunktionen der Neuronen

$$\phi(\xi) = \frac{1}{1 + e^{-\xi}}, \quad \phi(\xi) = \tanh(\xi)$$

Neuronale Netze – Beispiel

Lernverfahren ($m = 5$, Gewichte u_0, u_1)

$$\frac{1}{2} \sum_{i=1}^5 (z_i - \phi(u_1 y_i + u_0))^2$$



Definition

Definiert man allgemein die Abbildung ($m > n$)

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

kann das *nichtlineare Ausgleichsproblem* wie folgt formuliert werden:

Nichtlineares Ausgleichsproblem

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

Definition

Zur Erinnerung: 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:

Definition

Zur Erinnerung: 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:

1. $\nabla \phi(x^*) = 0$ (d.h., x^* ist kritischer Punkt von ϕ),

Definition

Zur Erinnerung: 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:

1. $\nabla \phi(x^*) = 0$ (d.h., x^* ist kritischer Punkt von ϕ),
2. $\phi''(x^*) \in \mathbb{R}^{n \times n}$ ist symmetrisch positiv definit.

Definition

Zur Erinnerung: 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:

1. $\nabla \phi(x^*) = 0$ (d.h., x^* ist kritischer Punkt von ϕ),
2. $\phi''(x^*) \in \mathbb{R}^{n \times n}$ ist symmetrisch positiv definit.

Es läßt sich durch Nachrechnen bestätigen, dass

$$\begin{aligned}\nabla \phi(x) &= F'(x)^T F(x), \\ \phi''(x) &= F'(x)^T F'(x) + \sum_{i=1}^m F_i(x) F_i''(x),\end{aligned}$$

mit Jacobi-Matrix $F'(x) \in \mathbb{R}^{m \times n}$ und

$$\text{Hesse-Matrix } F_i''(x) := \left(\frac{\partial^2 F_i(x)}{\partial x_j \partial x_k} \right)_{1 \leq j, k \leq n} \in \mathbb{R}^{n \times n}$$

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, bestimme $x^* \in \mathbb{R}^n$ so dass

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

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, bestimme $x^* \in \mathbb{R}^n$ so dass

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

Ansatz:

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, bestimme $x^* \in \mathbb{R}^n$ so dass

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

Ansatz:

1. Ersetze $F(x)$ durch lineare Approximation (Taylorentwicklung)

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, bestimme $x^* \in \mathbb{R}^n$ so dass

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

Ansatz:

1. Ersetze $F(x)$ durch lineare Approximation (Taylorentwicklung)
2. Schrittweise Annäherung an x^* durch Lösung linearer Probleme in jedem Schritt

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, bestimme $x^* \in \mathbb{R}^n$ so dass

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

Ansatz:

1. Ersetze $F(x)$ durch lineare Approximation (Taylorentwicklung)
2. Schrittweise Annäherung an x^* durch Lösung linearer Probleme in jedem Schritt

Zur Erinnerung:

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, bestimme $x^* \in \mathbb{R}^n$ so dass

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

Ansatz:

1. Ersetze $F(x)$ durch lineare Approximation (Taylorentwicklung)
2. Schrittweise Annäherung an x^* durch Lösung linearer Probleme in jedem Schritt

Zur Erinnerung:

- Wir bezeichnen die Lösung am Iterationsschritt k mit

$$x^k = (x_1^k, \dots, x_n^k)^T \in \mathbb{R}^n.$$

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, bestimme $x^* \in \mathbb{R}^n$ so dass

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

Ansatz:

1. Ersetze $F(x)$ durch lineare Approximation (Taylorentwicklung)
2. Schrittweise Annäherung an x^* durch Lösung linearer Probleme in jedem Schritt

Zur Erinnerung:

- Wir bezeichnen die Lösung am Iterationsschritt k mit

$$x^k = (x_1^k, \dots, x_n^k)^T \in \mathbb{R}^n.$$

- Taylorentwicklung

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

Das Gauß-Newton-Verfahren

Ansatz: Ersetze $\mathbf{F}(\mathbf{x})$ in

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

durch lineare Approximation

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{F}(\mathbf{x}^k) + \mathbf{F}'(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k)\|_2,$$

Das Gauß-Newton-Verfahren

Ansatz: Ersetze $F(x)$ in

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

durch lineare Approximation

$$\min_{x \in \mathbb{R}^n} \underbrace{\|F(x^k) + F'(x^k)(x - x^k)\|_2}_{\in \mathbb{R}^m},$$

Das Gauß-Newton-Verfahren

Ansatz: Ersetze $\mathbf{F}(\mathbf{x})$ in

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

durch lineare Approximation

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\| \underbrace{\mathbf{F}(\mathbf{x}^k)}_{\in \mathbb{R}^m} + \underbrace{\mathbf{F}'(\mathbf{x}^k)}_{\in \mathbb{R}^{m \times n}} (\mathbf{x} - \mathbf{x}^k) \right\|_2,$$

Das Gauß-Newton-Verfahren

Ansatz: Ersetze $F(x)$ in

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

durch lineare Approximation

$$\min_{x \in \mathbb{R}^n} \left\| \underbrace{F(x^k)}_{\in \mathbb{R}^m} + \underbrace{F'(x^k)}_{\in \mathbb{R}^{m \times n}} \underbrace{(x - x^k)}_{=s \in \mathbb{R}^n} \right\|_2,$$

Das Gauß-Newton-Verfahren

Ansatz: Ersetze $F(x)$ in

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

durch lineare Approximation

$$\min_{x \in \mathbb{R}^n} \left\| \underbrace{F(x^k)}_{\in \mathbb{R}^m} + \underbrace{F'(x^k)}_{\in \mathbb{R}^{m \times n}} \underbrace{(x - x^k)}_{=s \in \mathbb{R}^n} \right\|_2,$$

Wir setzen $s = x - x^k$ (bzw. $s^k = x^{k+1} - x^k$) und erhalten das
lineare Ausgleichsproblem:

N6.1

Das Gauß-Newton-Verfahren

Ansatz: Ersetze $F(x)$ in

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

durch lineare Approximation

$$\min_{x \in \mathbb{R}^n} \left\| \underbrace{F(x^k)}_{\in \mathbb{R}^m} + \underbrace{F'(x^k)}_{\in \mathbb{R}^{m \times n}} \underbrace{(x - x^k)}_{=s \in \mathbb{R}^n} \right\|_2,$$

Wir setzen $s = x - x^k$ (bzw. $s^k = x^{k+1} - x^k$) und erhalten das *lineare Ausgleichsproblem*:

N6.1

Finde $s^k \in \mathbb{R}^n$ mit minimaler 2-Norm, so dass

$$\|F'(x^k)s^k + F(x^k)\|_2 = \min_{s \in \mathbb{R}^n} \|F'(x^k)s + F(x^k)\|_2$$

Das Gauß-Newton-Verfahren

Ansatz: Ersetze $F(x)$ in

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

durch lineare Approximation

$$\min_{x \in \mathbb{R}^n} \underbrace{\|F(x^k)\|_2}_{\in \mathbb{R}^m} + \underbrace{F'(x^k)(x - x^k)}_{\substack{\in \mathbb{R}^{m \times n} \\ = s \in \mathbb{R}^n}}\|_2,$$

Wir setzen $s = x - x^k$ (bzw. $s^k = x^{k+1} - x^k$) und erhalten das *lineare Ausgleichsproblem*:

N6.1

Finde $s^k \in \mathbb{R}^n$ mit minimaler 2-Norm, so dass

$$\|F'(x^k)s^k + F(x^k)\|_2 = \min_{s \in \mathbb{R}^n} \|F'(x^k)s + F(x^k)\|_2$$

Anschließend berechnen wir

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

Das Gauß-Newton-Verfahren

Insgesamt erhält man folgendes Verfahren:

Algorithmus 6.3 (Gauß-Newton).

Wähle Startwert x^0 .

Für $k = 0, 1, 2, \dots$:

1. Berechne $F(x^k)$, $F'(x^k)$.
2. Finde $s^k \in \mathbb{R}^n$ mit minimaler 2-Norm, so dass
$$\|F'(x^k)s^k + F(x^k)\|_2 = \min_{s \in \mathbb{R}^n} \|F'(x^k)s + F(x^k)\|_2$$
3. Setze $x^{k+1} = x^k + s^k$.

Das Gauß-Newton-Verfahren

Insgesamt erhält man folgendes Verfahren:

Algorithmus 6.3 (Gauß-Newton).

Wähle Startwert x^0 .

Für $k = 0, 1, 2, \dots$:

1. Berechne $F(x^k), F'(x^k)$.
2. Finde $s^k \in \mathbb{R}^n$ mit minimaler 2-Norm, so dass
$$\|F'(x^k)s^k + F(x^k)\|_2 = \min_{s \in \mathbb{R}^n} \|F'(x^k)s + F(x^k)\|_2$$
3. Setze $x^{k+1} = x^k + s^k$.

Beachte:

- Schritt 2 erfordert die Lösung eines linearen Ausgleichsproblems (Normalgleichung, QR-Zerlegung, SVD)

Bemerkungen

- ▶ “Analogie” nichtlineare Gleichungssysteme.

N6.2

Bemerkungen

- ▶ “Analogie” nichtlineare Gleichungssysteme.
- ▶ In einem kritischen Punkt x^* von ϕ muss die Ableitung $\nabla\phi(x) = F'(x)^T F(x)$ Null sein. Als Abbruchkriterium für das Verfahren wird daher häufig

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

benutzt, wobei ε eine vorgegebene Toleranz ist.

N6.2

Bemerkungen

- ▶ “Analogie” nichtlineare Gleichungssysteme. N6.2
- ▶ In einem kritischen Punkt x^* von ϕ muss die Ableitung $\nabla\phi(x) = F'(x)^T F(x)$ Null sein. Als Abbruchkriterium für das Verfahren wird daher häufig

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

benutzt, wobei ε eine vorgegebene Toleranz ist.

- ▶ Der Erfolg des Gauß-Newton-Verfahrens hängt von der Wahl des Startwerts ab (vgl. Newton-Verfahren).

Bemerkungen

- ▶ “Analogie” nichtlineare Gleichungssysteme. N6.2
- ▶ In einem kritischen Punkt x^* von ϕ muss die Ableitung $\nabla\phi(x) = F'(x)^T F(x)$ Null sein. Als Abbruchkriterium für das Verfahren wird daher häufig

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

benutzt, wobei ε eine vorgegebene Toleranz ist.

- ▶ Der Erfolg des Gauß-Newton-Verfahrens hängt von der Wahl des Startwerts ab (vgl. Newton-Verfahren).
- ▶ Den Zusatz "mit minimaler 2-Norm" kann man weglassen, wenn $\text{Rang}(F'(x)) = n$ gilt (vgl. SVD).

Bemerkungen

- ▶ “Analogie” nichtlineare Gleichungssysteme. N6.2
- ▶ In einem kritischen Punkt x^* von ϕ muss die Ableitung $\nabla \phi(x) = F'(x)^T F(x)$ Null sein. Als Abbruchkriterium für das Verfahren wird daher häufig

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

benutzt, wobei ε eine vorgegebene Toleranz ist.

- ▶ Der Erfolg des Gauß-Newton-Verfahrens hängt von der Wahl des Startwerts ab (vgl. Newton-Verfahren).
- ▶ Den Zusatz "mit minimaler 2-Norm" kann man weglassen, wenn $\text{Rang}(F'(x)) = n$ gilt (vgl. SVD).
- ▶ Analyse Gauß-Newton, Vergleich mit Newton-Verfahren N6.3

Bemerkungen

- ▶ “Analogie” nichtlineare Gleichungssysteme. N6.2
- ▶ In einem kritischen Punkt x^* von ϕ muss die Ableitung $\nabla \phi(x) = F'(x)^T F(x)$ Null sein. Als Abbruchkriterium für das Verfahren wird daher häufig

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

benutzt, wobei ε eine vorgegebene Toleranz ist.

- ▶ Der Erfolg des Gauß-Newton-Verfahrens hängt von der Wahl des Startwerts ab (vgl. Newton-Verfahren).
- ▶ Den Zusatz "mit minimaler 2-Norm" kann man weglassen, wenn $\text{Rang}(F'(x)) = n$ gilt (vgl. SVD).
- ▶ Analyse Gauß-Newton, Vergleich mit Newton-Verfahren N6.3
- ▶ Vergleich lineares vs. nichtlineares Ausgleichsproblem N6.4(1-4)

Beispiel 6.4.

Lösen Sie das nichtlineare Ausgleichsproblem

N6.5

$$\min_{x \in \mathbb{R}} \|F(x)\|_2,$$

wobei

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

Beispiel 6.4.

Lösen Sie das nichtlineare Ausgleichsproblem

N6.5

$$\min_{x \in \mathbb{R}} \|F(x)\|_2,$$

wobei

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

► Für die Jacobi-Matrix erhält man

$$F'(x) = r \begin{pmatrix} -\sin x \\ \cos x \end{pmatrix}, \quad F'(x)^T F'(x) = r^2.$$

Beispiel 6.4.

Lösen Sie das nichtlineare Ausgleichsproblem

N6.5

$$\min_{x \in \mathbb{R}} \|F(x)\|_2,$$

wobei

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

- Für die Jacobi-Matrix erhält man

$$F'(x) = r \begin{pmatrix} -\sin x \\ \cos x \end{pmatrix}, \quad F'(x)^T F'(x) = r^2.$$

- Außerdem ergibt sich

$$\phi(x) = \frac{1}{2} \|F(x)\|_2^2 = \frac{1}{2} (a^2 + 2ar \cos x + r^2)$$

und damit

$$\nabla \phi(x) = -ra \sin x.$$

Beispiel 6.4.

- Für die Iterationsfunktion zu F erhält man schließlich

$$\begin{aligned}\Phi(x) &= x - (F'(x)^T F'(x))^{-1} \nabla \phi(x^k) \\ &= x + \frac{a}{r} \sin x\end{aligned}$$

Beispiel 6.4.

- Für die Iterationsfunktion zu F erhält man schließlich

$$\begin{aligned}\Phi(x) &= x - (F'(x)^T F'(x))^{-1} \nabla \phi(x^k) \\ &= x + \frac{a}{r} \sin x\end{aligned}$$

- Es gibt zwei kritische Punkte von ϕ

Beispiel 6.4.

- Für die Iterationsfunktion zu F erhält man schließlich

$$\begin{aligned}\Phi(x) &= x - (F'(x)^T F'(x))^{-1} \nabla \phi(x^k) \\ &= x + \frac{a}{r} \sin x\end{aligned}$$

- Es gibt zwei kritische Punkte von ϕ

$$x^* = 0 \quad (\text{lokales Maximum}),$$

$$x^* = \pi \quad (\text{lokales Minimum}).$$

Beispiel 6.4.

- Für die Iterationsfunktion zu F erhält man schließlich

$$\begin{aligned}\Phi(x) &= x - (F'(x)^T F'(x))^{-1} \nabla \phi(x^k) \\ &= x + \frac{a}{r} \sin x\end{aligned}$$

- Es gibt zwei kritische Punkte von ϕ

$$x^* = 0 \quad (\text{lokales Maximum}),$$

$$x^* = \pi \quad (\text{lokales Minimum}).$$

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

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

Beispiel 6.4.

- Für die Iterationsfunktion zu F erhält man schließlich

$$\begin{aligned}\Phi(x) &= x - (F'(x)^T F'(x))^{-1} \nabla \phi(x^k) \\ &= x + \frac{a}{r} \sin x\end{aligned}$$

- Es gibt zwei kritische Punkte von ϕ

$$x^* = 0 \quad (\text{lokales Maximum}),$$

$$x^* = \pi \quad (\text{lokales Minimum}).$$

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

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

und damit

$$|\Phi'(x^*)| = \frac{a+r}{r} > 1 \quad \text{für } x^* = 0 \text{ (lokales Max)}$$

$$|\Phi'(x^*)| = \frac{a-r}{r} = \frac{a}{r} - 1 \quad \text{für } x^* = \pi \text{ (lokales Min)}$$

Beispiel 6.4.

Das Gauß-Newton-Verfahren hat in diesem Beispiel folgende Eigenschaften:

1. Das lokale Maximum ist abstoßend

Beispiel 6.4.

Das Gauß-Newton-Verfahren 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

Beispiel 6.4.

Das Gauß-Newton-Verfahren 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 $a > 2r$).

Mann kann zeigen, dass ähnliche Eigenschaften in einem allgemeinen Rahmen gültig sind.

Beispiel 6.7.

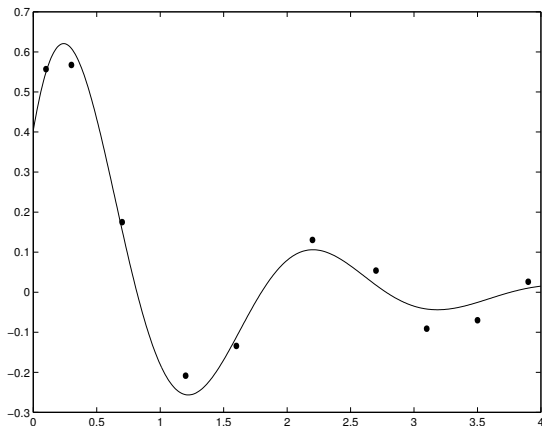
Das Gauß-Newton-Verfahren angewandt auf das Problem der gedämpften Schwingung in Beispiel 6.1.

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 der Tabelle sieht man das lineare Konvergenzverhalten des Gauß-Newton-Verfahrens.

Beispiel 6.7.

Die berechneten Parameterwerte $x^* = x^{12}$ liefern die Lösung $y(t; x^*) = x_1^* e^{-x_2^* t} \sin(x_3^* t + x_4^*)$ im folgenden Plot.



Levenberg-Marquardt-Verfahren

Zur Erinnerung: Berechnung der Korrektur (bzw. Schrittweite) beim Gauß-Newton-Verfahren

$$F'(x)^T F'(x) s = -F'(x)^T F(x)$$

Levenberg-Marquardt-Verfahren

Zur Erinnerung: Berechnung der Korrektur (bzw. Schrittweite) beim Gauß-Newton-Verfahren

$$F'(x)^T F'(x) s = -F'(x)^T F(x)$$

Idee: Einführung einer Dämpfung

$$[F'(x)^T F'(x) + \mu^2 I] s = -F'(x)^T F(x),$$

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

Levenberg-Marquardt-Verfahren

Zur Erinnerung: Berechnung der Korrektur (bzw. Schrittweite) beim Gauß-Newton-Verfahren

$$F'(x)^T F'(x) s = -F'(x)^T F(x)$$

Idee: Einführung einer Dämpfung

$$[F'(x)^T F'(x) + \mu^2 I] s = -F'(x)^T F(x),$$

wobei $\mu > 0$ ein zu wählender Parameter ist. Wir erhalten annähernd für

- ▶ μ gross: Gradientenverfahren (Verfahren des steilsten Abstiegs)
- ▶ μ klein: Gauß-Newton-Verfahren

Levenberg-Marquardt-Verfahren

Zur Erinnerung: Berechnung der Korrektur (bzw. Schrittweite) beim Gauß-Newton-Verfahren

$$F'(x)^T F'(x) s = -F'(x)^T F(x)$$

Idee: Einführung einer Dämpfung

$$[F'(x)^T F'(x) + \mu^2 I] s = -F'(x)^T F(x),$$

wobei $\mu > 0$ ein zu wählender Parameter ist. Wir erhalten annähernd für

- ▶ μ gross: Gradientenverfahren (Verfahren des steilsten Abstiegs)
- ▶ μ klein: Gauß-Newton-Verfahren

Variante: Berechne Schrittweite aus

$$[F'(x)^T F'(x) + \mu^2 \text{diag}(F'(x)^T F'(x))] s = -F'(x)^T F(x),$$

Bemerkungen

Lineares Ausgleichsproblem (Gauß-Newton)

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

Bemerkungen

Lineares Ausgleichsproblem (Gauß-Newton)

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

wird ersetzt durch

$$\min_{s \in \mathbb{R}^n} (\|F'(x^k)s + F(x^k)\|_2^2 + \mu^2 \|s\|_2^2),$$

oder, äquivalent,

$$\min_{s \in \mathbb{R}^n} \left\| \begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix} s + \begin{pmatrix} F(x^k) \\ \emptyset \end{pmatrix} \right\|_2.$$

Bemerkungen

Lineares Ausgleichsproblem (Gauß-Newton)

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

wird ersetzt durch

$$\min_{s \in \mathbb{R}^n} (\|F'(x^k)s + F(x^k)\|_2^2 + \mu^2 \|s\|_2^2),$$

oder, äquivalent,

$$\min_{s \in \mathbb{R}^n} \left\| \begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix} s + \begin{pmatrix} F(x^k) \\ \emptyset \end{pmatrix} \right\|_2.$$

Neue Annäherung: $x^{k+1} = x^k + s^k$

Bemerkungen

Lineares Ausgleichsproblem (Gauß-Newton)

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

wird ersetzt durch

$$\min_{s \in \mathbb{R}^n} (\|F'(x^k)s + F(x^k)\|_2^2 + \mu^2 \|s\|_2^2),$$

oder, äquivalent,

$$\min_{s \in \mathbb{R}^n} \left\| \begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix} s + \begin{pmatrix} F(x^k) \\ \emptyset \end{pmatrix} \right\|_2.$$

Neue Annäherung: $x^{k+1} = x^k + s^k$

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

Bemerkungen

- Für die Korrektur s^k gilt

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

d.h. man kann durch eine geeignete Wahl von μ eine "zu große" Korrektur s^k vermeiden.

Bemerkungen

- ▶ Für die Korrektur s^k gilt

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

d.h. man kann durch eine geeignete Wahl von μ eine "zu große" Korrektur s^k vermeiden.

- ▶ Wahl der Korrektur in der Praxis heuristisch
 - ▶ basierend auf Residuum $\|F(x^k)\|_2^2 - \|F(x^k + s^k)\|_2^2$
 - ▶ algebraisch (Quotientenregel)

Bemerkungen

- ▶ Für die Korrektur s^k gilt

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

d.h. man kann durch eine geeignete Wahl von μ eine "zu große" Korrektur s^k vermeiden.

- ▶ Wahl der Korrektur in der Praxis heuristisch
 - ▶ basierend auf Residuum $\|F(x^k)\|_2^2 - \|F(x^k + s^k)\|_2^2$
 - ▶ algebraisch (Quotientenregel)
- ▶ Levenberg-Marquardt-Verfahren kann auch als Fixpunktiteration formuliert werden

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

Zusammenfassung

Algorithmus 6.10 (Levenberg-Marquardt)

Wähle Startwert x^0 und Anfangswert für den Parameter μ .
Für $k = 0, 1, 2, \dots$:

1. Berechne $F(x^k)$, $F'(x^k)$
2. Löse das lineare Ausgleichsproblem

$$s^k = \arg \min_{s \in \mathbb{R}^n} \left\| \begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix} s + \begin{pmatrix} F(x^k) \\ \emptyset \end{pmatrix} \right\|_2.$$

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

Wenn ja, dann:

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