

# Numerische Mathematik für Maschinenbauer

## Nichtlineare Ausgleichsrechnung

A. Reusken

K.-H. Brakhage, Amira El Amouri, Thomas Jankuhn

Institut für Geometrie und Praktische Mathematik  
RWTH Aachen

Sommersemester 2019

# Heute in der Vorlesung

Themen: Dahmen & Reusken Kap 6.1-6.3

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

Was Sie mitnehmen sollten:

- ▶ Wie ist die Problemstellung bei einem nichtlinearen Ausgleichsproblem
- ▶ Wie funktioniert das Gauß-Newton-Verfahren
- ▶ Wie funktioniert das Levenberg-Marquardt-Verfahren
- ▶ Wichtige (Konvergenz-)Eigenschaften dieser Methoden

## Zur Erinnerung: Lineares Ausgleichsproblem (Beispiel 4.2.)

In der Fourieranalyse wird eine  $T$ -periodische Funktion  $f$  durch eine **Linearkombination** der  $T$ -periodischen trigonometrischen Polynome

$$1, \cos(ct), \sin(ct), \cos(2ct), \sin(2ct), \dots, \cos(Nct), \sin(Nct)$$

mit  $c := \frac{2\pi}{T}$  in der Form

$$g_N(t) = \frac{1}{2} \alpha_0 + \sum_{k=1}^N \left( \alpha_k \cos(kct) + \beta_k \sin(kct) \right)$$

approximiert.

# Zur Erinnerung: Lineares Ausgleichsproblem (Beispiel 4.2.)

## Annahme:

Nicht  $f$ , sondern nur eine Reihe vom Meßdaten

$$b_i \approx f(t_i), \quad 0 \leq t_1 < t_2 < \dots < t_m \leq T,$$

ist bekannt, wobei  $m > 2N + 1$ .

Daraus ergibt sich der

## Ansatz zur Bestimmung der Koeffizienten

$$x = (\alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \alpha_N, \beta_N)^T.$$

Gaußsche Fehlerquadratmethode: 
$$\min_{x \in \mathbb{R}^{2N+1}} \sum_{i=1}^m (g_N(t_i) - b_i)^2.$$

# Lineares vs. nichtlineares Ausgleichsproblem

## Definition

Zu gegebenen  $A \in \mathbb{R}^{m \times n}$  und  $b \in \mathbb{R}^m$ , bestimme  $x^* \in \mathbb{R}^n$ , für das

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

gilt.

**Notation:**

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

Diese Problemstellung heißt das **lineare Ausgleichsproblem**.

## Wesentliche Eigenschaften

Die unbekanntenen Koeffizienten/Parameter tauchen **linear** auf.

**Rang(A) = n**: eindeutige Lösung  $x^*$ .

# Beispiel 6.1.

Differentialgleichung einer gedämpften Schwingung:

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

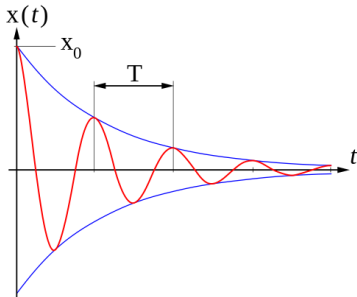
mit Masse  $m$ , Dämpfungskonstante  $b$  und Federkonstante  $D$ .

Lösungen haben die **nichtlineare** Form:

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

wobei:

- $u_0$  → Anfangswert
- $\varphi_0$  → Nullphasenwinkel
- $\delta$  → Abklingkonstante
- $\omega_d$  → 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  $x_1, \dots, x_4$ .

## Gesucht:

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

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

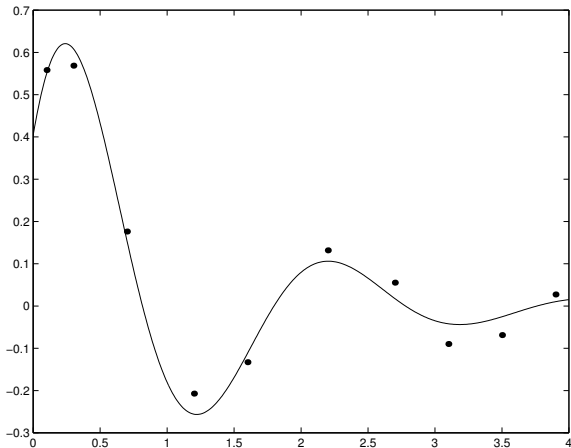
minimal wird.

Hierbei ist  $F : \mathbb{R}^4 \rightarrow \mathbb{R}^{10}$  nichtlinear in  $x_2, x_3, x_4$ .

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

## Beispiel 6.1.

Berechnete Lösung des nichtlinearen Ausgleichsproblems:





# 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 man das **nichtlineare Ausgleichsproblem** wie folgt formulieren:

## Nichtlineares Ausgleichsproblem

Bestimme  $x^* \in U \subseteq \mathbb{R}^n$ , so dass

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

**Notation:**  $x^* = \arg \min_{x \in U \subseteq \mathbb{R}^n} \|F(x)\|_2.$

**Äquivalent:**  $x^* = \arg \min_{x \in U \subseteq \mathbb{R}^n} \phi(x),$

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

# 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**, wenn die folgenden zwei Bedingungen erfüllt sind:

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

Es lässt sich durch Nachrechnen bestätigen, dass

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

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

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

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

# Das Gauß-Newton-Verfahren

## Nichtlineares Ausgleichsproblem (lokal)

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

$$x^* = \arg \min_{x \in U} \|F(x)\|_2.$$

Bei  $U \neq \mathbb{R}^n$  liegt ein **lokales**, ansonsten ein **globales** Problem vor.

### Ansatz:

1. Ersetze  $F(x)$  durch lineare Approximation (Taylor-Entwicklung)
2. Schrittweise Annäherung an  $x^*$  durch Lösung linearer Probleme

### Zur Erinnerung:

- ▶ Wir bezeichnen die Lösung am Iterationsschritt  $k$  mit

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

- ▶ Taylor-Entwicklung

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

# Das Gauß-Newton-Verfahren

## Ansatz:

Ersetze  $F(x)$  in  $\min_{x \in U} \|F(x)\|_2$ , durch lineare Approximation

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

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

Finde  $s^k \in \mathbb{R}^n$ , so dass

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

Anschließend berechnen wir  $x^{k+1} = x^k + s^k$ .

Es kann eventuell  $x^{k+1} \notin U$  sein.

# Das Gauß-Newton-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  $s^k$ , so dass

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

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

## Beachte

- ▶ Schritt 2 erfordert die Lösung eines linearen Ausgleichsproblems (Normalgleichung, QR-Zerlegung)
- ▶ Falls  $F'(x)$  nicht vollen Rang hat, hat das Ausgleichsproblem keine eindeutige Lösung.

# Bemerkungen

- ▶ “Analogie” nichtlineare Gleichungssysteme.
- ▶ In einem **kritischen Punkt**  $\mathbf{x}^*$  von  $\phi$  muss die **Ableitung**

$$\nabla\phi(\mathbf{x}) = \mathbf{F}'(\mathbf{x})^T \mathbf{F}(\mathbf{x})$$

gleich Null  $\in \mathbb{R}^{m \times n}$  sein.

Als Abbruchkriterium für das Verfahren wird daher

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

häufig benutzt, wobei  $\varepsilon$  eine vorgegebene Toleranz ist.

- ▶ Der Erfolg des Gauß-Newton-Verfahrens hängt von der Wahl des Startwerts ab (vgl. Newton-Verfahren).
- ▶ Falls **Rang** $(\mathbf{F}'(\mathbf{x})) \neq n$  muss der Zusatz “mit minimaler 2-Norm” aufgenommen werden.

# Analyse der Gauß-Newton-Methode

Sei  $x^*$  ein kritischer Punkt von  $\phi$ , der in einer Umgebung  $U$  eindeutig ist.

Annahme:

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

Für  $x^k \in U$  hat das lineare Ausgleichsproblem die eindeutige Lösung

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

# Analyse der Gauß-Newton-Methode

Deshalb gilt für die Gauß-Newton-Iteration:

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

mit

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

Es gilt:

$$\mathbf{x} = \Phi(\mathbf{x}) \Leftrightarrow \nabla \phi(\mathbf{x}) = \mathbf{0} \Leftrightarrow \mathbf{x} = \mathbf{x}^* .$$

Die Gauß-Newton-Methode ist also eine Fixpunktiteration.



## Beispiel 6.4.

Lösen Sie das nichtlineare Ausgleichsproblem

$$\mathbf{x}^* = \arg \min_{x \in [0, 2\pi]} \|\mathbf{F}(x)\|_2,$$

wobei

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

- ▶ Für die Jacobi-Matrix erhält man

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

- ▶ Außerdem ergibt sich

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

und damit

$$\nabla \phi(x) = -a r \sin(x).$$

## Beispiel 6.4.

- ▶ Für die Iterationsfunktion zu  $F$  erhält man schließlich

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

- ▶ Es gibt zwei kritische Punkte von  $\phi$

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

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

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

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

und damit

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

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

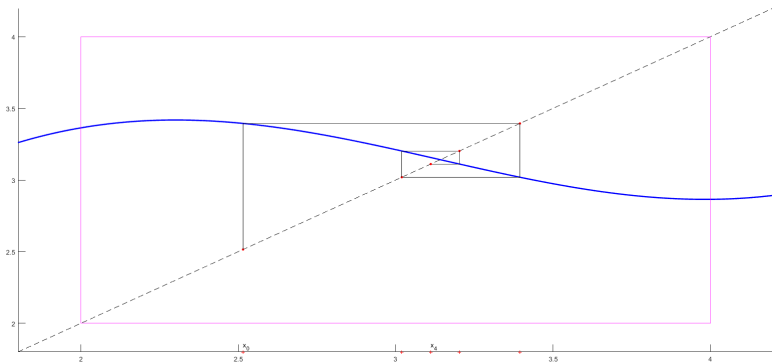
## Beispiel 6.4.

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

1. Das lokale Maximum ist abstoßend
2. Die Methode ist linear konvergent in einer Umgebung des lokalen Minimums (wenn  $a < 2r$ ),  
Falls  $a = r$  liegt sogar quadratische Konvergenz vor.  
oder
3. das lokale Minimum ist auch abstoßend (wenn  $a > 2r$ ).

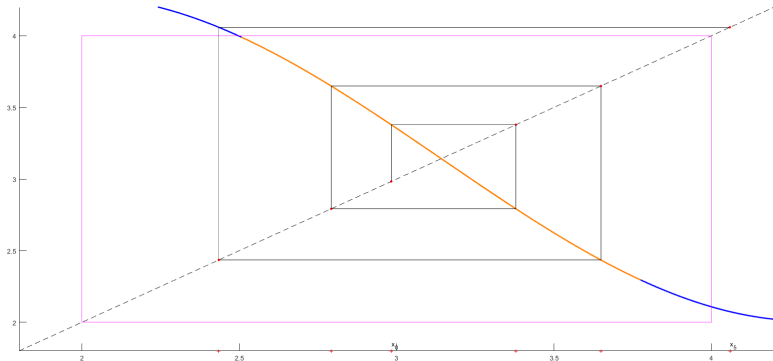
Man kann zeigen, dass ähnliche Eigenschaften in einem allgemeinen Rahmen gültig sind.

## Beispiel 6.4.

Konvergenter Fall:  $a < 2r$ 

# Beispiel 6.4.

Konvergenter Fall:  $a > 2r$



## Folgerung 6.6.

Für die allgemeine Gauß-Newton-Iterationsfunktion  $\Phi$  gilt

$$\|\Phi'(x^*)\|_A = \rho(K) \|F(x^*)\|_2,$$

$$\|\Phi'(x^*)\| \geq \rho(K) \|F(x^*)\|_2,$$

mit

$$K := A \left( \sum_{i=1}^m \frac{F_i(x^*)}{\|F(x^*)\|_2} F_i''(x^*) \right) A^{-1}$$

für jede Operatornorm  $\|\cdot\|$ .

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.

## Folgerung 6.6.

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.

### Beachte

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

## Folgerung 6.6.

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

Für ein lokales Minimum  $\mathbf{x}^*$  der Funktion  $\phi$  ist die lokale Konvergenz des Gauß-Newton-Verfahrens gesichert, falls die Bedingung

$$\rho(\mathbf{K}) \|F(\mathbf{x}^*)\|_2 < 1$$

erfüllt ist.

Sei  $\mathbf{x}^*$  ein lokales Minimum von  $\phi$ , wofür gilt:

$$\rho(\mathbf{K}) \|F(\mathbf{x}^*)\|_2 > 1.$$

Dann ist  $\|\Phi'(\mathbf{x}^*)\| > 1$  für jede Operatornorm  $\|\cdot\|$ .

Ein lokales Minimum von  $\phi$  kann für die Gauß-Newton-Methode also abstoßend sein.



## 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

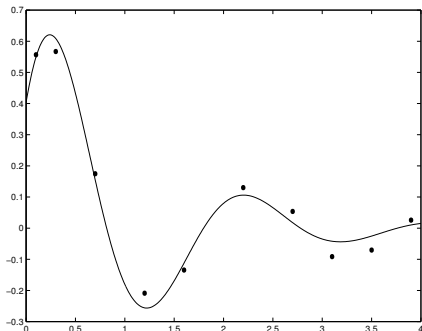
In der letzten Spalte der Tabelle sieht man das lineare Konvergenzverhalten des Gauß-Newton-Verfahrens.

## Beispiel 6.7.

Die berechneten Parameterwerte  $x^*$  aus dem 12. Iterationsschritt liefern die Lösung

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

die im folgenden Plot dargestellt ist.



# Levenberg-Marquardt-Verfahren

## Zur Erinnerung:

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

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

## Idee:

Einführung einer „Regularisierung“

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

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

# Levenberg-Marquardt-Verfahren

Lineares Ausgleichsproblem (Gauß-Newton)

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

wird **ersetzt** durch

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

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**.

# Levenberg-Marquardt-Verfahren

## Weitere günstige Eigenschaft

Es gilt

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

d.h.  $\mu$  groß  $\Rightarrow$  Korrektur  $s^k$  klein.

Die Methode erlaubt eine **Dämpfungsstrategie**.

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

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

Konvergenzordnung ist **1**, wie bei der Gauß-Newton-Methode.

**Geeignete Wahl von  $\mu$ : Einzugsbereich wird vergrößert.**

# Levenberg-Marquardt-Methode

## Algorithmus 6.10. (Levenberg-Marquardt)

Wähle Startwert  $x^0$  und Anfangswert für den Parameter  $\mu$ .

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  $\mu$  vergrößert und **Schritt 2** wiederholt.
4. Setze  $x^{k+1} = x^k + s^k$ .
5. Sind bestimmte Kriterien erfüllt sind, wird  $\mu$  verkleinert.

# Zusammenfassung

- ▶ Beim Gauß-Newton-Verfahren wird das **nichtlineare Ausgleichsproblem** über eine Folge **linearer Ausgleichsprobleme** gelöst.
- ▶ Lokale Konvergenz im allgemeinen nur **1. Ordnung**, falls  $F(x^*) = \mathbf{0}$ , sogar von **2. Ordnung**.
- ▶ Es kann lokale **Divergenz** auftreten.
- ▶ Matrix  $F'(x^k)$  kann  $\text{Rang} < n$  haben.
- ▶ Bei Levenberg-Marquardt:
  - ▶ Parameter  $\mu$  zur **Vergrößerung des Einzugsbereichs**.
  - ▶ Matrix  $\begin{pmatrix} F'(x^k) \\ \mu I \end{pmatrix}$  hat **Rang = n**.

# Verständnisfragen

Es sei  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  mit  $m > n$ .

Wir betrachten das (nichtlineare) Ausgleichsproblem:

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

- f** Die Gauß-Newton-Methode ist immer konvergent in einer hinreichend kleinen Umgebung von  $x^*$ .
- w** Beim Levenberg-Marquardt-Verfahren hat die Matrix des linearisierten Ausgleichsproblems in jedem Schritt stets vollen Rang.
- w** Die Gauß-Newton-Methode kann man als Fixpunktiteration darstellen.
- f** Die Konvergenzordnung der Gauß-Newton-Methode ist in der Regel 2.