

# 1 Wiederholung?

Wir starten mit einer kurzen Wiederholung zur Fixpunktiteration zum Lösen von Gleichungen der Form  $Tx = x$  durch  $x_{n+1} = Tx_n$ .

**Satz 1.1 (Banach 1922).** Sei  $M$  eine abgeschlossene nichtleere Teilmenge in einem vollständig metrischem Raum  $(X, d)$ . Sei  $T : M \rightarrow M$  eine Selbstabbildung und  $k$ -kontraktiv, d.h.  $d(Tx, Ty) \leq k \cdot d(x, y) \forall x, y \in M$  mit  $0 \leq k < 1$ . Dann folgt:

1. Existenz und Eindeutigkeit: die Gleichung  $Tx = x$  hat genau eine Lösung, d.h.  $T$  hat genau einen Fixpunkt in  $M$ .
2. Konvergenz der Iteration  $x_{k+1} = Tx_k$ . Die Folge  $(x_k)_{k \in \mathbb{N}}$  konvergiert gegen den Fixpunkt  $x^*$  für einen beliebigen Startpunkt  $x_0 \in M$ .
3. Fehlerabschätzung: Für alle  $n = 0, 1, \dots$  gilt
  - a-priori:  $d(x_n, x^*) \leq k^n(1 - k)^{-1}d(x_0, x_1)$
  - a-posteriori:  $d(x_{n+1}, x^*) \leq k(1 - k)^{-1}d(x_n, x_{n+1})$
4. Konvergenzrate: Für alle  $n \in \mathbb{N}$  gilt  $d(x_{n+1}, x^*) \leq k \cdot d(x_n, x^*)$

*Beweis.*

2. Wir zeigen, dass  $(x_n)$  eine Cauchy-Folge ist. Für den Abstand zweier benachbarter Folgeglieder  $x_n$  und  $x_{n+1}$  gilt

$$d(x_n, x_{n+1}) = d(Tx_{n-1}, Tx_n) \leq k \cdot d(x_{n-1}, x_n) \leq \dots \leq k^n \cdot d(x_0, x_1)$$

Mehrfache Anwendung der Dreiecksungleichung liefert daher für  $n, m \in \mathbb{N}$ :

$$\begin{aligned} d(x_n, x_{n+m}) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \dots + d(x_{n+m-1}, x_{n+m}) \\ &\leq (k^n + k^{n+1} + \dots + k^{n+m}) \cdot d(x_0, x_1) \\ &\leq k^n(1 + k + k^2 + \dots) \cdot d(x_0, x_1) \\ &= k^n \cdot (1 - k)^{-1}d(x_0, x_1) \end{aligned}$$

Demnach folgt  $d(x_n, x_{n+m}) \rightarrow 0$  für  $n \rightarrow \infty$  und da  $X$  vollständig ist konvergiert  $(x_n)$  gegen ein  $x^* \in X$ .

1. Da  $T$  stetig ist (aufgrund  $k$ -Kontraktivität) folgt für die konvergente Folge  $(x_n)$ , dass

$$x^* = \lim_{n \rightarrow \infty} x_{n+1} = \lim_{n \rightarrow \infty} Tx_n = Tx^*$$

Da  $M$  abgeschlossen ist existiert also ein Fixpunkt in  $M$ .

Dieser ist eindeutig, denn für  $x, y$  mit  $Tx = x$  und  $Ty = y$  gilt  $d(x, y) = d(Tx, Ty) \leq kd(x, y)$ , also  $d(x, y) = 0$ .

3. Aus dem Beweis zu 2. haben wir  $d(x_n, x_{n+m}) \leq k^n(1 - k)^{-1}d(x_0, x_1)$ , wegen der Stetigkeit der Metrik folgt die a-priori-Fehlerabschätzung aus  $m \rightarrow \infty$ .

Die a-posteriori-Fehlerabschätzung folgt analog aus dem Ansatz

$$\begin{aligned} d(x_{n+1}, x_{n+1+m}) &\leq d(x_{n+1}, x_{n+2}) + \dots + d(x_{n+m}, x_{n+1+m}) \\ &\leq (k + \dots + k^m) \cdot d(x_n, x_{n+1}) \\ &\leq k \cdot (1 - k)^{-1}d(x_n, x_{n+1}) \end{aligned}$$

4. Folgt direkt durch  $d(x_{n+1}, x^*) = d(Tx_n, Tx^*) \leq k \cdot d(x_n, x^*)$

**Beispiel 1.2.** Wir betrachten das Nullstellenproblem  $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \cos x - x = 0$ . Umformung ergibt  $\underbrace{\cos x}_{Tx} = x$  und somit die Fixpunktiteration  $x_{k+1} = Tx_k = \cos(x_k)$

**Abb. 1.1**

Prüfung der Voraussetzungen des Banach'schen FP-Satzes:

Wir wählen als Einschränkung  $M = [0, 1]$ , dies liefert uns eine Selbstabbildung auf einer abgeschlossenen Teilmenge  $M$  des vollständig metrischen Raum  $\mathbb{R}$  mit der Abstandsfunktion  $d(x, y) = |x - y|$ . Weiter ist die Abbildung  $k$ -kontraktiv: Nach Mittelwertsatz der Differentialrechnung gilt

$$|\cos x - \cos y| = \underbrace{|\sin \xi|}_{\leq \sin(1)} \cdot |x - y| \leq \underbrace{0,85}_{=:k} \cdot |x - y|, \quad \text{für } \xi \in [0, 1]$$

Wir können also nach Banach die Existenz und Eindeutigkeit eines Fixpunkt  $x^*$  folgern, diesen Fixpunkt finden wir durch die konvergente Folge  $x_{k+1} = \cos x_k$ .

Wir betrachten im folgenden die Idee der Umwandlung eines Nullstellenproblems in Fixpunkt-Gleichung noch etwas allgemeiner. Für eine Gleichung  $f(x) = 0$  mit  $f : \mathbb{R} \rightarrow \mathbb{R}$  haben wir verschiedene Möglichkeiten zur Umformung:

- a) Betrachte  $Tx := x - f(x)$  gefolgt aus  $f(x) = 0 \Leftrightarrow -f(x) = 0 \Leftrightarrow x - f(x) = x$ .
  - b) Betrachte  $Tx := x - \omega \cdot f(x)$  mit  $\omega \neq 0$  (lineare Relaxation)
  - c) Betrachte  $Tx := x - \omega \cdot g(f(x))$  mit  $\omega \neq 0$  und geeigneter Funktion  $g$  (nichtlineare Relaxation). Wenn  $g(0) \neq 0$  dann betrachte  $Tx := x - \omega \cdot (g(f(x)) + g(0))$
  - d) Betrachte  $Tx := x - (f'(x))^{-1}f(x)$  (Newtonverfahren)  
Newton hat teils Probleme, bei falschen Startwerten:
- Abb 1.2**
- e) Betrachte  $Tx := h^{-1}(f(x) - g(x))$ , wobei  $f(x) = h(x) + g(x)$  (Splitting-Verfahren)

## 2 Iteratives Vorgehen zur Lösung linearer Gleichungssysteme

### 2.1 Splittingverfahren

Gegeben sei das LGS  $Ax = b$  für  $A \in \mathbb{K}^{n \times n}, b \in \mathbb{K}^n, x \in \mathbb{K}^n$ , wobei  $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ . Wir wollen dieses LGS nun in ein FP-Problem umformen, sei hierfür  $A$  nicht singulär (sonst nicht lösbar).

Wir schreiben  $A = M - N$ , wobei  $M$  invertierbar und häufig sogar eine Diagonalmatrix ist (damit  $M$  leicht zu invertieren ist). Dies liefert:

$$Ax = b \Leftrightarrow (M - N)x = b \Leftrightarrow Mx = Nx + bx = \underbrace{M^{-1} \cdot (Nx + b)}_{\tilde{T}x}$$

$\tilde{T}$  ist affin-linear. Wir erhalten also unser FP-Problem  $x = \tilde{T}x = Tx + c$  mit  $T = M^{-1}N$  und  $c = M^{-1}b$

#### Algorithmus 1: Splittingverfahren

- 1 **Initialisierung:**  $A = M - N$  mit  $N \in GL(n, \mathbb{K})$
- 2 Wähle  $x^{(0)} \in \mathbb{K}^n$  beliebig
- 3 **for**  $k = 1, 0, \dots$
- 4 | löse  $Mx^k = Nx^{k-1} + b$
- 5 **until** stop (beliebiges Stopkriterium)

Konvergenz dieses Algorithmus folgt aus Banachschen Fixpunktsatz.

**Bemerkung 2.1.** Nach gleicher Überlegung lässt sich auch unser obiges Splittingverfahren für Nullstellenbestimmung herleiten:

$$f(x) = 0 \Leftrightarrow h(x) + g(x) := f(x) = 0 \Leftrightarrow h(x) = f(x) - g(x) \Leftrightarrow x = h^{-1}(f(x) - g(x))$$

*Wiederholung:* Eine Matrixnorm ist eine Norm auf dem Vektorraum der Matrizen, d.h.  $\|\cdot\| : \mathbb{K}^{n \times n} \rightarrow \mathbb{R}$ , bereits bekannte Matrixnormen sind:

- Frobeniusnorm:  $\|A\|_F := \left( \sum_{i,j} |a_{ij}|^2 \right)^{1/2}$
- Spaltensummennorm  $\|A\|_1 := \max_j \sum_i |a_{ij}|$
- Zeilensummennorm  $\|A\|_\infty := \max_i \sum_j |a_{ij}|$
- Spektralnorm  $\|A\|_2 := \sqrt{\lambda_{\max}(A^H A)}$ , ( $A^H := \overline{A}^T$ )

Im allgemeinen induziert eine Vektornorm auch immer eine Matrixnorm, diese nennen wir auch Operatornorm:

$$\|A\| := \max_{\|x\|=1} \|Ax\|$$

Die oben aufgelisteten Normen  $\|\cdot\|_1, \|\cdot\|_2$  und  $\|\cdot\|_\infty$  sind die Operatornormen zu der jeweiligen  $p$ -Normen.

## 2.1 Splittingverfahren

Eine Norm  $\|\cdot\|$  auf  $\mathbb{K}^{n \times n}$  heißt submultiplikativ, falls  $\|AB\| \leq \|A\| \cdot \|B\|$  und sie heißt verträglich mit einer Vektornorm  $\|\cdot\|_V$ , falls  $\|Ax\|_V \leq \|A\| \cdot \|x\|_V$ .

Operatornormen sind immer submultiplikativ und verträglich zu der Vektornorm, aus welcher sie abgeleitet wurden.

**Satz 2.2.** Ist  $\|\cdot\|$  eine Norm auf  $\mathbb{K}^{n \times n}$ , die mit einer Vektornorm verträglich ist, und ist  $\|M^{-1}N\| < 1$ , dann konvergiert der Algorithmus für jedes  $x^{(0)} \in \mathbb{K}^n$  gegen  $A^{-1}b$ , d.h. gegen die Lösung des linearen Gleichungssystems  $Ax = b$ .

*Beweis.* Sei  $\tilde{T}(x) := Tx + c$  mit  $T = M^{-1}N$  und  $c = M^{-1}b$ .

Offensichtlich gilt  $\tilde{T} : \mathbb{K}^n \rightarrow \mathbb{K}^n$ , sowie

$$\|\tilde{T}(x) - \tilde{T}(y)\| = \|Tx - Ty\| \leq \|T\| \cdot \|x - y\|$$

Da  $\|T\| = \|M^{-1}N\| < 1$  ist  $\tilde{T}$  eine  $k$ -kontraktive Selbstabbildung und somit konvergiert die Folge  $(x^k)$  aus dem Algorithmus gegen den eindeutigen Fixpunkt  $x^*$  mit  $\tilde{T}(x^*) = x^*$ .

Einsetzen der Definition von  $\tilde{T}$  liefert:

$$x^* = Tx + c = M^{-1}(Nx + b) \Rightarrow Mx = Nx + b \Rightarrow Ax = (M - N)x = b$$

**Korollar 2.3.** Sei  $A$  invertierbar, so konvergiert der obige Algorithmus genau dann für alle Startwerte  $x^{(0)} \in \mathbb{K}^n$  gegen  $x^* = A^{-1}b$ , wenn für den Spektralradius  $\rho(T) = \max\{|\lambda| : \lambda \in \sigma(T)\}$  die Ungleichung  $\rho(T) < 1$  erfüllt ist.

*Beweis.*

$\Leftarrow$ : Falls  $\rho(T) < 1$  dann existiert eine Norm  $\|\cdot\|_\varepsilon$  auf  $\mathbb{K}^n$  und eine dadurch induzierte Operatornorm  $\|\cdot\|_\varepsilon$  auf  $\mathbb{K}^{n \times n}$  mit  $\|T\|_\varepsilon \leq \rho(T) + \varepsilon < 1$  **Warum?**

Satz 2.2 liefert dann die Konvergenz des Algorithmus.

$\Rightarrow$ : Angenommen  $\rho(T) \geq 1$ , d.h. es existiert ein Eigenwert  $\lambda$  von  $T$  mit  $|\lambda| \geq 1$  und zugehörigem Eigenvektor  $z$ . Für  $x^{(0)} = x^* + z$  und festes  $k$  sich der Iterationsfehler

$$x^{(k)} - x^* = Tx^{(k-1)} + c - x^* = Tx^{(k-1)} - Tx^* = T(x^{(k-1)} - x^*)$$

Induktiv folgt dann  $x^{(k)} - x^* = T^k(x^{(0)} - x^*) = T^k z = \lambda^k z$ , demnach gilt  $\|x^{(k)} - x^*\| = |\lambda|^k \cdot \|z\|$ . Für größer werdendes  $k$  kann  $x^{(k)}$  also nicht gegen  $x^*$  konvergieren.

**Satz 2.4.** Unter gleichen Voraussetzungen des obigen Korollars gilt

$$\max_{x^{(0)} \in \mathbb{K}^n} \limsup_{k \rightarrow \infty} \|x^* - x^{(k)}\|^{1/k} = \rho(T)$$

*Beweis.* Aus dem Beweis von Korollar 2.3 sehen wir

$$\max_{x^{(0)} \in \mathbb{K}^n} \limsup_{k \rightarrow \infty} \|x^* - x^{(k)}\|^{1/k} \geq \limsup_{k \rightarrow \infty} \|T^k z\|^{1/k} = \limsup_{k \rightarrow \infty} |\lambda| \cdot \|z\|^{1/k} = |\lambda| = \rho(T)$$

Für jeden Startwert  $x^{(0)} \in \mathbb{K}^n$  gilt nun

$$\|x^{(k)} - x^*\|_\varepsilon = \|T^k(x^{(0)} - x^*)\|_\varepsilon \leq \|T\|_\varepsilon^k \cdot \|x^{(0)} - x^*\|_\varepsilon$$

Da im  $\mathbb{K}^n$  alle Normen äquivalent sind, also insbesondere auch  $\|\cdot\|_\varepsilon$  und  $\|\cdot\|$ , existiert eine Konstante  $c_\varepsilon > 0$ , so dass

$$\|x^{(k)} - x^*\|^{1/k} \leq \left(c_\varepsilon \cdot \|x^{(k)} - x^*\|_\varepsilon\right)^{1/k} \leq \|T\|_\varepsilon \cdot \left(c_\varepsilon \cdot \|x^{(0)} - x^*\|_\varepsilon\right)^{1/k} \xrightarrow{k \rightarrow \infty} \|T\|_\varepsilon$$

## 2.1 Splittingverfahren

Folglich ist

$$\varrho(T) \leq \max_{x^{(0)}} \limsup_{k \rightarrow \infty} \|x^{(k)} - x^*\|^{1/k} \leq \|T\|_\varepsilon$$

□Dieser Satz ermöglicht es nun einen sinnvollen Begriff der Konvergenzrate zu definieren:

### Definition 2.5.

Die Zahl  $\varrho(T)$  heißt (asymptotischer) Konvergenzfaktor von der Iteration  $x^{(k)} = Tx^{(k-1)} + c$ . Die (asymptotische) Konvergenzrate lässt sich dadurch ausdrücken mit  $r = -\log_{10} \varrho(T)$

Mittels der Zerlegung  $A = D + L + R$ , wobei  $D$  die Diagonale,  $L$  die untere (linke) Hälfte und  $R$  die obere (rechte) Hälfte der Matrix  $A$  sind, erhalten wir einen Spezialfall der Splitting-Verfahren. Durch die Wahl  $M = D$  und  $N = L + R$  ergibt sich  $x^{(k+1)} = D^{-1}(b - (L + R)x^{(k)})$ , bzw. in algorithmischer Form:

### Algorithmus 2: Jacobi / Gesamtschritt Verfahren

Gegeben sei das Lineare Gleichungssystem  $Ax = b$  mit  $a_{ii} \neq 0$ .

```

1 Initialisierung: Wähle beliebigen Startvektor  $x^{(0)} \in \mathbb{K}^n$ 
2 for  $k = 1, 0, \dots$ 
3   for  $i = 1, \dots, n$ 
4      $x_i^{(k+1)} \leftarrow \frac{1}{a_{ii}} \left( b_i - \sum_{j \neq i} a_{ij} x_j^{(k)} \right)$ 
5   end
6 until stop (beliebiges Stopkriterium)
```

Die zugehörige Iterationsmatrix ist hierbei  $J = M^{-1}N = D^{-1}(L + R)$  und nennt sich (beim Jacobi Verfahren) Gesamtschrittoperator.

Einen weitere Version des Splitting-Verfahren ergibt sich durch die Wahl  $M = D - L$  und  $N = R$ . Hierbei bildet  $D - L$  eine obere Dreiecksmatrix und die Inversion ergibt sich mittels Vorwärtssubstitution:

### Algorithmus 3: Gauss-Seidel / Einzelschritt Verfahren

Gegeben sei das Lineare Gleichungssystem  $Ax = b$  mit  $a_{ii} \neq 0$ .

```

1 Initialisierung: Wähle beliebigen Startvektor  $x^{(0)} \in \mathbb{K}^n$ 
2 for  $k = 1, 0, \dots$ 
3   for  $i = 1, \dots, n$ 
4      $x_i^{(k+1)} \leftarrow \frac{1}{a_{ii}} \left( b_i - \sum_{j < i} a_{ij} x_j^{(k+1)} - \sum_{j > i} a_{ij} x_j^{(k)} \right)$ 
5   end
6 until stop (beliebiges Stopkriterium)
```

Die hier erhaltene Iterationsmatrix nennen wir Einzelschrittoperator  $L = (D - L)^{-1}R$ . Mittels der Zeilensummennorm erhalten wir nun ein leicht prüfbares Konvergenzkriterium:

**Satz 2.6.** Ist  $A \in \text{GL}_n(\mathbb{K})$  strikt diagonaldominant, d.h.  $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ , dann konvergieren Jordan und Gauss-Seidel Verfahren für alle Startwerte  $x^{(0)} \in \mathbb{K}^n$  gegen die eindeutige Lösung von  $Ax = b$ .

## 2.1 Splittingverfahren

*Beweis.*

Da  $A$  strikt diagonaldominant ist, muss  $a_{ii} \neq 0$  und damit sind beide Verfahren wohldefiniert.

a) Jacobi Verfahren: Für die Iterationsmatrix gilt

$$\|J\|_{\infty} = \|D^{-1}(L + R)\|_{\infty} = \max_{i \in [n]} \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| =: q < 1$$

Nach Satz 2.2 folgt damit die Konvergenz des Jacobi Verfahren.

b) Gauss-Seidel Verfahren: Um  $\|L\|_{\infty} < 1$  zu zeigen, nutzen wir, dass die Zeilensummennorm die Operatornorm induziert durch die Maximumsnorm ist, d.h.

$$\|L\|_{\infty} = \max_{\|x\|_{\infty}=1} \|Lx\|_{\infty}$$

Sei nun  $y = Lx$  für ein  $x \in \mathbb{K}^n$  mit  $\|x\|_{\infty} = 1$ .

Induktiv folgt nun  $y_i \leq q < 1$ , der Induktionsanfang folgt dabei aus dem Beweisteil a).

Unter der Induktionsvoraussetzung gilt für  $j < i$ , dass  $|y_j| \leq q$  und damit:

$$\begin{aligned} \|y_i\| &\leq \frac{1}{|a_{ii}|} \left( \sum_{j < i} |a_{ij}| \cdot \underbrace{|y_j|}_{\leq q} + \sum_{j > i} |a_{ij}| \cdot \underbrace{|x_j|}_{\leq \|x\|_{\infty}} \right) \\ &\leq \frac{1}{|a_{ii}|} \left( \sum_{j < i} |a_{ij}| \cdot q + \sum_{j > i} |a_{ij}| \cdot \|x\|_{\infty} \right) \\ &< \frac{1}{|a_{ii}|} \left( \sum_{j < i} |a_{ij}| + \sum_{j > i} |a_{ij}| \right) \\ &= \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| \\ &= q \end{aligned}$$

Da dies für alle Einträge von  $y$  gilt folgt  $\|y\|_{\infty} = \|Lx\|_{\infty} \leq q$  für alle  $x$  mit  $\|x\|_{\infty} = 1$  und damit  $\|L\|_{\infty} \leq q < 1$   $\square$

**Beispiel 2.7.** Gegeben sei das LGS  $Ax = b$  mit

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 1 & -4 & 1 \\ 0 & -1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}$$

Dieses System hat die eindeutige Lösung  $x^* = (1, -1, -1)^T$ .

Durch die Wahl  $x^{(0)} = (1, 1, 1)^T$  erhalten wir beim Jacobi Verfahren:

$$\begin{aligned} x^{(1)} &= D^{-1}(b - (L + R)x^{(0)}) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \cdot \left[ \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right] = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ 0 \end{pmatrix} \\ x^{(2)} &= D^{-1}(b - (L + R)x^{(1)}) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \cdot \left[ \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ -\frac{1}{2} \\ 0 \end{pmatrix} \right] = \begin{pmatrix} \frac{1}{2} \\ -1 \\ -\frac{3}{4} \end{pmatrix} \\ &\vdots \end{aligned}$$

## 2.2 Gradientenverfahren

**Motivation:** Eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  soll minimiert werden. Von einem Startpunkt  $x^{(0)}$  ausgehen bewegen wir uns nun Stück für Stück in Richtung des steilsten Abstiegs, intuitiv sollten wir so ein Minimum finden.

Als Iterationsvorschrift ergibt sich  $x^{(k+1)} = x^{(k)} + \alpha^{(k)} \cdot d^{(k)}$ ,  $k = 0, 1, \dots$

dabei ist  $\alpha^{(k)} > 0$  die Schrittweite und Abstiegsrichtung  $d^{(k)} \in \mathbb{R}^n$ . (Eine typische Wahl der Abstiegsrichtung ist  $d^{(k)} = -\partial f / \partial x(x^{(k)}) = -\nabla f(x^{(k)})$ )

Das Ziel ist des Verfahren ist es, dass sich der Wert von  $f$  in jedem Schritt verbessert, d.h.  $f(x^{(k+1)}) < f(x^{(k)})$ . Es ergibt sich ein 1-dim. Optimierungsproblem für die Schrittweite  $\alpha^{(k)}$ :

$$\alpha^{(k+1)} = \min_{\alpha \neq 0} \{f(x^{(k)} + \alpha \cdot d^{(k)})\}$$

Ein Nachteil des Verfahren ist, dass ein sogenannter „Zick-Zack-Kurs“ entstehen kann.

**Verfahren der konjugierten Gradienten:** Die obige Idee kann zur effizienten Lösung linearer Gleichungssysteme genutzt werden. Gegeben sei das LGS  $Ax = b$  mit (zunächst)  $A \in \mathbb{R}^{n \times n}$  hermitisch, d.h.  $a_{ij} = \overline{a_{ji}}$  (hieraus folgt insbesondere, dass die Hauptdiagonale reell ist). Zur Lösung wird hierbei die Minimierung des quadratischen Funktionals

$$\phi(x) = \frac{1}{2} x^* A x - x^* b$$

Sollte eine Lösung  $\hat{x} = A^{-1}b$  des LGS  $Ax = b$  existieren, so gilt für alle  $x \in \mathbb{R}^{n \times n}$ :

$$\begin{aligned} \phi(x) - \phi(\hat{x}) &= \frac{1}{2} x^* A x - x^* b - \left( \frac{1}{2} \hat{x}^* A \hat{x} - \hat{x}^* b \right) \\ &\vdots \\ &= \frac{1}{2} (x - \hat{x})^* A (x - \hat{x}) \geq 0 \end{aligned}$$

Die Funktion hat demnach ein eindeutiges Minimum bei  $\hat{x}$ .

**Definition 2.8.** Ist  $A \in \mathbb{R}^{n \times n}$  hermitisch und pos. definitiv, dann wird durch  $\|x\|_A = \sqrt{x^* A x}$ ,  $x \in \mathbb{R}^{n \times n}$  eine Norm in  $\mathbb{R}^n$  definiert, die sogenannte Energienorm. Zur Energienorm gehört ein inneres Produkt, nämlich  $\langle x, y \rangle_A = x^* A y$ ,  $x, y \in \mathbb{R}^n$ . Mithilfe dieser Definition und obiger Erkenntnis ergibt sich die Abweichung des Funktionals von seinem Minimum:

$$\phi(x) - \phi(\hat{x}) = \frac{1}{2} \|x - \hat{x}\|_A^2$$

**geometrische Interpretation:** Der Graph von  $\phi$  bezüglich der Energienorm ist ein kreisförmiger Paraboloid, welcher über dem Mittelpunkt  $\hat{x}$  liegt.

**Idee:** Konstruktion eines Verfahrens, welches die Lösung  $\hat{x}$  von  $Ax = b$  iterativ approximiert, indem das Funktional  $\phi$  zukzessiv minimiert wird. Es gilt

$$\phi(x^{(k)} + \alpha d^{(k)}) = \phi(x^{(k)}) + \alpha d^{(k)*} A x^{(k)} + \frac{1}{2} \alpha^2 d^{(k)*} A d^{(k)} - 2 d^{(k)*} \cdot b \quad (4)$$

Durch Differentiation und Null setzen der Ableitung ergibt sich die Schrittweite  $\alpha^{(k)}$ :

$$\alpha^{(k)} = \frac{r^{(k)*} d^{(k)}}{d^{(k)*} A d^{(k)}}, \quad \text{mit } r^{(k)} = b - A x^{(k)} \quad (5)$$

Weiter ergibt sich die Suchrichtung  $d^{(k+1)}$ :

$$d^{(k+1)} = r^{(k+1)} + \beta^{(k)} d^{(k)}, \quad (6)$$

$$\text{mit } \beta^{(k)} = - \frac{r^{(k+1)*} A d^{(k)}}{d^{(k)*} A d^{(k)}} \quad (7)$$

Die Gleichungen (5) und (7) sind wohldefiniert, wenn  $d^{(k)*} A d^{(k)} \neq 0$ , aufgrund der positiv Definitheit von  $A$  ist dies genau dann der Fall wenn  $d^{(k)} \neq 0$ . Nach (6) ist  $d^{(k)} = 0$  jedoch nur dann möglich,

## 2.2 Gradientenverfahren

wenn  $r^{(k)}$  und  $d^{(k-1)}$  linear abhängig sind, doch nach Definition verläuft die Suchrichtung tangential zur Niveaufläche von  $\phi$ , also orthogonal zum Gradienten  $r^{(k)}$ . Somit folgt  $d^{(k)} = 0$  nur wenn  $r^{(k)} = 0$ , was  $x^{(k)} = \hat{x}$  implizieren würde.

Wegen der zusätzlichen Orthogonalitätsbedingung  $\langle d^{(k+1)}, d^{(k)} \rangle_A = 0$  nennt man die Suchrichtungen zueinander  $A$ -konjugiert und das Verfahren, Verfahren der konjugierten Gradienten (CG-Verfahren).