Seminar *Newton- und Newton-ähnliche Verfahren*

Anwendung affin invarianter Konvergenzsätze auf newtoneske Verfahren

Hans Joachim Ferreau

`ferreau@urz.uni-heidelberg.de`

03. November 2004

Inhalt

Residuen-basierte Algorithmen

- **Residuen-basierte Algorithmen**

  Konvexe Optimierung

  Literatur

Einleitung

Einleitung · Newton-Verfahren

- Lösen der Gleichung

$$F(x) = 0$$

- z. B. mit dem gewöhnlichen Newton-Verfahren

$$F'(x^k)\Delta x^k = -F(x^k)\,,\; x^{k+1} = x^k + \Delta x^k \quad \forall\, k = 0, 1, \ldots$$

- Transformiertes Problem

$$G(y) := F(By) = 0\,,\; x = By\,,\; B \in \mathsf{GL}(n, \mathbb{R})$$

Einleitung · Affine Transformation

- Dann gilt

$$G'(y) = F'(By)B$$

- Und gewöhnliches Newton-Verfahren transformiert sich

$$G'(y^k)\Delta y^k = -G(y^k) \iff F'(By^k)B\Delta y^k = -F(By^k)$$

$$\iff F'(x^k)\underbrace{B\Delta y^k}_{=\Delta x^k} = -F(x^k)$$

- Korrekturen werden exakt so wie der gesamte Definitionsbereich transformiert

Einleitung · Affine Kontravarianz

- Mit $y^0 := B^{-1}x^0$ gilt dies auch für die Iterierten selbst
- Diese Eigenschaft bezeichnet man als **affine Kontravarianz**
- Dagegen bleiben die Residuen $F(x^k)$ bzw. ihre Norm $\|F(x^k)\|$ unverändert:

$$G(y^k) := F(By^k) = F(x^k)$$

- Deshalb **Residuen-basierte** Konvergenzkriterien

Gewöhnliches Newton-Verfahren

Gewöhnliches Newton-Verfahren · Newton-Mysovskikh

- Satz (Newton-Mysovskikh, affin kontravariante Version)
  Sei $F : \mathcal{D} \to \mathbb{R}^n$ eine differenzierbare Abbildung, $\mathcal{D} \subseteq \mathbb{R}^n$ offen und konvex, und gelten folgende Bedingungen:
  1. $F'(x)$ ist für alle $x \in \mathcal{D}$ invertierbar.
  2. Es gilt die affin kontravariante Lipschitz-Bedingung:
  $$\exists\, \omega > 0 : \ \left\| \left(F'(y) - F'(x)\right)(y - x) \right\| \leq \omega \left\| F'(x)(y - x) \right\|^2 \ \forall\, x, y \in \mathcal{D}\,.$$
  3. Durch
  $$\mathcal{L}_\omega := \left\{ x \in \mathcal{D} \;\middle|\; \|F(x)\| < \frac{2}{\omega} \right\}$$
  ist eine offene Menge definiert und $\overline{\mathcal{L}_\omega} \subset \mathcal{D}$ ist beschränkt.

Gewöhnliches Newton-Verfahren · Newton-Mysovskikh

- 4. Es gilt $x^0 \in \mathcal{L}_\omega$, also

$$h_0 := \omega \left\| F(x^0) \right\| < 2\,.$$

Dann gilt:

1. Die Iterierten $\{x_k\}$ des gewöhnlichen Newton-Verfahrens bleiben in $\mathcal{L}_\omega$ und konvergieren gegen eine Lösung $x^* \in \mathcal{L}_\omega$.
2. Die Residuen $\left\{F(x^k)\right\}$ konvergieren gegen 0 mit der Konvergenzrate

$$\left\| F(x^{k+1}) \right\| \leq \frac{\omega}{2} \left\| F(x^k) \right\|^2\,.$$

Gewöhnliches Newton-Verfahren · Kriterien

- Wir definieren die Kontraktionsfaktoren

$$\Theta_k := \frac{\left\| F(x^{k+1}) \right\|}{\left\| F(x^k) \right\|}$$

- und die Kantorowitsch-Werte

$$h_k := \omega \left\| F(x^k) \right\|$$

- Es gilt

$$h_{k+1} \leq \frac{1}{2} h_k^2$$

Gewöhnliches Newton-Verfahren · Kriterien

- Damit folgt

$$\Theta_k = \frac{h_{k+1}}{h_k} \leq \frac{1}{2} h_k < 1$$

und wir erhalten das

- **Startwert-Kriterium:** Falls

$$\Theta_0 \geq 1$$

gilt, weisen wir den Startwert x^0 als *nicht hinreichend nahe* zurück.

Gewöhnliches Newton-Verfahren · Kriterien

- Man kann außerdem

$$\Theta_{k+1} \approx \Theta_k^2 < \Theta_k < \ldots < \Theta_0$$

zeigen. Dies ergibt das

- **Divergenz-Kriterium 1:** Wir brechen die gewöhnliche Newton-Iteration als *divergent* ab, falls

$$\Theta_{k+1} > \Theta_0 \, .$$

Gewöhnliches Newton-Verfahren · Kriterien

- **Divergenz-Kriterium 2:** Wir brechen die gewöhnliche Newton-Iteration als *divergent* ab, falls

$$\Theta_{k+1} \geq 2\Theta_k^2 \,.$$

- **Konvergenz-Kriterium:** Wir beenden die gewöhnliche Newton-Iteration, falls

$$\left\| F(x^k) \right\| \leq FTOL$$

erfüllt ist und geben x^k als Lösung aus.

Vereinfachtes Newton-Verfahren

Vereinfachtes Newton-Verfahren · Grundlagen

- Benutze Jacobi-Matrix im Startpunkt x^0 während der gesamten Iteration als Näherung für $F'(x^k)$:

$$F'(x^0)\overline{\Delta x}^k = -F(x^k)\,,\ x^{k+1} = x^k + \overline{\Delta x}^k \quad \forall\, k = 0, 1, \ldots$$

- Weniger Rechenaufwand pro Iteration (keine Inversenberechnung)
- Dafür mehr Iterationen nötig

Vereinfachtes Newton-Verfahren · Newton-Kantorowitsch

- Satz (Newton-Kantorowitsch, affin kontravariante Version)
  Sei $F : \mathcal{D} \to \mathbb{R}^n$ eine stetig-differenzierbare Abbildung, $\mathcal{D} \subseteq \mathbb{R}^n$ konvex, und gelten folgende Bedingungen:

  1. Es exitstiert ein $0 \leq \omega < \infty$ so, dass $\forall\, x \in \mathcal{D}$, $v \in \mathbb{R}^n$:
  $$\left\| \Big(F'(x) - F'(x^0) \Big) v \right\| \leq \omega \left\| F'(x^0)(x - x^0) \right\| \left\| F'(x^0) v \right\| .$$
  2. Der Abschluss $\overline{\mathcal{L}_\omega} \subseteq \mathcal{D}$ ist beschränkt, wobei
  $$\mathcal{L}_\omega := \left\{ x \in \mathbb{R}^n \;\middle|\; \|F(x)\| \leq \frac{1}{2\omega} \right\} .$$
  3. Für den Startwert x^0 gilt
  $$h_0 := \omega \left\| F(x^0) \right\| \leq \frac{1}{2} .$$

Vereinfachtes Newton-Verfahren · Newton-Kantorowitsch

- Dann gilt:
  1. Die Iterierten $\{x_k\}$ des vereinfachten Newton-Verfahrens bleiben in $\mathcal{L}_\omega$ und konvergieren gegen eine Lösung $x^* \in \mathcal{L}_\omega$.
  2. Die Norm der Residuen konvergiert mit der Rate

$$\frac{\left\|F(x^{k+1})\right\|}{\left\|F(x^k)\right\|} \leq \frac{1}{2}(t_k + t_{k+1}) < 1 - \sqrt{1 - 2h_0}$$

gegen 0, wobei die $\{t_k\}$ wie folgt definiert sind:

$$t_0 := 0\,, \quad t_{k+1} = h_0 + \frac{1}{2}t_k^2$$

Vereinfachtes Newton-Verfahren · Kriterium

- Man kann also i. Allg. nur noch *lineare* Konvergenz erwarten.
- **Startwert-Kriterium:** Falls

$$\Theta_0 > \frac{1}{4}$$

gilt, weisen wir den Startwert x^0 als *nicht hinreichend nahe* zurück.

Quasi-Newton-Verfahren

Quasi-Newton-Verfahren · Grundlagen

- Broyden stellte 1965 Quasi-Newton-Methoden vor:

$$J_k \delta x^k = -F(x^k)\,,\; x^{k+1} = x^k + \delta x^k \quad \forall\, k = 0, 1, \dots$$

- Dabei wird die Jacobi-Matrix im Punkt x_k durch eine Näherung J_k approximiert.
- Eine spezielle Variante, die Broyden auf Grund praktischer Tests als „schlecht“ bezeichnete, spiegelt affine Kontravarianz wider

Quasi-Newton-Verfahren · Eine Iteration

- Sei

$$J_{k+1}^{-1} = J_k^{-1}\left(Id - \frac{F(x^{x+1})(\delta F_{k+1})^T}{\|\delta F_{k+1}\|^2}\right)$$

($\delta F_{k+1} := F(x^{k+1}) - F(x^k)$) das affin kontravariante schlechte Broyden-Rang-1-Update und gelte die Residuen-Kontraktion

$$\Theta_k := \frac{\left\|F(x^{k+1})\right\|}{\left\|F(x^k)\right\|} < 1\,.$$

Definiert man

$$E_k(J) := Id - J_k J^{-1}\,,$$

dann gilt:

Quasi-Newton-Verfahren · Eine Iteration

1. Die Update-Matrix J_{k+1} erfüllt

$$\begin{aligned} \|E_k(J_{k+1})\| &\leq \|E_k(J)\| \quad \forall\, J : E_k(J)\delta F_{k+1} = F_{k+1} \quad \text{und} \\ \|E_k(J_{k+1})\| &\leq \frac{\Theta_k}{1-\Theta_k}\,. \end{aligned}$$

2. Die Update-Matrix J_{k+1} ist regulär, wenn J_k regulär ist und besitzt die Darstellung

$$J_{k+1} = \left(Id - \frac{F(x^{x+1})(\delta F_{k+1})^T}{(\delta F_{k+1})^T F(x^k)} \right) J_k\,.$$

3. Die nächste Quasi-Newton-Korrektur ist

$$\delta x^{k+1} = -J_{k+1}^{-1} F(x^{k+1}) = \left(1 - \frac{(\delta F_{k+1})^T F(x^{x+1})}{\|\delta F_{k+1}\|^2} \right) \left(-J_k^{-1} F(x^{k+1}) \right)\,.$$

Quasi-Newton-Verfahren · Gesamte Iteration

- Sei $F : \mathcal{D} \to \mathbb{R}^n$ eine stetig-differenzierbare Abbildung, $\mathcal{D} \subseteq \mathbb{R}^n$ konvex, und x^* der eindeutige Lösungspunkt von F mit $F'(x^*)$ regulär. Gelten ferner folgende Bedingungen:
  1. Es exitstiert ein $\omega < \infty$ so, dass $\forall\, x, y \in \mathcal{D}$
  $$\left\| \left(F'(x) - F'(x^*)\right)(y - x)\right\| \leq \omega \left\|F'(x^*)(x - x^*)\right\| \left\|F'(x^*)(y - x)\right\| .$$
  2. Für obige Quasi-Newton-Iteration gibt es ein $0 < \overline{\Theta} < 1$ so, dass
     (a) die Näherung J_0 der Jacobi-Matrix der Ungleichung
     $$\overline{\eta}_0 := \|E_0\| < \overline{\Theta}$$
     genügt (wobei $E_k := Id - F'(x^*)J_k^{-1}$) und
     (b) der Startwert x^0 erfüllt die Ungleichung
     $$t_0 := \omega \left\|F'(x^*)(x^0 - x^*)\right\| \leq \frac{\overline{\Theta} - \overline{\eta}_0}{1 + \overline{\eta}_0 + \frac{4}{3}(1 - \overline{\Theta})^{-1}}$$

Quasi-Newton-Verfahren · Gesamte Iteration

- Dann gilt:
  1. Die Quasi-Newton-Iterierten $\left\{x^k\right\}$ konvergieren gegen x^* mit
  $$\left\|F(x^{k+1})\right\| \leq \overline{\Theta} \left\|F(x^k)\right\| \, .$$
  2. Die Konvergenz ist sogar *superlinear*.
  3. Die Abweichung der Näherungen der Jacobi-Matrix ist beschränkt durch
  $$\|E_k\| \leq \overline{\eta}_0 + \frac{t_0}{(1-t_0)(1-\overline{\Theta})} \leq \overline{\Theta}$$
  und es gilt die asymptotische Gleichung
  $$\lim_{k\to\infty} \frac{\|E_k \delta F_{k+1}\|}{\|\delta F_{k+1}\|} = 0 \, .$$

Quasi-Newton-Verfahren · QNRES

```
// QNRES (Codefragment)

F[0] = func(x);                       // F_0 := F(x^0) = func(x^0)
sigma[0] = F[0]*F[0];                 // σ_0 := ||F_0||^2

J0 = GetJacobiMatrix(x);              // J_0 = F'(x_0)
J0.Solve(dx,-F[0]);                   // Löse J_0 δx^0 = -F_0

double kappa = 1;                     // κ := 1
int k = 0;

while (k < kmax);
{
    ++k;

    x = x + dx;                       // x^{k+1} := x^k + δx^k
```

Kommentare: $F_0 := F(x^0) = func(x^0)$; $\sigma_0 := \|F_0\|^2$; $J_0 = F'(x_0)$; Löse $J_0 \delta x^0 = -F_0$; $\kappa := 1$; $x^{k+1} := x^k + \delta x^k$

Quasi-Newton-Verfahren · QNRES

```
F[k+1] = func(x);                               // F_{k+1} := F(x^{k+1})
dF[k+1] = F[k+1] - F[k];                        // δF_{k+1} := F_{k+1} − F_k

sigma[k+1] = F[k+1]*F[k+1];                     // σ_{k+1} := ||F_{k+1}||^2
if (sigma[k+1] <= FTOL*FTOL)                    // σ_{k+1} ≤ FTOL^2?
    { status = solutionfound; break; }          // Lösung gefunden!

theta[k] = sqrt(sigma[k+1]/sigma[k]);           // Θ_k := sqrt(σ_{k+1}/σ_k)
if (theta[k] >= thetamax)                       // Θ_k ≥ Θ_max?
    { status = noconvergence; break; }          // Keine Konvergenz!

kappa = kappa/(1-2*theta[k]);                   // κ := κ/(1 − 2Θ_k)
if (kappa >= kappamax)                          // κ ≥ κ_max?
    { status = illconditioned; break; }         // Schlecht konditioniert!

gamma[k] = dF[k+1]*dF[k+1];                     // γ_k := ||δF_{k+1}||^2
```

Quasi-Newton-Verfahren · QNRES

```
    v = F[k+1]*(1 - dF[k+1]*F[k+1]/gamma[k]);   // v := F_{k+1}(1 - <δF_{k+1}, F_{k+1}> / γ_k)
    double beta = 0;
    for (int j=k-1; j>=0; --j)
    {
        beta = dF[j+1]*v/gamma[j];              // β := <δF_{j+1}, v> / γ_j
        v = v - F[j+1]*beta;                    // v := v - βF_{j+1}
    }
    J0.Solve(dx,-v);                            // Löse J_0 δx^{k+1} = -v
}
```

- Nur die Residuen $F_0, \dots, F_{k+1}$ und die Differenzen $\delta F_1, \dots, \delta F_{k+1}$ müssen gespeichert werden.
- Für Θ_{max} lässt sich ein sinnvoller Wert von $\approx \frac{1}{4}$ ableiten.

Inexaktes Newton-Verfahren

Inexaktes Newton-Verfahren · Grundlagen

- Wenn die Zahl der Unbekannten sehr groß ist oder
- wenn die Iterierten noch weit von der Lösung entfernt sind, verwendet man inexakten Newton-Verfahren:

$$F'(x^k)\delta x^k = -F(x^k) + r^k \,, \text{ mit } \frac{\|r_k\|}{\|F(x^k)\|} \leq \eta_k \,,$$

$$x^{k+1} = x^k + \delta x^k \quad \forall\, k = 0, 1, \ldots$$

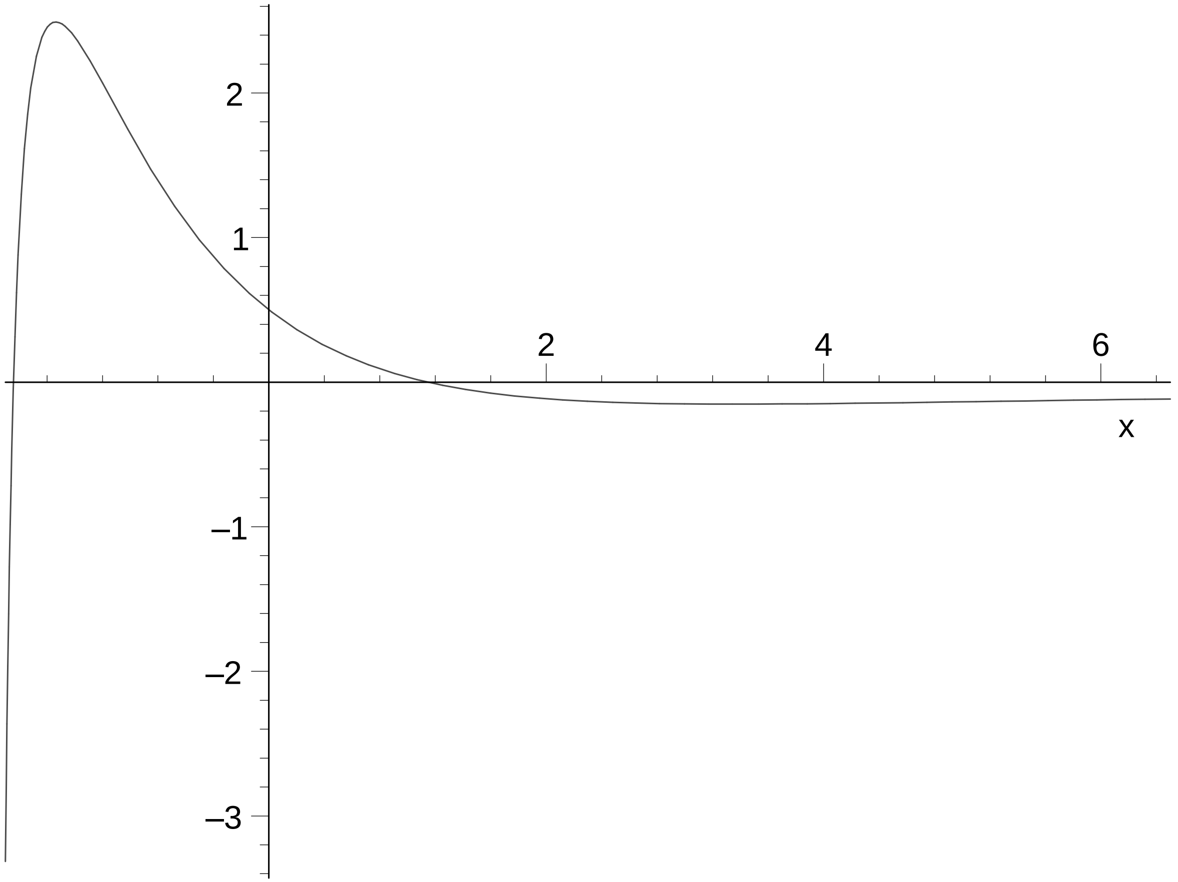
- 1982 wurde in gezeigt, dass das inexakte Newton-Verfahren mit $\eta_k < 1 \,\forall\, k$ lokal konvergiert

Inexaktes Newton-Verfahren · Grundlagen

- Die Wahl des Parameters η_k ist entscheidend:
  - Sehr genaues Lösen des inneren Gleichungssystems unnötig, falls die (äußeren) Newton-Iterierten noch weit von der Lösung entfernt sind
  - Durch zu ungenaues Lösen des inneren Gleichungssystems kann die Konvergenz des Newton-Verfahrens verlangsamt werden.
- **Ziel:** η_k so steuern, dass eine möglichst schnelle Konvergenz erreicht wird

Numerische Tests

Numerische Tests · Funktion 1



$$f(x) = \exp(-x) + \frac{1}{x+2}$$

Numerische Tests · Funktion 1

$$x^* \approx 1.146$$

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*KonK1:*
exaktes NV	0.0	nein	nein	ja	ja (5)
	0.5	nein	nein	nein	ja (4)
	1.7	nein	nein	nein	ja (5)
	1.8	ja	nein	nein	ja (5)
	2.3	ja	nein	ja	ja (7)
	2.5	ja	nein	ja	nein[a]

[a] Konvergenz gegen andere Nullstelle

Numerische Tests · Funktion 1

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*KonK1:*
vereinfachtes NV	0.0	ja	–	–	ja (48)
	0.3	nein	–	–	ja (32)
	1.4	nein	–	–	ja (15)
	1.7	ja	–	–	ja (657)
	1.8	ja	–	–	nein
QNRES	0.0	–	nein	–	ja (7)
	0.5	–	nein	–	ja (6)
	1.6	–	nein	–	ja (6)
	1.7	–	nein	–	ja (6)
	3.0	–	nein	–	nein

Numerische Tests · Funktion 2

$$\begin{aligned} f_1(x_1, x_2) &:= (x_1 + 3)(x_2^3 - 7) + 18 \\ f_2(x_1, x_2) &:= \sin(x_2)\exp(x_1) - 1 \end{aligned}$$

$$x^* \approx (0.128, 1.076)$$

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*KonK1:*
exaktes NV	$(0,0)$	nein	nein	nein	ja (6)
	$(0,1)$	nein	nein	nein	ja (4)
	$(0,2.2)$	nein	nein	ja	ja (5)
	$(0,3.2)$	nein	ja	ja	ja (6)
	$(0,3.5)$	ja	ja	ja	nein

Numerische Tests · Funktion 2

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*KonK1:*
vereinfachtes NV	$(0,0)$	ja	–	–	nein
	$(0,0.4)$	ja	–	–	ja (314)
	$(0,1.1)$	nein	–	–	ja (8)
	$(0,2)$	ja	–	–	ja (22)
	$(0,3)$	ja	–	–	ja (134)
	$(0,3.2)$	ja	–	–	nein
QNRES	$(0,0)$	–	ja	–	ja (14)
	$(0,1)$	–	ja	–	ja (7)
	$(0,1.1)$	–	nein	–	ja (6)
	$(0,2)$	–	ja	–	ja (9)
	$(0,3)$	–	ja	–	ja (12)
	$(0,3.2)$	–	ja	–	nein

Numerische Tests · Funktion 3

$$
\begin{aligned}
f_1(x_1,\ldots,x_n) &:= -(3+\alpha x_1)x_1 + 2x_2 - \beta \\
f_i(x_1,\ldots,x_n) &:= x_{i-1} - (3+\alpha x_i)x_i + 2x_{i+1} - \beta \\
f_n(x_1,\ldots,x_n) &:= x_{n-1} - (3+\alpha x_n)x_n - \beta
\end{aligned}
$$

$$\alpha := -0.5,\ \beta := 1 :\ x^* \approx (-1.032, -1.315, \ldots, -0.968, -0.597)$$

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*KonK1:*
exaktes NV	$(0,\ldots,0)$	ja	nein	ja	ja (8)
	$(-0.7,\ldots,-0.7)$	nein	nein	ja	ja (5)
	$(-0.81,\ldots,-0.81)$	nein	nein	nein	ja (4)
	$(-1,\ldots,-1)$	nein	nein	nein	ja (4)
	$(-1.2,\ldots,-1.2)$	nein	nein	ja	ja (4)
	$(-100,\ldots,-100)$	nein	nein	ja	ja (10)

Numerische Tests · Funktion 3

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*KonK1:*
vereinfachtes NV	$(0,\ldots,0)$	ja	–	–	nein
	$(-0.6,\ldots,-0.6)$	ja	–	–	ja (4535)
	$(-0.93,\ldots,-0.93)$	nein	–	–	ja (24)
	$(-1000,\ldots,-1000)$	nein	–	–	ja (14766)
QNRES	$(0,\ldots,0)$	–	ja	–	nein
	$(-0.19,\ldots,-0.19)$	–	ja	–	ja (26)
	$(-0.82,\ldots,-0.82)$	–	nein	–	ja (12)
	$(-3,\ldots,-3)$	–	nein	–	ja (16)
	$(-3.7,\ldots,-3.7)$	–	ja	–	ja (20)
	$(-100,\ldots,-100)$	–	ja	–	ja (72)

Konvexe Optimierung

Residuen-basierte Algorithmen

- **Konvexe Optimierung**

Literatur

Einleitung

Einleitung · Affine Konjugiertheit

- Minimierung eines strikt konvexen, zweimal stetig-differenzierbaren Funktionals

$$f : \mathcal{D} \longrightarrow \mathbb{R}, \quad \mathcal{D} \subseteq \mathbb{R}^n \text{ konvex}$$

- Äquivalent mit dem Lösen der Geichung

$$F(x) := \text{grad}\, f(x) = 0, \quad x \in \mathcal{D}$$

- $G(y) = B^T F(By),\ x = By \implies G'(y) = B^T F'(x) B,\ x = By$
- Die Matrix $G'(y)$ ist konjugiert zur Matrix $F'(x)$, daher spricht man von **affiner Konjugiertheit**

Einleitung · Energienorm

- $F'(x) = f''(x)$ ist symmetrisch und (strikt) positiv definit
- Daher kann man durch

$$\langle \cdot\,,\,\cdot \rangle_{F'(x)} :\ \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}\,, \quad \langle u, v\rangle_{F'(x)} \longmapsto \left\langle u, F'(x)v \right\rangle$$

$$\|\cdot\|_{F'(x)} :\ \mathbb{R}^n \longrightarrow \mathbb{R}\,, \quad \|u\|_{F'(x)} \longmapsto \left\| F'(x)^{1/2} u \right\|$$

ein Energie-Skalarprodukt bzw. eine Energienorm definieren.

Einleitung · Energienorm

- Mit beliebigen $u = B\overline{u},\ v = B\overline{v},\ x = By$ gilt

$$\begin{aligned}\langle u, v\rangle_{F'(x)} &= u^T F'(x) v = \overline{u}^T B^T F'(x) B\overline{v} \\ &= \overline{u}^T G'(y)\overline{v} = \langle \overline{u}, \overline{v}\rangle_{G'(y)}\end{aligned}$$

- Daher Aussagen über
  - die Energienorm der Korrekturen $\left\|F'(x)^{1/2}\Delta x\right\|$,
  - die Energienorm des Fehlers $\left\|F'(x)^{1/2}(x^k - x^*)\right\|$ und
  - die Funktionswerte $f(x^k)$
- Letztere bleiben wegen $g(y^k) := f(By^k) = f(x^k)$ invariant

Gewöhnliches Newton-Verfahren

Gewöhnliches Newton-Verfahren · Newton-Mysovskikh

- Satz (Newton-Mysovskikh, affin konjugierte Version)
  Sei $f : \mathcal{D} \to \mathbb{R}$ ein strikt konvexes, zweimal stetig-differenzierbares Funktional über einer offenen und konvexen Menge $\mathcal{D} \subseteq \mathbb{R}^n$. Sei $F(x) := f'(x)^T$ und $F'(x) = f''(x)$ eine symmetrische und (strikt) positiv definite Matrix und gelten folgende Bedingungen:
  1. Es exitstiert ein $0 \leq \omega < \infty$ so, dass $\forall$ kollineare $x, y, z \in \mathcal{D}$:
  $$\left\| F'(z)^{-1/2} \left(F'(y) - F'(x)\right)(y - x) \right\| \leq \omega \left\| F'(x)^{1/2}(y - x) \right\|^2 .$$
  2. $\mathcal{L}_0 := \left\{ x \in \mathcal{D} \,\middle|\, f(x) \leq f(x^0) \right\}$ ist kompakt.
  3. Für den Startwert x^0 gilt
  $$h_0 := \omega \left\| F'(x^0)^{1/2} \Delta x^0 \right\| < 2 .$$

Gewöhnliches Newton-Verfahren · Newton-Mysovskikh

- Dann gilt:

  1. Die Iterationen des gewöhnlichen Newton-Verfahrens bleiben in $\mathcal{L}_0$ und konvergieren gegen das Minimum x^* von f mit der Rate

  $$\left\| F'(x^{k+1})^{1/2} \Delta x^{k+1} \right\| \leq \frac{\omega}{2} \left\| F'(x^k)^{1/2} \Delta x^k \right\|^2 .$$

  Oder, mit den Definitionen

  $$\varepsilon_k := \left\| F'(x^k)^{1/2} \Delta x^k \right\|^2 , \quad h_k := \omega \left\| F'(x^k)^{1/2} \Delta x^k \right\| ,$$

  mit den Abschätzungen

  $$-\frac{1}{6} h_k \varepsilon_k \leq f(x^k) - f(x^{k+1}) - \frac{1}{2}\varepsilon_k \leq \frac{1}{6} h_k \varepsilon_k ,$$

  $$\frac{1}{6}\varepsilon_k \leq f(x^k) - f(x^{k+1}) \leq \frac{5}{6}\varepsilon_k .$$

Gewöhnliches Newton-Verfahren · Newton-Mysovskikh

- 2. Der Abstand zum Minimum ist beschränkt durch

$$f(x^0) - f(x^k) \leq \frac{\frac{5}{6}\varepsilon_0}{1 - \frac{h_0}{2}}.$$

- Mit den Kontraktionsfaktoren

$$\Theta_k := \sqrt{\frac{\varepsilon_{k+1}}{\varepsilon_k}} = \frac{\left\| F'(x^{k+1})^{1/2} \Delta x^{k+1} \right\|}{\left\| F'(x^k)^{1/2} \Delta x^k \right\|}$$

folgert man wieder

$$\Theta_k = \frac{h_{k+1}}{h_k} \leq \frac{1}{2} h_k < 1 \text{ und } \Theta_{k+1} \approx \Theta_k^2 < \Theta_0 < 1$$

Gewöhnliches Newton-Verfahren · Kriterien

- **Startwert-Kriterium:** Falls

$$\Theta_0 \geq 1$$

gilt, weisen wir den Startwert x^0 als *nicht hinreichend nahe* zurück.

- **Divergenz-Kriterium 1:** Wir brechen die gewöhnliche Newton-Iteration als *divergent* ab, falls

$$f(x^{k+1}) - f(x^k) > -\frac{1}{6}\varepsilon_k\,.$$

Gewöhnliches Newton-Verfahren · Kriterien

- **Divergenz-Kriterium 2:** Wir brechen die gewöhnliche Newton-Iteration als *divergent* ab, falls

$$\Theta_k \geq \Theta_0 \text{ für } k > 0\,.$$

- **Divergenz-Kriterium 3:** Wir brechen die gewöhnliche Newton-Iteration als *divergent* ab, falls

$$\Theta_{k+1} \geq \frac{\Theta_k^2}{\Theta_0}\,.$$

Gewöhnliches Newton-Verfahren · Kriterien

- Aus $f(x^k) - f(x^{k+1}) \longrightarrow \frac{1}{2}\varepsilon_k$ folgern wir:
- **Konvergenz-Kriterium 1:** Wir beenden die gewöhnliche Newton-Iteration, falls
$$f(x^k) - f(x^{k+1}) \leq \frac{1}{2} ETOL^2$$
erfüllt ist und geben x^{k+1} als Lösung aus.
- **Konvergenz-Kriterium 2:** Wir beenden die gewöhnliche Newton-Iteration, falls
$$\varepsilon_k \leq ETOL^2$$
erfüllt ist und geben x^k als Lösung aus.

Vereinfachtes Newton-Verfahren

Vereinfachtes Newton-Verfahren · Satz

- Sei $f : \mathcal{D} \to \mathbb{R}$ ein strikt konvexes, zweimal stetig-differenzierbares Funktional über einer konvexen Menge $\mathcal{D} \subseteq \mathbb{R}^n$. Sei $F(x) := f'(x)^T$ und $F'(x) = f''(x)$ eine symmetrische und (strikt) positiv definite Matrix und gelten folgende Bedingungen:
  1. Es exitstiert ein $0 \leq \omega < \infty$ so, dass $\forall\, z \in \mathcal{D}$, $v \in \mathbb{R}^n$
  $$\left\| F'(x^0)^{-1/2} \left(F'(z) - F'(x^0) \right) v \right\| \leq \omega \left\| F'(x^0)^{1/2}(z - x^0) \right\| \left\| F'(x^0)^{1/2} v \right\| .$$
  2. Für den Startwert x^0 gilt
  $$h_0 := \omega \left\| F'(x^0)^{1/2} \overline{\Delta x}^0 \right\| \leq \frac{1}{2} .$$
  Sei ferner $t^* := 1 - \sqrt{1 - 2h_0}$.

Vereinfachtes Newton-Verfahren · Satz

- Dann gilt:

  1. Die Iterationen des vereinfachten Newton-Verfahrens konvergieren gegen ein x^* mit
  $$\omega \left\| x^* - x^0 \right\| \leq t^* .$$
  2. Die Konvergenzrate kann durch
  $$-\frac{1}{6}\varepsilon_k(t_{k+1} + 2t_k) \leq f(x^k) - f(x^{k+1}) - \frac{1}{2}\varepsilon_k \leq \frac{1}{6}\varepsilon_k(t_{k+1} + 2t_k) ,$$
  $$\Theta_k = \left(\frac{\varepsilon_{k+1}}{\varepsilon_k} \right)^{1/2} \leq \frac{1}{2}(t_{k+1} + t_k)$$
  abgeschätzt werden, wobei
  $$t_0 := 0 , \quad t_{k+1} = h_0 + \frac{1}{2} t_k^2 < t^* .$$

Vereinfachtes Newton-Verfahren · Kriterien

- **Startwert-Kriterium:** Falls

$$\Theta_0 > \frac{1}{4}$$

gilt, weisen wir den Startwert x^0 als *nicht hinreichend nahe* zurück.

- **Konvergenz-Kriterium:** Wir beenden die vereinfachte Newton-Iteration, falls

$$\varepsilon_{k-1} = \left\| F'(x^0)^{1/2} \overline{\Delta x}^{k-1} \right\|^2 \leq ETOL^2$$

erfüllt ist und geben x^k als Lösung aus.

Vereinfachtes Newton-Verfahren · Kriterien

- **Divergenz-Kriterium 1:** Wir brechen die vereinfachte Newton-Iteration als *divergent* ab, falls

$$\Theta_k \geq 1 \,.$$

- **Divergenz-Kriterium 2:** Wir brechen die vereinfachte Newton-Iteration als *divergent* ab, falls

$$\Theta_k \geq [t^*] := 1 - \sqrt{1 - 4\Theta_0} \,.$$

Inexaktes Newton-PCG-Verfahren

Inexaktes Newton-PCG-Verfahren · Grundlagen

- Jacobi-Matrix der Gradientenabbildung ist symmetrisch und positiv definit
- Daher bietet sich zur Lösung ein (vorkonditioniertes) konjugiertes Gradientenverfahren ($\hat{=}$ PCG) an
- Weitere Details im Handout

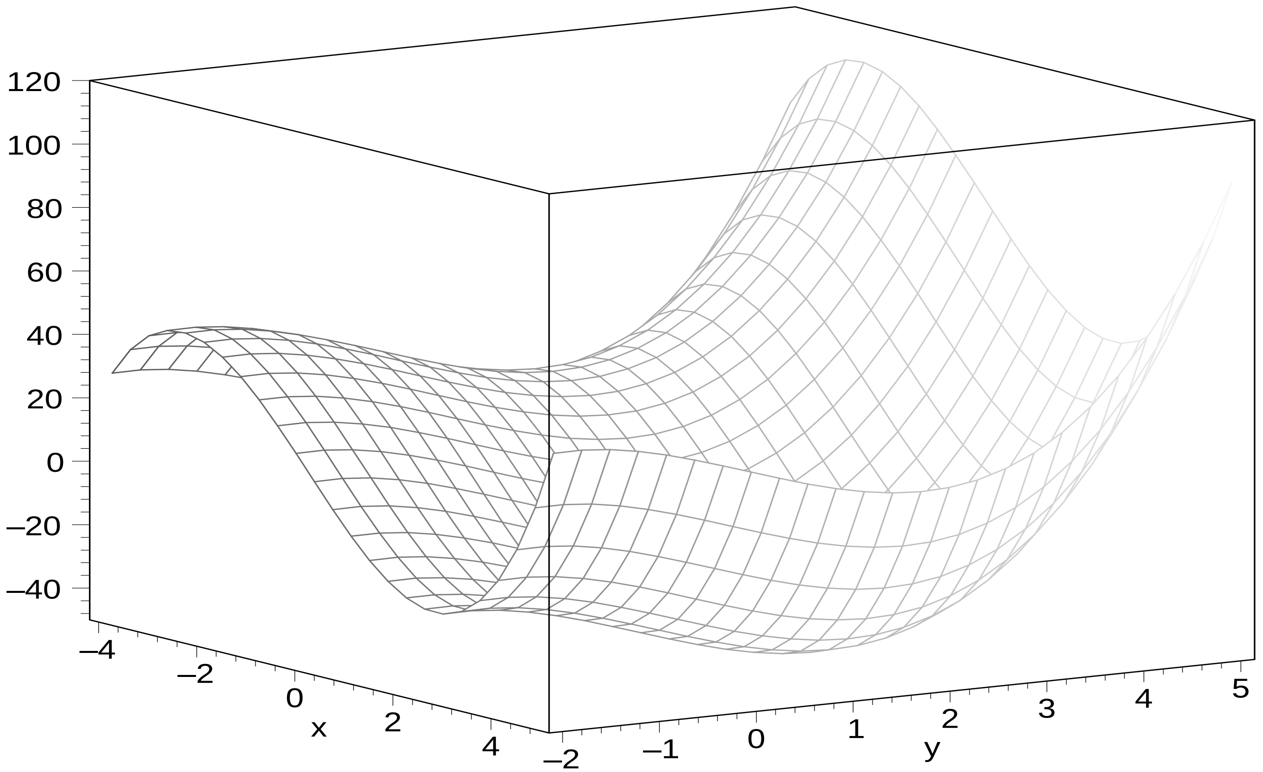
Numerische Tests

Numerische Tests · Funktion 1

$$f(x_1, x_2) := x_1^2 + x_1 + x_2^2$$
$$x^* = (-0.5, 0)$$

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*DivK3:*	*KonK1:*	*KonK2:*
exaktes NV	$(0, 0)$	nein	nein	nein	nein	ja (2)	ja (1)
	$(10, -10)$	nein	nein	nein	nein	ja (3)	ja (2)
	$(-10, 100)$	nein	nein	nein	nein	ja (3)	ja (2)
	$(1000, 10)$	nein	nein	nein	nein	ja (5)	ja (4)
vereinf. NV	$(0, 0)$	nein	nein	nein	–	ja (1)	–
	$(10, -10)$	nein	nein	nein	–	ja (2)	–
	$(-10, 100)$	nein	nein	ja	–	ja (3)	–
	$(100, 10)$	nein	nein	ja	–	ja (4)	–
	$(700, 10)$	ja	nein	nein	–	ja (64)	–

Numerische Tests · Funktion 2



$$f(x_1, x_2) = x_1^3 - 20x_1 + x_2^3 - 10x_2$$

im Bereich des lokalen Minimums $x^* \approx (2.582, 1.826)$.

Numerische Tests · Funktion 2

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*DivK3:*	*KonK1:*	*KonK2:*
exaktes NV	$(2,2)$	nein	nein	nein	nein	ja (3)	ja (3)
	$(3,2)$	nein	nein	nein	nein	ja (4)	ja (3)
	$(10,2)$	nein	nein	nein	nein	ja (6)	ja (5)
	$(2,1)$	ja	nein	nein	ja	ja (4)	ja (4)
	$(2,0.1)$	ja	nein	nein	ja	ja (6)	ja (8)
vereinf. NV	$(2,2)$	nein	nein	nein	–	ja (10)	–
	$(3,2)$	nein	nein	ja	–	ja (6)	–
	$(10,2)$	nein	nein	ja	–	ja (44)	–
	$(2,1)$	ja	nein	nein	–	ja (64)	–
	$(2,0.1)$	ja	ja	nein	–	nein	–

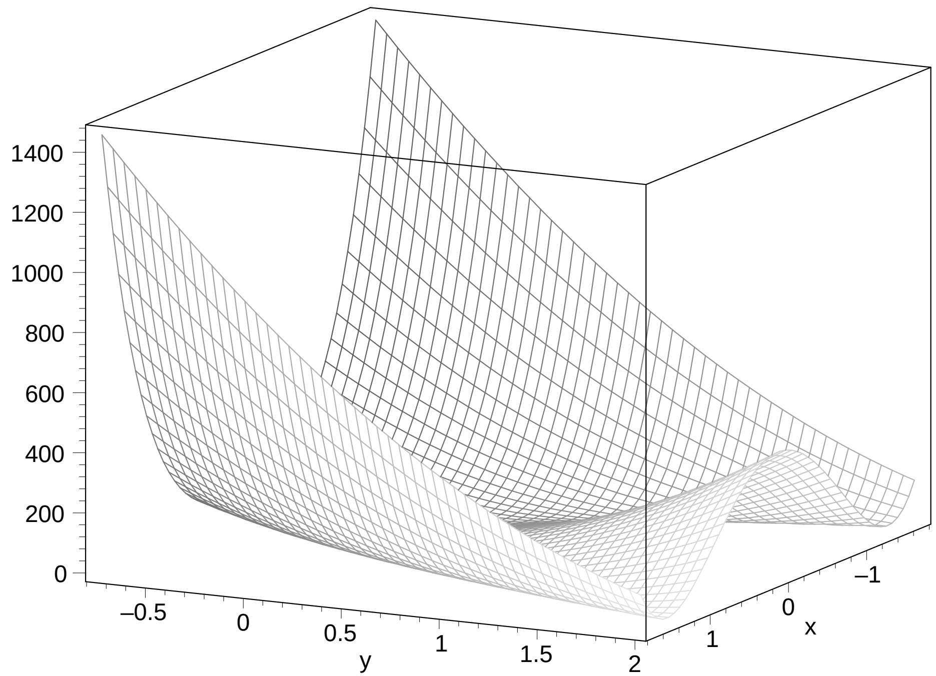
Numerische Tests · Funktion 3

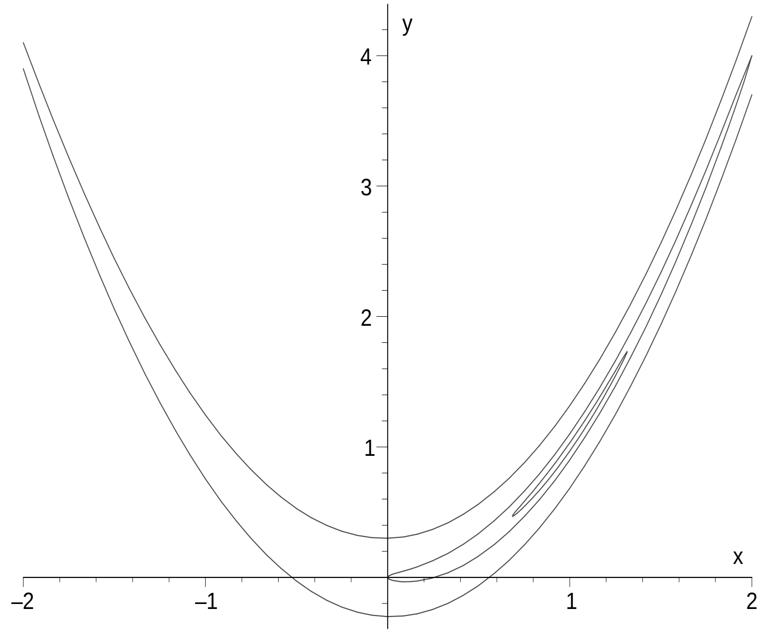
$$
\begin{aligned}
f(x_1,x_2,x_3,x_4) &:= (x_1+10x_2)^2+5(x_3-x_4)^2 \\
&\quad +(x_2-2x_3)^4+10(x_1-x_4)^4 \\
x^* &= (0,0,0,0)
\end{aligned}
$$

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*DivK3:*	*KonK1:*	*KonK2:*
exaktes NV	$(3,-1,0,1)$	nein	ja	ja	ja	ja (23)	ja (21)
	$(1,-1,0,1)$	ja	ja	nein	ja	ja (26)	ja (24)
	$(0,0,0,1)$	ja	ja	nein	ja	ja (32)	ja (31)
vereinf. NV	$(3,-1,0,1)$	nein	nein	ja	–	ja (19340)	–
	$(1,-1,0,1)$	ja	ja	nein	–	nein	–
	$(0,0,0,1)$	ja	nein	nein	–	nein	–

Numerische Tests · Funktion 4

Rosenbrocks Funktion mit parabelförmigem Tal (siehe [12] und [7]).





Links: Dreidimensionale Darstellung.

Rechts: Niveaulinien für die Werte 10, 1 und $\frac{1}{10}$ (von außen nach innen).

Numerische Tests · Funktion 4

$$
\begin{aligned}
f(x_1, x_2) &:= 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\
x^* &= (1, 1)
\end{aligned}
$$

Verfahren:	*Startwert:*	*SWK:*	*DivK1:*	*DivK2:*	*DivK3:*	*KonK1:*	*KonK2:*
exaktes NV	$(-1.2, 1)$	nein	ja	ja	ja	ja (9)	ja (7)
	$(0, 0)$	ja	nein	nein	ja	ja (3)	ja (3)
	$(0.8, 0.8)$	nein	ja	ja	ja	ja (7)	ja (5)
	$(2, 2)$	nein	ja	ja	ja	ja (7)	ja (6)
vereinf. NV	$(-1.2, 1)$	nein	ja	ja	–	nein	–
	$(0, 0)$	ja	ja	nein	–	nein	–
	$(0.8, 0.8)$	nein	ja	nein	–	nein	–
	$(2, 2)$	nein	nein	ja	–	nein	–

Literatur

Residuen-basierte Algorithmen

Konvexe Optimierung

- **Literatur**

Teil 1

[1] Broyden, C. G.: *A Class of Methods for Solving Nonlinear Simultaneous Equations*. Mathematics of Computation, Vol. 19, No. 92, 577-583 (1965)

[2] Dembo, R. S.; Eisenstat, S. C.; Steihaug, T.: *Inexact Newton Methods*. SIAM Journal on Numerical Analysis, Vol. 19, No. 2, 400-408 (1982)

[3] Deuflhard, P.: *Newton Methods for Nonlinear Problems*. Springer, New York (2004).

[4] Deuflhard, P.: *Cascadic Conjugate Gradient Methods for Elliptic Partial Differential Equations I. Algorithm and Numerical Results*. Konrad-Zuse-Zentrum für Informationstechnik Berlin, Preprint SC 93-23 (1993)

[5] Deuflhard, P.; Weiser, M.: *Local Inexact Newton Multilevel FEM for Nonlinear Elliptic Problems*. Konrad-Zuse-Zentrum für Informationstechnik Berlin, Preprint SC 96-29 (1996)

[6] Fletcher, R.; Powell, M. J. D.: *A rapidly convergent descent method for minimization*. The Computer Journal, Vol. 6, Issue 2, 163-168 (1963)

Teil 2

[7] FLETCHER, R.; REEVES, C. M.: *Function minimization by conjugate gradients*. The Computer Journal, Vol. 7, Issue 2, 149-154 (1964)

[8] HOHMANN, A.: *Inexact Gauss Newton Methods for Parameter Dependent Nonlinear Problems*. Dissertation, Freie Universität Berlin (1994)

[9] KIRCHES, C.: *Fehlergesteuerte Algorithmen*. Seminarvortrag, Universität Heidelberg (2004).

[10] NOCEDAL, J.; WRIGHT, S. J.: *Numerical Optimization*. Springer, New York (1999).

[11] PONTA, O.: *P. Deuflhard, Newton Methods for Nonlinear Problems, Kapitel 1*. Seminarvortrag, Universität Heidelberg (2004).

[12] ROSENBROCK, H. H.: *An automatic method for finding the greatest or the least value of a function*. The Computer Journal, Vol. 3, Issue 3, 175-184 (1960)

Hinweis

- Das Handout, die Präsentationsfolien und die Quelltexte stehen unter

  `www.mathi.uni-heidelberg.de/~ferreau/newton/`

  zum Download bereit.

Fragen

Fragen?