Globale Newton Verfahren

- Betrachten: System von n nichtlinearen Gleichungen:

$$F : D \rightarrow \mathbb{R}^n,\ F \in C^1(D)$$

- Gesucht: $\hat{x} \in D$, sodass $F(\hat{x}) = 0$.

- Vorher: Bedingungen für Startwert x^0 wie z.B.

$$\|\hat{x} - x^0\| < 2/\omega$$

garantieren die Konvergenz des lokalen Newton-Verfahrens zu $\hat{x}$

Globale Newton Verfahren

- Jetzt: “**Globalisierung**” der lokalen Verfahren, z.B:
  - steepest descent methods
  - trust region methods (Levenberg-Marquardt-Modell)
  - Newton method with damping strategie

- Ziele:
  - Konvergenz auch für schlechte Anfangsvektoren x^0
  - Annäherung an lokales Verfahren in der Nähe von $\hat{x}$, um dessen quadratische bzw. superlineare Konvergenz auszunutzen

Globale Newton Verfahren
Einfache Globalisierungsmöglichkeiten

- Homotopie-Verfahren für Probleme der Form $F(x)=0$:

$$\hat{F}(x,\tau)=0,\ \tau\in[0,1]$$

- Für Probleme der Form $x'=F(x),\ x(0)=x^0$:
  - Explizites Euler-Verfahren:

$$x^{k+1}-x^k=\Delta x^k=\tau F(x^k)$$

  - Lineares implizites Euler-Verfahren:

$$(Id-\tau F'(x^k))\,\Delta x^k=\tau F(x^k)$$

Globale Newton Verfahren komponentenweise konvexe Funktionen

Komponentenweise konvexe Funktionen können durch eine der drei äquivalenten Eigenschaften charakterisiert werden:

(1) $$F(\lambda x+(1-\lambda)y)\leq \lambda F(x)+(1-\lambda)F(y)$$

(2) $$F(y)-F(x)\geq F'(x)(y-x)$$

(3) $$(F'(y)-F'(x))(y-x)\geq 0$$

Für $x, y\in D\cap\mathbb{R}^n,\ D\ konvex,\ \lambda\in[0,1]$. Um monotone Konvergenz sicherzustelle, nehmen wir zusätzlich an:

$$F'(z)^{-1}\geq 0,\ z\in D$$

Globale Newton Verfahren
komponentenweise konvexe Funktionen

(4) $$F'(x)^{-1}F(\lambda x+(1-\lambda)y)\leq F'(x)^{-1}(\lambda F(x)+(1-\lambda)F(y))$$

(5) $$F'(x)^{-1}(F(y)-F(x))\geq y-x$$

(6) $$F'(x)^{-1}(F'(y)-F'(x))(y-x)\geq 0$$

Lemma 3.1 Sei $F:D\rightarrow\mathbb{R}^n$ eine stetig differenzierbare Funktion mit $D\subseteq\mathbb{R}^n$ offen und konvex. Wenn die Funktion eine der Eigenschaften (4)-(6) erfüllt, so konvergiert das gewöhnliche Newton Verfahren mit Startpunkt $x^0\in D$ monoton und global, sodass komponentenweise gilt:

$$\hat{x}\leq x^{k+1}\leq x^k,\ \ k=1,2,\ldots$$

Globale Newton Verfahren
'residual level function'

- *Residual level function:*

$$T(x) := 1/2 \|F(x)\|_2^2 \equiv 1/2\, F(x)^T F(x)$$

- *Eigenschaften:*

$$T(x) = 0 \Leftrightarrow x = \hat{x}$$

$$T(x) > 0 \Leftrightarrow x \neq \hat{x}$$

- *Monotonie Kriterium:*

$$T(x^k) \neq 0 \Rightarrow T(x^{k+1}) < T(x^k)$$

$$(lokale\, Verfahren: \; x^k \neq \hat{x} \Rightarrow \|x^{k+1} - \hat{x}\| < \|x^k - \hat{x}\|)$$

Globale Newton Verfahren 'level sets'

- *Level sets:*

$$G(z) := \{ x \in D \subseteq \mathbb{R}^n \mid T(x) \leq T(z) \}$$

- *Eigenschaften von T implizieren:*

$$\hat{x} \in G(x) \, , \; x \in D$$

- *Monotonie Kriterium*:

$$\mathring{G}(x^k) \neq \emptyset \Rightarrow x^{k+1} \in \mathring{G}(x^k)$$

Globale Newton Verfahren
'steepest descent method'

$$\Delta x^k := -gradT(x^k) = -F'(x^k)^T F(x^k)$$

$$x^{k+1} := x^k + s_k \Delta x^k, \quad s_k > 0. \ Schrittlängenparameter$$

Lemma 3.2 Sei $F: D \rightarrow \mathbb{R}^n$ eine stetig differenzierbare Funktion mit $D \subseteq R^n$. Sei $\Delta x \neq 0$. Dann gibt es ein $\mu > 0$, sodass

$$T(x + s\Delta x) < T(x), \ 0 < s < \mu$$

(Der Index 'k' wurde in der Notation weggelassen.)

Globale Newton Verfahren 'steepest descent method' - Schrittlängestrategie

- Reduktionsstrategie:

  Wenn $T(x^k + s_k^i \Delta x^k) > T(x^k),\ i \in \{0,1,\ldots\},$ dann setze $s_k^{i+1} := \kappa\, s_k^i,\ \kappa < 1$. Lemma 3.2 sichert die Existenz von $\hat{i}$, sodass $s_k^{\hat{i}} > 0$ ein geeigneter Schrittlängenfaktor ist.

- Prediktionsstrategie

$$s_{k+1}^0 := \begin{Bmatrix} min(s_{max}, s_k^{\hat{i}}/\kappa), & wenn\ s_{k-1}^{\hat{i}} \le s_k^{\hat{i}} \\ s_k^{\hat{i}} & sonst \end{Bmatrix}$$

Globale Newton Verfahren Levenberg-Marquardt-Modell

- Definiere den Korrekturvektor Δx durch

(7)
$$min_{\Delta x} \| F(x) + F'(x) \Delta x \|_2$$
$$s.t.\ \| \Delta x \|_2 \leq \delta,\ \delta > 0$$

- Sei $p \geq 0$ der *Lagrange Multiplikator,* so erhält man das nebenbedingungsfreie Optimierungsproblem

$$min_{\Delta x} \| F(x) + F'(x) \Delta x \|_2^2 + p \| \Delta x \|_2^2$$

- Dies führt auf das *Levenberg-Marquardt-Modell:*

(8) $$(F'(x^k)^T F'(x^k) + p\, Id) \Delta x = -F'(x^k)^T F(x^k),\ x^{k+1} = x^k + \Delta x^k$$

Globale Newton Verfahren
'Trust region strategies'

- Grenzfälle für den Korrekturvektor $\Delta x^k(p)$

  $$p \to 0^+ : \Delta x^k(0) = -F'(x^k)^{-1} F(x^k), \; wenn\, F'(x^k) invertierbar$$

  $$p \to \infty : \Delta x^k(p) \to -1/p\, grad\, T(x^k)$$

- **Lemma 3.4** Unter den üblichen Voraussetzungen sei $\Delta x(p) \neq 0$ die Levenberg-Marquardt-Korrektur definiert in (8). Dann gibt es ein $p_{min} \geq 0$, sodass

  $$T(x + \Delta x(p)) < T(x), \; p > p_{min}$$

Globale Newton Verfahren
'Trust region strategies'

- Affin kovariantes Variante des L.-M.-Modells

$$min \left\| F'(x)^{-1} (F(x) + F'(x) \Delta x) \right\|_2$$

$$s.t. \left\| \Delta x \right\|_2 \le \delta_0$$

- Daraus erhält man die Newton Methode mit Dämpfungsfaktor

$$F'(x^k) \Delta x^k = -F(x^k), \; x^{k+1} := x^k + \lambda_k \Delta x^k, \; 0 < \lambda_k \le 1$$

- Affin kontravariante Variante des L.-M.-Modells

$$min \left\| F(x) + F'(x) \Delta x \right\|_2$$

$$s.t. \left\| F'(x) \Delta x \right\|_2 \le \delta_0$$

Globale Newton Verfahren 'Trust region strategies'

- **Lemma 3.5** Unter den üblichen Voraussetzungen sei $F'(x)$ invertierbar und $F \neq 0$. Δx sei die Newton Richtung. Dann gibt es ein $\mu > 0$, sodass

$$T(x + \lambda \Delta x) < T(x),\ 0 < \lambda < \mu$$

- **Armijo strategie** Sei $\Lambda \subset \{1, 1/2, 1/4, \ldots, \lambda_{min}\}$ ein Sequenz sodass gilt:

$$T(x^k + \lambda_k \Delta x^k) \leq (1 - 1/2\, \lambda_k) T(x^k),\ \lambda \in \Lambda_k$$

und definiere den optimalen Dämpfungsfaktor durch:

$$T(x^k + \lambda_k \Delta x^k) = min_{\lambda \in \Lambda_k} T(x^k + \lambda \Delta x^k)$$

Globale Newton Verfahren
Newton Pfad

- Erinnerung: $T(x|A) = 1/2 \| A F(x) \|_2^2 \quad A \in Gl(n)$
- Definition: $G(z|A) := \{ x \in D \subseteq \mathbb{R}^n \mid T(x|A) \leq T(z|A) \}$
- Monotonie-Kriterium: $\mathring{G}(x^k|A) \neq \emptyset \;\Rightarrow\; x^{k+1} \in \mathring{G}(x^k|A)$
- Definition: $\bar{G}(x) := \cap_{A \in Gl(n)} G(x|A)$

Globale Newton Verfahren
Newton Pfad

Theorem 3.6 Sei $F \in C^1(D),\ D \subseteq \mathbb{R}^n,\ F'(x)$ invertierbar für alle $x \in D$. Für $\hat{A} \in Gl(n)$ sei die pfadzusammenhängende Komponente von $G(x^0|\hat{A})$ in x^0 kompakt und in *D* enthalten. Dann ist die pfadzusammenhängende Komponente von $\bar{G}(x^0)$ ein topologischer Pfad $\bar{x}:[0,2] \to \mathbb{R}^n$, der so genannte *Newton-Pfad,* der die folgenden Bedingungen erfüllt:

$$F(\bar{x}(\lambda)) = (1-\lambda) F(x^0)$$

$$T(\bar{x}(\lambda)|A) = (1-\lambda)^2 T(x^0|A)$$

$$d\bar{x}/d\lambda = -F'(\bar{x})^{-1} F(x^0)$$

$$\bar{x}(0) = x^{0,}\ \bar{x}(1) = \hat{x}$$

$$d\bar{x}/d\lambda|_{\lambda=0}. = -F'(x^0)^{-1} F(x^0) \equiv \Delta x^0$$

Globale Newton Verfahren
Newton Pfad

- Der Newton-Pfad $\bar{x}$ hat die Eigenschaft das alle Funktion $T(x|A)$ für alle $A \in Gl(n)$ entlang $\bar{x}$ fallen.

- In der Praxis terminieren alle Newton-Pfad $\bar{x}$ (auch für $F'(\mathring{x})$ singulär), die in x^0 starten, in einem der 3 Fälle:

  - Im Punkt $\hat{x}$
  - An einem Punkt $\mathring{x}$ mit singulärer Jakobi-Matrix, der nahe genug an $\hat{x}$ liegt
  - An einem Punkt auf dem Rand δD des Definitions-bereiches von F