



Adaptive Systeme

Optimierung

Etwas Mathematik zu Optimierungsaufgaben



Optimierungsverfahren

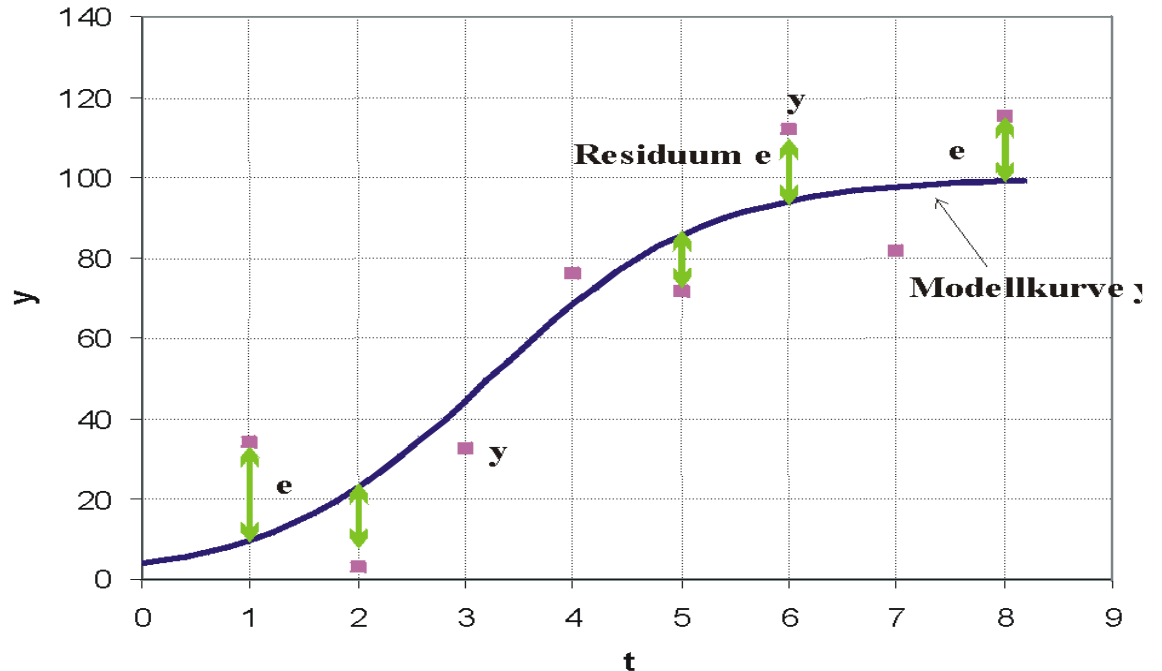
- Bei der einfachen Optimierung geht es darum die Minima oder Maxima einer geeigneten Ziel- oder Kostenfunktion zu finden.
- Eng verwandt ist die Aufgabe eine Funktion zu approximieren, dann wird meist als Zielfunktion eine geeignete (Vektor)Norm optimiert.
- Ein Optimum wird wie in der Kurvendiskussion in einer Dimension durch die Nullstelle der ersten Ableitung ermittelt.
- Geeignete Methoden zur Nullstellensuche sind daher wichtige Hilfsmittel der Optimierung.



Problemstellungen

- Gesucht sind meistens *optimale Lösungen*.
 - Was heißt hier optimal...?
- Gelingt es ein Problem mathematisch zu formulieren, so ist die optimale Lösung als das Minimum einer Kostenfunktion beschrieben.
- Häufig reicht als Lösung eine Approximation aus, da
 - die Eingangsdaten nicht exakt vorliegen
 - nur ein vereinfachtes Modell vorliegt
 - das Modell mathematisch nicht geschlossen lösbar ist.
- Die Lösung muss dann innerhalb einer vorgegebenen Fehlerschranke $\varepsilon > 0$ exakt sein.

Daten fitten



- Häufig muss zu gegebenen Daten eine geeignete Modellfunktion gefunden und parametrisiert werden.
- Hierzu wird versucht die Fehler (Residuen) zwischen den Daten und der Modellfunktion zu minimieren.



Es kommt drauf an...

- Es gibt verschiedene Möglichkeiten eine Kurve an die Datenpunkte anzupassen, z. B.:
- Polynom-Fit: Suche ein Polynom, das exakt(!) durch alle Datenpunkte geht.

$$\sum_k |p_n(x_k) - y_k| = 0$$

- => Lagrange- oder Newton-Polynome.
- Führt häufig zu heftig oszillierenden Lösungen...
- Suche eine Funktion, die „glatt“ durch die Datenpunkte geht, d.h. dort eine stetige Ableitung besitzt.
 - => Kubische Spline-Funktion.
 - stückweise Polynome dritten Grads.



Beispiel: Messreihen fitten

- Gegeben sei eine Messreihe $(x_k, y_k)_{k \in \mathbb{N}}$ an die eine Modellfunktion $f_a(x)$ optimal angepasst werden soll.
 - $(x_k, y_k) \in \mathbb{R}^p \times \mathbb{R}^q$ können höher dimensional sein.
- Meist wird hierzu der quadratische Fehler minimiert:

$$\chi^2 = \sum_{k=1}^m (y_k - f(x_k; a))^2 = \text{Min}$$

- Die Funktion f_a wird mittels p Variablen a_k bzw. einem Vektor $\vec{a} \in \mathbb{R}^p$ parametrisiert, der geeignet zu bestimmen ist. (Das Vektorsymbol fehlt häufig...)
- $\chi^2 : \mathbb{R}^p \rightarrow \mathbb{R}$ ist eine reellwertige Funktion mit $\chi^2 \geq 0$ und gesucht ist ein a mit $\chi^2(a) = \text{Min}$



Vektorschreibweise

- Schreibweise für Vektoren a, b, c des \mathbb{R}^n für $n=3$.

$$\vec{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \vec{c} = \vec{a} + \vec{b} = \begin{pmatrix} x+1 \\ y+2 \\ z+3 \end{pmatrix}$$

- Wichtig ist es den Unterschied zum transponierten Vektor zu beachten.

$$\vec{a}^T = (1, 2, 3) \neq \vec{a}$$

- Entsprechendes gilt für vektorielle Funktionen z.B.:

$$\vec{f}(t)^T = (t, \sin(t), e^{-t^2})$$



Methode der kleinsten Quadrate

- Ziel der Ausgleichsrechnung ist es die unbekanntes p Parameter $a=(a_1, \dots, a_p)$ einer Modellfunktion f an eine Reihe von (Mess)Daten $(x, y)_k$ *optimal* anzupassen.
- Dies geschieht mit der Methode der kleinsten Quadrate:

$$\chi^2(\vec{a}) = \sum_k (f(\vec{a}, x_k) - y_k)^2 = \text{minimal}$$

- Ableiten führt auf p Bedingungsgleichungen $g_j(a)$

$$g_j(\vec{a}) = (\nabla_{\vec{a}} \chi^2)_j \equiv \frac{\partial \chi^2}{\partial a_j} = 0 \quad \text{mit } j = 1, \dots, p$$



Beispiel „Schwingende Saite“

- Die Differentialgleichung einer Saite eingespannt in den Punkten $x=0$ und $x = \pi$ lautet:

$$\frac{\partial^2 y}{\partial t^2} = \omega^2 \frac{\partial^2 y}{\partial x^2} \quad \Rightarrow \quad \begin{aligned} f'' + \lambda f &= 0 \\ \ddot{g} + \lambda \omega^2 g &= 0 \end{aligned}$$

- Wg. $f(0) = f(\pi) = 0 \Rightarrow \lambda_n = n^2$ ($n \in \mathbb{N}$) lautet die Lösung der Einspannbedingung

$$\bar{y}(x, t) = \sum_{n=0}^{\infty} \sin(nx) (a_n \cos(n\omega t) + b_n \sin(n\omega t))$$

- Es gilt auch noch die Anfangsbedingungen

$$y(x, 0) = g(x) \quad \partial_t y(x, 0) = v(x)$$

- mit geeigneten a_n und b_n zu erfüllen.



Generische Optimierung

- Bei vielen Problemen geht es darum (lokale oder globale) Optima zu finden. D. h. es gibt eine Abbildung:

$$f : U \rightarrow V$$

$$x \rightarrow y = f(x)$$
- und gesucht ist ein Wert $x_0 \in U$ der die Abbildung f bezüglich der Zielfunktion – meist versehen mit einer Norm $\| \cdot \|$ – optimiert.
- Bei dem Messreihenbeispiel wurde das Minimum der Abstandsquadrate verwendet und die schwingende Saite führt auf die Konvergenz im quadratischen Mittel und zur Theorie der Fourierreihen.



Normierte Räume

- Ein linearer Raum V heißt normierter Raum, wenn jedem $f \in V$ eine reelle Zahl $\|f\|$ zugeordnet ist mit den Eigenschaften:

$$\|f\| \geq 0 \quad \text{und} \quad \|f\| = 0 \Leftrightarrow f = 0 \quad \text{ist Nullvektor}$$

$$\|\alpha f\| = |\alpha| \|f\| \quad (\alpha \in \mathbb{C})$$

$$\|f + g\| \leq \|f\| + \|g\|$$

- Bemerkung: Der Raum V muss nicht immer der \mathbb{R}^n bzw. \mathbb{C}^n sein, sondern auch Funktionsräume, wie die Menge der stetigen (p-mal differenzierbaren) Funktionen bilden „Räume“ : $C(X), C^p(X)$



Beispiele von Normen

- Bekannte Normen auf \mathbb{R}^n sind die „Manhattan Norm“

$$\|x\|_1 := \sum_{k=1}^n |x_k|$$

- die euklidische Norm (per Skalarprodukt)

$$\|x\|_2 := \sqrt{\langle x, x \rangle} = \sqrt{x^T \cdot x} = \sqrt{\sum_{k=1}^n x_k^2}$$

- und die Maximumsnorm

$$\|x\|_\infty := \max(|x_1|, \dots, |x_n|)$$

- allesamt Spezialfälle der generischen p-Norm

$$\|x\|_p := \left(\sum_{k=1}^n |x_k|^p \right)^{1/p}$$



Normen für Funktionsräume

- Die Menge $B(T)$ der beschränkten Funktionen f auf einer Menge T

$$\|f\| := \sup_{t \in T} |f(t)|$$

- Die Menge $C(T)$ der stetigen Funktionen f auf kompakten T

$$\|f\| := \max_{t \in T} |f(t)|$$

- Die Menge $L^p(a, b)$ der auf dem Interval $[a, b]$ meßbaren (integrierbaren) Funktionen

$$\|f\|_p := \left(\int_a^b |f(t)|^p dt \right)^{(1/p)}$$



Norm und Metrik

- Einem normierten Raum ist ein topologischer Abstands-begriff mittels der kanonischen Metrik $d(x,y)$, zugeordnet:

$$d(x, y) := \|x - y\|$$
- Die bis lang betrachteten Normen induzieren für diese Metrik, dass in ihnen alle Cauchy-Folgen gegen ihren Grenzwert konvergieren und Addition, Multiplikation und die Norm selber stetig sind:
 - Aus: $\vec{x}_n \rightarrow \vec{x}$, $\vec{y}_n \rightarrow \vec{y}$, $\alpha_n \rightarrow \alpha$
 - Folgt: $\vec{x}_n + \vec{y}_n \rightarrow \vec{x} + \vec{y}$, $\alpha_n \vec{x}_n \rightarrow \alpha \vec{x}$, $\|\vec{x}_n\| \rightarrow \|\vec{x}\|$
 - Wichtige Voraussetzung für unsere Approximationen.



Norm und Innenprodukt

- Ist auf dem Raum V eine Abbildung $\langle x|y \rangle$ definiert (auch $\langle x,y \rangle$ oder $x^T y$) mit den Eigenschaften:

$$\langle x+y|z \rangle = \langle x|z \rangle + \langle y|z \rangle$$

$$\langle \alpha x|y \rangle = \alpha \langle x|y \rangle$$

$$\langle x|y \rangle = \langle y|x \rangle^*$$

$$\langle x|x \rangle \geq 0 \text{ und } \langle x|x \rangle = 0 \text{ genau wenn } x=0$$

- So induziert dieses Innenprodukt eine Norm:

$$\|x\| := \langle x|x \rangle^{1/2}$$

- Z.B für Vektoren auf \mathbb{R}^n oder „Vektoren“ des $L^p(a,b)$

$$\vec{x}^T \cdot \vec{y} = \sum_{k=1}^n x_k y_k \quad \langle x|y \rangle = \int dt x(t) \cdot y(t)$$

Bracket Schreibweise $\langle | \rangle$
Mit Bra- und Ket-Vektoren
 $\langle |$ und $| \rangle$.

Berechnen Sie für die beiden (Spalten)Vektoren

$$\vec{a} = (x, 4 - x, x^2)^T$$

$$\vec{b} = (4x, x^2, x + 2)^T$$

1. Das Produkt $\vec{a}^T \vec{b}$
2. sowie das Produkt $\vec{a} \vec{b}^T$
3. Differenzieren Sie 1. und 2. nach x .
4. Verifizieren Sie die „Kettenregel“

$$\frac{d}{dx} (\vec{a}^T \vec{b}) = \frac{d \vec{a}^T}{dx} \vec{b} + \vec{a}^T \frac{d \vec{b}}{dx}$$



Datenfit abstrakt...

- Mit Hilfe der Definition des Skalarprodukts

$$x^T \cdot y = \sum_{k=1}^n x_k y_k$$

- lässt sich der χ^2 -Fit als Minimierung der L^2 -Norm, d.h. des Residuen Vektors $(f-y)$ der Länge m schreiben:

$$\chi^2(a) = \sum_k (f_k(a) - y_k)^2$$

$$\chi^2 = (f(a) - y)^T \cdot (f(a) - y) = \langle f_a - y | f_a - y \rangle$$

$$\chi^2 \equiv \|f_a\|^2 + \|y\|^2 - 2 y^T \cdot f_a$$

$$\nabla_a \chi^2 = 2 (f_a - y)^T \cdot \nabla_a f_a = 0$$



Diskret versus stetig

- Anstatt zu fordern, dass die Modellfunktion die diskreten Datenpunkte möglichst gut approximiert

$$\chi^2(\vec{a}) = \sum_k (f(\vec{a}, x_k) - y_k)^2$$

- ist es auch möglich bei bekannter Funktion $y(x)$ das Quadrat der $L^2(a, b)$ (Integral)norm zu minimieren:

$$\chi^2(\vec{a}) = \int_{x_0}^{x_n} dx (f(\vec{a}, x) - y(x))^2 = \|f - y\|_2^2$$

$$\Rightarrow \int_{x_0}^{x_n} dx f(\vec{a}, x) \frac{\partial f}{\partial a_k} = \int_{x_0}^{x_n} dx y(x) \frac{\partial f}{\partial a_k}$$

„Skalarprodukt“: $\langle f | \partial f \rangle = \langle y | \partial f \rangle$



Optimierung per Zielfunktion

- Wir nehmen im Folgendem an, dass eine zu optimierende Zielfunktion $f : X \rightarrow \mathbb{R}$ existiert.
- Bemerkung: Maxima von f sind Minima von $-f$.
- Häufig wird das Problem so formuliert, dass nur noch Parameter zu bestimmen sind und $x \in X$ sind n -Tupel des \mathbb{R}^n wofür mit den Methoden der Numerischen Mathematik zahlreiche Verfahren zur Verfügung stehen.
- Selten stehen exakte geschlossene Lösungen zur Verfügung und es werden Nährungsverfahren verwendet, um eine (Cauchy)Folge, die gegen die gesuchte Lösung konvergiert zu erzeugen.



Iterationsverfahren

- Iterationsverfahren beruhen meist auf der Anwendung des **Banachschen Fixpunktsatzes**. Sei A eine kontrahierende Selbstabbildung von X mit $x, y \in X$

$$\|Ax - Ay\| \leq \lambda \|x - y\|$$
 und $0 < \lambda < 1$ fest dann besitzt A einen Fixpunkt ζ in X .
- Zur näherungsweise Bestimmung der Nullstelle einer Funktion $f : X \rightarrow Y$ wird mit einer Iterationsfunktion $\phi_f : X \rightarrow X$ eine konvergente Folge x_k konstruiert

$$x_{k+1} := \phi_f(x_k) \quad \lim_{k \rightarrow \infty} x_k = \zeta$$
- Der Fixpunkt $\phi_f(\zeta) = \zeta$ dieser Folge ist dann die gesuchte Nullstelle der Funktion f .



Lineare Abbildungen

- Ein Operator $A: V \rightarrow W$ des Vektorraums V heißt linear, wenn auf seiner Definitionsmenge für alle x, y und Skalare α gilt:

$$A(x + y) = Ax + Ay$$

$$A(\alpha x) = \alpha Ax$$

- (Für Techniker) Wird der *Input* x in den *Output* y mit beliebigem $0 < \lambda < 1$ vermöge

$$\|Ax\| \leq \lambda \|x\|$$

- transformiert, so ist A ein stetiger Operator, d.h. für alle Folgen $x_n \rightarrow x$ gilt $Ax_n \rightarrow Ax$.



Beispiele linearer Abbildungen

- Differentiation auf $C^{(1)}(T)$: $D_x f = \nabla f(x)$

$$D_x(f + g) = D_x f + D_x g$$

$$D_x(\alpha f) = \alpha D_x f$$

- Oder die Integration: $I_x(f) = (If)(x) = \int^x f(u) du$

$$I_x(\alpha f) = \int^x \alpha f(u) du = \alpha I_x(f)$$

$$I_x(f + g) = \int^x f(u) + g(u) du = I_x(f) + I_x(g)$$

- Bemerkung $C^{(n)}[a,b]$ und $L^p(a,b)$ sind Vektorräume mit den darauf definierten Funktionen als Vektoren!



Banachscher Fixpunktsatz

- Ist der Operator $A: V \rightarrow V$ eine kontrahierende Selbstabbildung, auf einer abgeschlossenen Teilmenge X von V , d.h für alle $x, y \in X$ existiert ein $0 < \lambda < 1$ mit

$$d(Ax, Ay) < \lambda d(x, y) \quad \text{bzw.} \quad \|Ax - Ay\| < \lambda \|x - y\|$$
- So ist A stetig und jeder Startwert $x_0 \in X$ konvergiert gegen den Fixpunkt x von A , d.h. für $x_{n+1} := Ax_n$ gilt:

$$\|Ax_n - Ax\| \leq \frac{\lambda^n}{1 - \lambda} \|x_1 - x_0\|$$

$$\|Ax_n - Ax\| \leq \frac{\lambda}{1 - \lambda} \|x_n - x_{n-1}\|$$



Bemerkung

- Der „Operator“ A des Banachschen Fixpunktsatzes kann „alles Mögliche“ sein:
 - Eine Matrix: $\vec{x}_{n+1} = A \vec{x}_n$
 - Eine Funktion: $x_{n+1} = f(x_n)$
 - Die Differentiation: $x_{n+1} = D_x f(x_n)$
 - Die Integration: $x_{n+1} = \int^{x_n} f(u) du$
 - Der Kern eines Integrals: $x_{n+1} = \int k(x_n, u) f(u) du$
 - Etc...
- Entscheidend ist die Definition einer für das jeweilige Problem geeigneten Norm mit $\|A\| < 1$.



Iterationsbeispiel

- Beispiel: $f(x) = x - \cos(x)$
- Es liegt nahe $\phi(x) = \cos(x)$ zu setzen und der Fixpunkt $\cos(\zeta) = \zeta$ ist zugleich Nullstelle von f .
- Da $|\cos(x)| \leq 1$ ist $\phi(x)$ eine Abbildung des Intervalls $X = [-1, 1]$ auf sich selbst. Mit dem Startpunkt $x_0 = 0.5$ ergibt sich die Iterationsfolge:

$$x_1 = 0.877$$

$$x_5 = 0.767$$

$$x_2 = 0.639$$

$$x_6 = 0.719$$

$$x_3 = 0.802$$

....

$$x_4 = 0.695$$

$$x_{50} = 0.73908513... \cong x_{\infty} =: \zeta$$



Entwicklung von Iterationsverfahren

- Das Auffinden der Funktion $\phi(x)$ geschah durch scharfes Hingucken oder Raten und der Lösungsweg war nicht sehr systematisch.
- Wenn über die Zielfunktion f weitere Eigenschaften wie Linearität, Periodizität, Darstellung als Polynom oder bestimmte Symmetrieeigenschaften bekannt sind, so stellt die Mathematik einen besseren Zugang zur Verfügung.
- Insbesondere gibt es für stetige und differenzierbare reelle Funktionen Methoden wie „Regular Falsi“, Intervallhalbierung oder das Newton-Verfahren, zunächst in einer Dimension.



Newton-Verfahren

- Ausgehend von der Taylorentwicklung der reellwertigen Funktion f wird zur Suche der Nullstelle ζ eine Iterationsfunktion ϕ bestimmt:

$$f(x) = f(\zeta) + (x - \zeta) f'(\zeta) + \frac{1}{2} f''(\zeta)(x - \zeta)^2 + \dots$$

- Abbruch nach dem linearen Term liefert:

$$\zeta = x - \frac{f(x) - f(\zeta) \equiv 0}{f'(\zeta)} + R(x)$$

- so dass sich als Iterationsfunktion ϕ ergibt:

$$\phi(x) = x - \frac{f(x)}{f'(x)}$$



Iterationsbeispiel

- Beispiel: $f(x) = x - \cos(x)$ $\phi(x) = x - \frac{x - \cos(x)}{1 + \sin(x)}$
 $f'(x) = 1 + \sin(x)$
- Auch dies ist eine kontrahierende Selbstabbildung für $x \in]0, 1]$, da $|\phi(x)| \leq |\cos(x)| < 1$
- Mit dem Startpunkt $x_0 = 0.5$ lautet die Iterationsfolge:
 - $x_1 = 0,7552224171$ $x_5 = 0,739085133(2) =: \zeta$
 - $x_2 = 0,7391416661$
 - $x_3 = 0,7390851339$
 - $x_4 = 0,7390851332$
- Offensichtlich konvergiert die mit dem Newton-Verfahren gewonnene Folge wesentlich schneller.



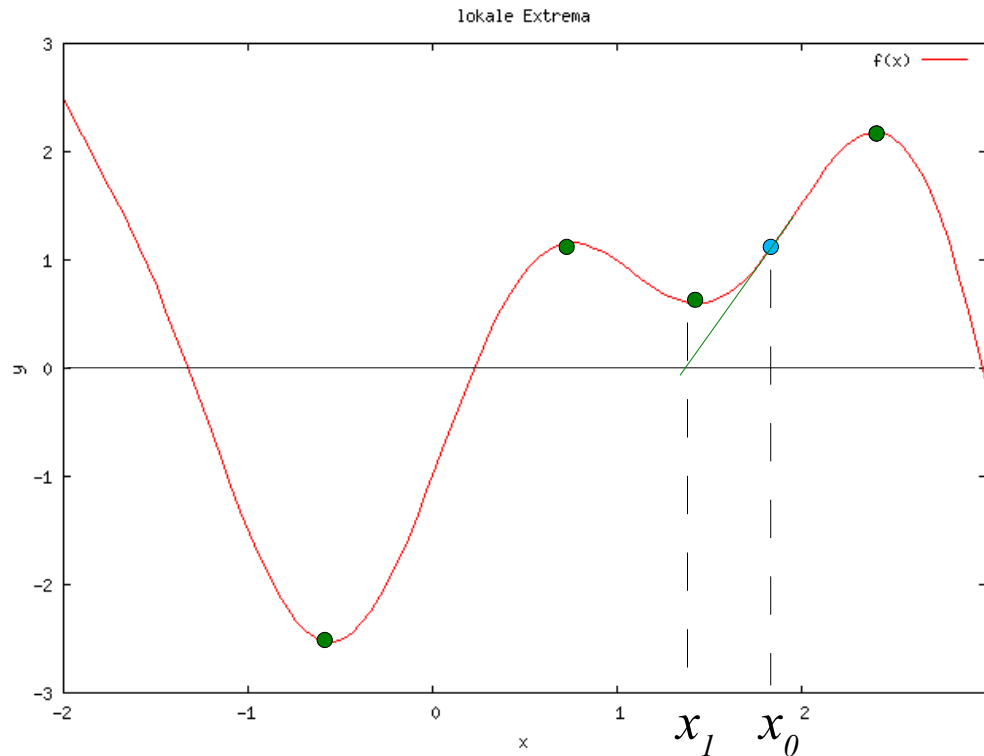
Bemerkungen

- Für eine einfache Nullstelle zeigt das Newton-Verfahren quadratische Konvergenz, wenn der Startwert x_0 hinreichend nahe bei ζ liegt.

$$|x_{n+1} - \zeta| \leq K (x_n - \zeta)^2$$

- Wichtig ist, dass die Ableitung von f in der Umgebung von ζ nicht verschwindet. Für mehrfache Nullstellen gilt $f'(\zeta) = 0$ und die Konvergenz verschlechtert sich.
- Liegen mehrere Nullstellen in der Nähe des Startwerts x_0 (oder x_k) so konvergiert das Verfahren unter Umständen nicht und/oder oszilliert.

Lokale Extrema



- Wird ein ungünstiger Startwert gewählt, konvergiert die Iteration gegen das falsche Minimum oder ein Maximum...



Newton Verfahren reloaded

- Die Konvergenz des Newton Verfahrens hängt von dem gewählten Startvektor x_0 ab. Wird dieser zufällig gewählt und existieren mehrere Lösungen, so ist eine interessante Frage, gegen welche Lösung konvergiert das Verfahren in Abhängigkeit vom Startwert?
- Als Beispiel das Newton-Verfahren im Komplexen, gesucht sind die Einheitswurzeln von $f(z) = z^3 - 1$.

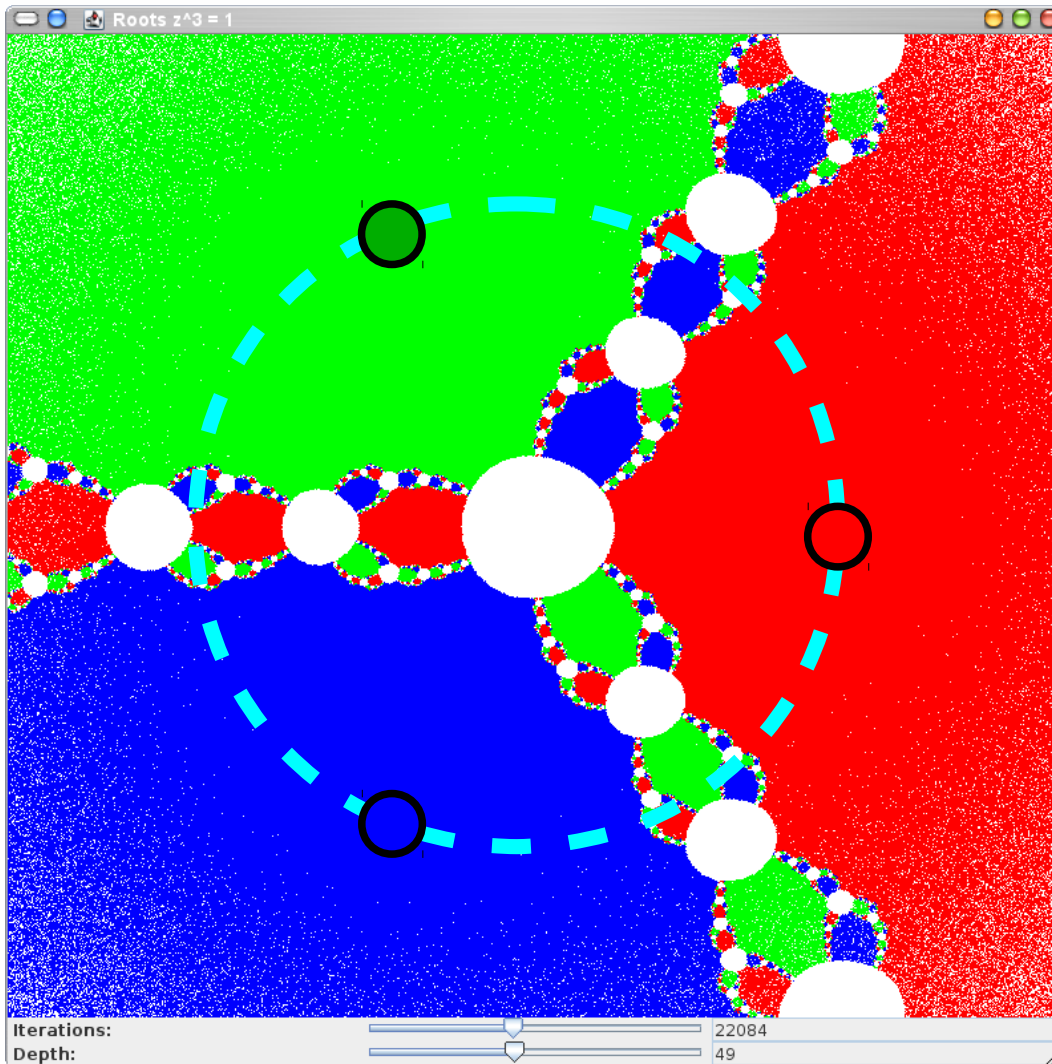
Bekanntlich sind dies 1 , $e^{i \frac{2}{3} \pi}$ und $e^{-i \frac{2}{3} \pi}$.

Komplexe Iterationsfolge mit random $z_0 \in \mathbb{C}$

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)} \qquad z_{n+1} = z_n - \frac{z_n^3 - 1}{3z_n^2}$$



Fraktale Lösungsmenge



Farbliche Markierung
der Konvergenzradien
der drei Lösungen zu
 $z^3 = 1$.

Die Abbildung zeigt die
typische fraktale Struktur
einer *Julia*-Menge.



Eindimensionale Optimierung

- Kochrezept zum Auffinden der Extrema von
 $f : X \subseteq \mathbb{R} \rightarrow \mathbb{R}$
- Ermittlung der Ableitung $f'(x)$.
- Finden der Nullstellen x_0 von $f'(x)$.
- Iterative Lösung von $f'(x) = 0$ mit geeigneter
 Abbildung $\phi_f : X \rightarrow X$
- Untersuchung der Eigenschaften von $f''(x_0)$.
- Untersuchung ob auf dem Rand von X weitere
 Extrema vorliegen.



Übungen II

1. Wo liegt das Maximum von $x^n e^{-x}$
2. Wie lautet die Iterationsfunktion Φ zur Berechnung von $\sqrt[\alpha]{x}$ mit $x > 0$ und $\alpha \in \mathbb{R}^+$?
3. Berechne mit der Lösung 2. $\sqrt[3]{5}$ auf 5 Stellen genau durch Iterationen mit Zwischenergebnissen.



Partielle Ableitungen

- Analog zum Rechnen in einer Variablen lassen sich die (lokalen) Extrema einer Funktion f durch die Nullstellen der ersten partiellen Ableitungen $\partial_{x_k} f$ bestimmen.

- Wie im Eindimensionalen sind diese definiert als

$$\frac{\partial f(x)}{\partial x_k} = \lim_{h \rightarrow 0} \frac{f(x + h * e_k) - f(x)}{h}$$

- wobei e_k ein Einheitsvektor in die k -Richtung ist.



Gradient einer Funktion

- Es sei $X \subset \mathbb{R}^n$ eine Teilmenge und $f : X \rightarrow \mathbb{R}$ eine p -mal stetig differenzierbare Funktion mit $p > 0$ $f \in C^p(X)$.
- Dann existieren die partiellen Ableitungen $\partial_{x_k} f(\zeta)$ für $\zeta \in X$ und der Gradient

$$(\text{grad } f)(\zeta) \equiv \nabla f(\zeta) := \begin{pmatrix} \frac{\partial f(\zeta)}{\partial x_1} \\ \vdots \\ \frac{\partial f(\zeta)}{\partial x_n} \end{pmatrix}$$

- definiert eine Funktion $\nabla f : X \rightarrow \mathbb{R}^n$.



Hessematrix

- Die Hessematrix H_f einer Funktion f im Punkt x ist definiert als die symmetrische $n \times n$ -Matrix der zweiten partiellen Ableitungen:

$$H_f = (\text{Hess } f)(x) = (D_i D_j f(x))_{1 \leq i \leq n, 1 \leq j \leq n}$$

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$



Jakobimatrix

- Die Jakobimatrix J_f einer Funktion $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ im Punkt x ist definiert als $m \times n$ die Matrix Df der ersten partiellen Ableitungen:

$$J_f(x) = (D_i f_j(x))_{1 \leq i \leq n, 1 \leq j \leq m}$$

$$J_f(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$



Taylor für $n > 1$ Dimensionen

- Ist $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Abbildung mehrere Veränderlicher in mehreren Dimensionen, so gilt es das Verfahren entsprechend zu verallgemeinern.
- Um das Gleichungssystem $f(x) = 0$, ($x \in \mathbb{R}^n$) zu lösen müssen simultan die m Gleichungen $f_k(x) = 0$ bzw. die Lösung von $\|f(x)\| = 0$ bestimmt werden.
- Ausgangspunkt ist auch hier wieder die Taylor-Entwicklung in n -Dimensionen.
- Für ein n -Tupel $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ natürlicher Zahlen sei:

$$|\alpha| := \alpha_1 + \dots + \alpha_n$$

$$\alpha! := \alpha_1! \alpha_2! \dots \alpha_n!$$



Satz von Taylor in n-Dimensionen

- Ist f eine $|\alpha|$ -mal stetig differenzierbare Funktion so setzt man

$$D^\alpha f = D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$$

- wobei $D_k^{\alpha_k} = \underbrace{D_k \cdots D_k}_{\alpha_k \text{-mal}}$

- Für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ sei $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$
- dann gilt für eine offene Umgebung $U_\xi \subset \mathbb{R}^n$, $x \in U_\xi$

$$f(x + \xi) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x)}{\alpha!} \xi^\alpha + O(\|\xi\|^k)$$



Maclaurin-Reihe

- Um die Formel zu illustrieren werde eine reellwertige Funktion f mit Variablen x, y um den Punkt $x_0 = (0, 0)$ bis zur 2.-ten Ordnung entwickelt:

$$f(x, y) \approx f(x_0) + \frac{\partial f(x_0)}{\partial x} x + \frac{\partial f(x_0)}{\partial y} y + \frac{1}{2} \left(\frac{\partial^2 f(x_0)}{\partial^2 x} x^2 + 2 \frac{\partial^2 f(x_0)}{\partial x \partial y} x y + \frac{\partial^2 f(x_0)}{\partial^2 y} y^2 \right)$$

- oder in Kurzform für $x, x_0 \in \mathbb{R}^n$

$$f(x) \approx f(x_0) + \sum_{j=1}^n x_j \partial_{x_j} f + \frac{1}{2} \sum_{i,j} x_i x_j \partial_{x_i x_j}^2 f$$

$$f(x) \approx f(x_0) + x^T \cdot \nabla f + \frac{1}{2} x^T H_f x$$



Mehrdimensionale Optimierung

- Kochrezept zum Auffinden der Extrema von

$$f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$$
- Ermittlung des Gradienten $\text{grad } f(x)$.
- Finden der Nullstellen x_0 von $\text{grad } f(x)$.
 - Evt. iterative Lösung von $\text{grad } f(x) = 0$ mit geeigneter Funktion $\Phi : X \rightarrow X$
- Untersuchung der Eigenschaften von $\text{Hess } f(x_0)$:
 - positiv definit \Rightarrow lokales Minimum,
 - negativ definit \Rightarrow lokales Maximum,
 - Indefinit \Rightarrow Sattelpunkt..
 - Untersuchen des Rand von X auf weitere Extrema.



Definitheit einer Matrix

Eine symmetrische Matrix $A \in M(n \times n, \mathbb{R})$ heißt

- *positiv definit*, falls

$$x^T A x > 0 \quad \forall x \in \mathbb{R}^n \setminus 0$$

—

- *positiv semidefinit*, falls

$$x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n \setminus 0$$

—

- *indefinit*, falls es Vektoren $x, y \in \mathbb{R}^n$ gibt mit

$$x^T A x > 0 \quad \text{und} \quad y^T A y < 0$$



Eigenwertkriterium

Vollkommen analog heißt die Matrix A

- ***negativ (semi)definit*** wenn die Matrix $-A$ positiv (semi)definit ist.

Es sei $v_k \in \mathbb{R}^n$ ein Eigenvektor von A mit Eigenwert λ_k , d.h. $A v_k = \lambda_k v_k$. Entwickeln von x nach den Eigenvektoren liefert:

$$x^T A x = \sum_{k=1} \alpha_k^2 \lambda_k$$

- A ist ***positiv (negativ) definit*** wenn alle Eigenwerte **positiv (negativ)** sind und ***positiv (negativ) semi definit*** wenn sie ≥ 0 (bzw. ≤ 0) sind.



Abschnittskriterium

Es sei Δ_k die k -te Abschnittsdeterminante der symmetrischen $n \times n$ Matrix $A := (a_{ij})$

$$\Delta_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix}$$

Dann gilt A ist

- **positiv definit**, falls $\Delta_k > 0$ für alle $k=1, \dots, n$
- **negativ definit**, falls die Vorzeichen der Δ_k alternieren beginnend mit $a_{11} = \Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0$ usw.



Bemerkungen

- Die Definitheitskriterien zu überprüfen ist für $n \geq 3$ nicht ganz einfach, da i. A. die Eigenwerte nicht bekannt sind. Auch ist die Berechnung der Abschnittsdeterminanten meist aufwendig...
- Die Kriterien der Hessematrix H_f versagen, wenn die Voraussetzung, dass die Funktion f zweimal stetig differenzierbar ist und alle partiellen Ableitungen in den Extrempunkten existieren nicht erfüllt sind.
- Das Versagen der Kriterien bedeutet aber nicht, dass nicht doch Extrempunkte vorliegen können. Z.B. wenn H_f semi definit ist...



Beispiel I

- Es sei $f(x) = x_1^2 + x_2^4 + a$ $x = (x_1, x_2) \in \mathbb{R}^2$

$$\nabla f = \begin{pmatrix} 2x_1 \\ 4x_2^3 \end{pmatrix} \Rightarrow \nabla f(x_0) = 0 \Leftrightarrow x_0 = (0, 0)$$

$$H_f = \begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix} \Rightarrow H_f(x_0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

- H_f ist positiv semi definit, was auch die Eigenwerte $\lambda_1 = 2$ und $\lambda_2 = 0$ bestätigen. \Rightarrow so nicht entscheidbar
- Dennoch ist x_0 lokales und globales Minimum von f
da $a = f(0, 0) < f(x_1, x_2) \quad \forall x = (x_1, x_2) \in \mathbb{R}^2 \setminus 0$



Beispiel II

- Es sei $f(x) = x_1^2 + x_2^3 + a$ $x = (x_1, x_2) \in \mathbb{R}^2$

$$\nabla f = \begin{pmatrix} 2x_1 \\ 3x_2^2 \end{pmatrix} \Rightarrow \nabla f(x_0) = 0 \Leftrightarrow x_0 = (0, 0)$$

$$H_f = \begin{pmatrix} 2 & 0 \\ 0 & 6x_2 \end{pmatrix} \Rightarrow H_f(x_0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

- Gleiche H_f wie im Beispiel I jedoch x_0 ist ein Sattelpunkt, denn $\forall (x_1, x_2) \in \mathbb{R}^2 \setminus \{x_0\}$ gilt:
 $a < f(\pm x_1, 0) \quad f(0, -x_2) < a < f(0, x_2)$



Messreihen fitten

- Die Minimierung von $\chi^2 : \mathbb{R}^p \rightarrow \mathbb{R}$ führt auf das p -dimensionale Gleichungssystem mit dem Vektor a :

$$g_j(a) = \sum_{k=1}^m (y_k - f(x_k; a)) \frac{\partial f(x_k)}{\partial a_j} = 0; \quad 1 \leq j \leq p$$

- Es gilt daher die Nullstelle von $g : \mathbb{R}^p \rightarrow \mathbb{R}^p$ zu bestimmen. Was je nach Modellfunktion f_a meistens nur mit Näherungsmethoden möglich ist.

– Ausnahmen sind Polynome.

- **Bemerkung:** Die vektorielle Funktion g lässt sich prägnant per Gradient Operator zum Argument a darstellen: $g(a) = \nabla_a \chi_f^2 = (y - f) \nabla_a f$



Lineare Regression

- Mit $f(x, a) = a_0 + a_1 x$ lautet ergibt sich das Gleichungssystem

$$\sum_{k=1}^n (y_k - (a_0 + a_1 x_k)) = 0$$

$$\sum_{k=1}^n (y_k - (a_0 + a_1 x_k)) x_k = 0$$

- und mit dem Mittelwert $\bar{x} := 1/n \sum_{j=1}^n x_j$ verkürzt sich dies zur Matrixgleichung

$$\begin{pmatrix} 1 & -\bar{x} \\ \bar{x} & -\bar{x}^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \bar{x}\bar{y} \end{pmatrix}$$



Fourier-Transformation

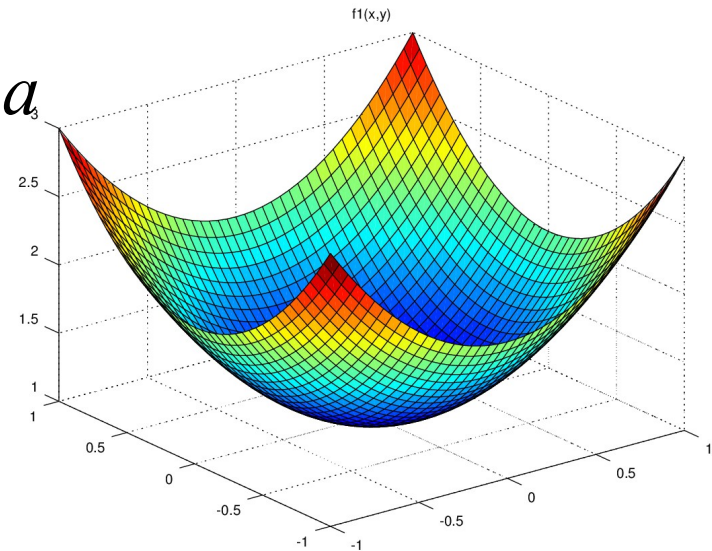
- Häufig wird ein Polynom oder eine Gauss-Funktion als Modell verwendet.
- Wird stattdessen ein Polynom in *sin* und *cos* Termen gewählt, so liefert der Fit die FFT Koeffizienten a_k .

$$f(\vec{a}, x) = \sum_{k=0}^{p/2} a_{2k} \cos(k \omega x) + a_{2k+1} \sin(k \omega x)$$

- Allerdings wird bei der FFT die $L^2(0, 2\pi)$ Norm verwendet.

Diskutiere die möglichen Extrema der folgenden Funktionen $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

1. $f(x,y) = a + x^2 + y^2$
2. $f(x,y) = a + x^2 - y^2$
3. $f(x,y) = x^3 + y^3 - 3xy$
4. $f(x,y) = x^2 + y^2 - 2xy + a$
5. $f(x,y) = (x^2 + y^2)^{0.5}$





Das Newton-Verfahren im \mathbb{R}^n

- Gesucht sind Lösungen der Gleichung $f(x) = 0$ für $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$
- Ganz analog zum eindimensionalen Fall wird in der Nähe der Nullstelle eine Iterationsfunktion $\phi : U \rightarrow U$ definiert

$$\phi(x) = x - [J_f(x)]^{-1} f(x)$$

$$x_{k+1} = x_k - [J_f(x_k)]^{-1} f(x_k)$$

- Der „Quotient“ $f(x)/f'(x)$ des eindimensionalen Newton-Verfahrens geht auf im Matrixprodukt der inversen Jakobi-Matrix $Df(x)^{-1}$ mit der Funktion f .



Optimierung im \mathbb{R}^n

- Ist als Zielfunktion eine Norm $f = \|g\|: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ zu optimieren, so führt dies immer dazu die Nullstelle des Gradienten ∇f zu bestimmen. Muss hierzu das Newton-Verfahren herangezogen werden, so enthält die Jakobi-Matrix die zweiten Ableitungen von f , ist also die Hessematrix und die Iterationsvorschrift lautet:

$$x_{k+1} = x_k - [H_f(x_k)]^{-1} \nabla f(x_k)$$

Bemerkung:

- Neben den Konvergenzeigenschaften erweist sich hierbei die Inverse der Hessematrix für große n als numerisch ungünstig.



Weitere Verfahren

- Um das Newton-Verfahren zu beschleunigen gibt es mehrere Möglichkeiten.
- Verzicht auf die Invertierung der Jakobi-Matrix, statt dessen mit $y_{k+1} = x_{k+1} - x_k$ Lösung der Gleichung

$$J_f(x_k) y_{k+1} = -f(x_k)$$

z.B. mit dem Gauss-Seidel-Verfahren ermitteln

- Approximation der Inversen Jakobi-Matrix.
- Konjugierte Gradienten Methoden etc.