

## Optimierungsalgorithmen II (Faktsetzung)

### Methoden basierend auf dem Newtonverfahren

Bislang haben wir nur die Ableitung erster Ordnung zur Richtungsbestimmung genutzt, jetzt eine Ordnung höher (Taylorreihen.)

$$f(\underline{x}) = f(\underline{x}_0) - \underset{\substack{\uparrow \\ \text{negativ} \\ \text{Gradient}}}{g} \cdot (\underline{x} - \underline{x}_0) + \frac{1}{2} (\underline{x} - \underline{x}_0) \cdot \underset{\substack{\sim \\ \text{Hess'sche Matrix} \\ \text{(vollständige zweite Abl.)}}}{H} \cdot (\underline{x} - \underline{x}_0)$$

Die erste Ableitung muss bei Minimum verschwinden,  
→ Approximiert:

$$f'(\underline{x}_0)|_{\underline{x}=\underline{x}_0} = -g + \frac{1}{2} \underline{H} \cdot (\underline{x} - \underline{x}_0) + \frac{1}{2} (\underline{x} - \underline{x}_0) \cdot \underline{H}$$

$\underline{H}$  ist symmetrisch

$$= -g + \underline{H} \cdot (\underline{x} - \underline{x}_0) = 0$$

$$\Rightarrow \underline{H} \cdot (\underline{x} - \underline{x}_0) = g$$

$$\underbrace{(\underline{x} - \underline{x}_0)}_{\Delta \underline{x}} = \underline{H}^{-1} \cdot g \quad \left( \text{vorausgesetzt } \underline{H} \text{ ist invertierbar} \right)$$

Diese Gleichung erlaubt die Richtung  $\underline{x} - \underline{x}_0$  zu bestimmen die am schnellsten zum Minimum führt:

- 1) Zur Illustration, überführen  $\Delta \underline{x}$  in das Koordinatensystem der Eigenvektoren von  $\underline{H}$

$$\Delta \underline{x}' = (\Delta x_1', \dots, \Delta x_i', \dots, \Delta x_n')$$

$$\Delta x_i' = \frac{f_i}{z_i} \leftarrow \begin{array}{l} \text{Projektion des Gradienten entlang des} \\ \text{Eigenvektors der H-Matrix} \end{array}$$

$\nearrow$  zugehörige Eigenwert

Positiv in der Nähe einströmen

- Ist ein negativer Wert dabei, ist es ein Sattelpunkt
- 2) Nahe eines Minimums können Eigenwerte klein werden  
 $\Rightarrow$  divergent Schritt! In dem Fall wird teilweise ein Verstrahlungsradius für die Stepsize vorgegeben, ist der Schritt darüber  $\Rightarrow$  Linesearch
- 3) Die Schrittweite sollte den Bereich, wo die Taylor-Näherung gut approximiert nicht verlassen.

Falls Hess'sche Matrix negative Eigenwert, hilft ein Schrittperzent

$$\Delta x' = - \frac{f_i}{z_i} \leftarrow \begin{array}{l} \text{kleiner als der kleinste Eigenwert} \end{array}$$

- 4) Es ist sehr aufwendig die Hessian Matrix zu bestimmen. In dem Fall gibt es iterative Methoden, die mit Näherung der Hess'schen Matrix arbeiten z.B. BFGS Methode

$$\underline{H}_0 = Id \Rightarrow \text{erster Schritt} \Rightarrow \text{Steepest Descent Linesearch}$$

$$\underline{H}_{k+1} = \underline{H}_k + \underline{U}_k + \underline{V}_k \quad \text{wobei } \underline{y}_k = \underline{g}_k - \underline{g}_{k+1}$$

$$\text{und } \underline{s}_k = \underline{x}_{k+1} - \underline{x}_k$$

$$\text{mit } \underline{U}_k = \frac{\underline{y}_k \underline{y}_k^T}{\underline{y}_k^T \underline{y}_k} \quad \underline{V}_k = \frac{\underline{H}_k \underline{s}_k \underline{s}_k^T \underline{H}_k}{\underline{s}_k^T \underline{H}_k \underline{s}_k}$$

Die Entwicklung kann dann ohne die Anwendung einer Matrix angesetzt werden (Formel ohne Herleitung)

$$\underline{H}_{k+1}^{-1} = \underline{H}_k^{-1} + \frac{(\underline{s}_k \underline{y}_k^T + \underline{y}_k \underline{H}_k^{-1} \underline{y}_k^T)(\underline{s}_k - \underline{s}_k)}{(\underline{s}_k^T \underline{y}_k)^2} - \frac{\underline{s}_k \underline{H}_k^{-1} \underline{y}_k^T + \underline{y}_k \underline{H}_k^{-1} \underline{s}_k^T}{\underline{s}_k^T \underline{y}_k}$$

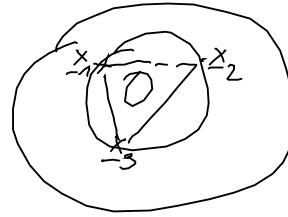
Wichtig: Hier muss keine Matrix gespeichert werden!

Meist eine gute Wahl, die Bedeutung der Hesseschen Matrix unklar ...

### Beispiel für gradientenfreie Methode Nelder-Mead

Man findet oft den Gradienten aufwendig, eine Möglichkeit ist dann Nelder-Mead

Idee: Wenn wir ein  $N$ -Dimensionales Problem haben, dann wird ein Simplex von  $N+1$  Dimensionen verwendet.  
z.B.  $N=2$  Dreieck



Die Vektoren sind sortiert, d.h.

$x_1, x_2, x_3, \dots, x_{N+1}$  gilt

$$f(x_1) \leq f(x_2) \leq \dots \leq f(x_{N+1})$$

Der schlechteste Wert wird bei jeder Iteration ersetzt:

$$x_{N+1}^{\text{neu}} = (1+\mu) \frac{1}{N} \sum_{i=1}^N x_i - \mu x_{N+1}$$

Das wird für verschiedene Werte von  $\mu$  versucht

$\mu = \mu_0, \mu = 2\mu_0, \frac{1}{2}\mu_0, -\frac{1}{2}\mu_0$ , ergibt sich ein Verhaltensgesetz.

Entspricht Reflektion, Expansion, Kontraktion, Shrinkage

(Achtung, kann auch zu nicht linearer Konvergenz  $\Rightarrow$  Problem)

### Allgemeine Bemerkung zu Optimierung

1) Das Koordinatensystem spielt wichtige Rolle, die richtigen Koordinaten können die Konvergenz beschleunigen:

Beispiel

$$E = [1 - e^{-x}]^2$$

Hermiten in  $x$   
 Mittelstufe (5. Iteration)  $y = e^{-x}$  schlechte (2. Iteration) und  $z = e^x$  Tangent (7. Iteration)

Sonst spielt die richtige Basis eine Rolle.

2.) Oft gibt es auch Nebenbedingungen.

1) Man addiert zu zu optimierender Funktion  $C[f]$  eine Kostenfunktion  $C_{NB}[f]$  hinzu, die Abwärtswert von  $C_{NB}$  mit höherem Funktionswert bestrahlt und optimiert  $C[f] + C_{NB}[f]$   
 Nachteil: Nebenbed. werden in Approximation eingeschaltet.

2) Lagrange Multiplikatoren  
 $L(x) = f(x) - \sum_i \lambda_i (g_i(x) - c_i)$   
 Nebenbed.

$\lambda_i$  wird zusätzlich optimiert und  $L$  minimiert  
 Oft ist  $\lambda_i$  ein physikalisch relevante Größe.

3) Wichtig, sollte das global Minimum wichtig sein, so muss on top weitere Methode verwendet werden.

IV.M Zeitliche Dynamik: Schrödingergl., Blochgleichung, Dichtematrixtheorie etc.

Bislang haben wir nur Eigenwertprobleme und stationäre Probleme angeschaut.

Jetzt schauen wir uns zeitabhängige Probleme an

1. Beispiel zeitabhängige Schrödingergleichung, im Raum diskretisieren:

$$i\hbar \partial_t |\psi\rangle = H |\psi\rangle$$

z.B. für wave Exzitonen

$$H = \underbrace{H_0}_{\substack{\text{bislang} \\ \text{Exziton} \\ \text{freie Bewegung} \\ \text{+ Coulomb}}} + \underbrace{H_{\text{el-Licht}}}_{\substack{\text{Wechselwirkung mit} \\ \text{dem Licht} \\ \text{(lineare Optik)}}$$

$$H_{\text{el-Licht}} = \delta(\underline{r}_e - \underline{r}_h) d \cdot \underline{E}(\underline{r}) \\ \equiv \delta(\underline{r}) d \cdot \underline{E}$$

↑  
Kanal

Bewerger:

$$i\hbar \partial_t \psi(\underline{r}) = \left( -\frac{\hbar^2}{2m} \Delta_r + V_{\text{ext}}(\underline{r}) \right) \psi(\underline{r}, t) + \delta(\underline{r}) d \cdot \underline{E}(t)$$

Delta Funktion,  
kann durch Zerlegen  
in Intervalle und Integri-  
erung sein.

Damit ist es z.B. möglich, die zeitliche  
Dynamik der optischen Anregung zu berechnen.

Vorteil: im Vergleich zu Eigenwertproblem in  $\psi$  (nicht auf je  
nach Zeit lösen) muss im Spindel gehalten werden.

Zeit lösen löst Eigenwertproblem für relevante Zustände  
implizit.

Nachteil: Eigenwerte und -vektoren nicht bekannt

Kann leicht auf Nichtlineare Optik erweitert werden (erfordert  
Bisexitonen), Deplatz z.B. Quanten-trajectories und  
Jump-Operatoren.