

Fortsetzung Optimierungsverfahren

Letzte Woche Gradientenbasierte Verfahren,
jetzt:

Methoden basierend auf Newton Verfahren

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

$$f(x) = f(x_0) - \underset{\substack{\uparrow \\ \text{negative} \\ \text{Gradient}}}{g} \cdot (x - x_0) + \frac{1}{2} (x - x_0) \cdot \underset{\substack{\sim \\ \text{Hessische} \\ \text{Matrix}}}{H} \cdot (x - x_0)$$

Die erste Ableitung muss bei Minimum
verschwinden, Approximation ableite
und 0 setzen.

$$\begin{aligned} f'(x) \big|_{x=x_0} &= -g + \frac{1}{2} H \cdot (x - x_0) + \frac{1}{2} (x - x_0) \cdot H \\ &\quad H \text{ symmetrisch} \\ &= -g + \frac{1}{2} H \cdot (x - x_0) = 0 \end{aligned}$$

$$\Rightarrow H \cdot (x - x_0) = g$$

$$\Leftrightarrow \underbrace{(x - x_0)}_{\Delta x} = H^{-1} \cdot g \quad (\text{vorausgesetzt } H^{-1} \text{ existiert})$$

Diese Gleichung Δx erlaubt die Bestimmung der Richtung
 $x - x_0$ zu zum Minimum.

1) Zur Illustration, wird Δx in das Koordinatensystem
der Eigenvektoren zu H überführt

$$\Delta x' = (\Delta x'_1, \dots, \Delta x'_i, \dots, \Delta x'_n)$$

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

Positiv in der Nähe ein Minimum.

Falls ein negativer Wert dabei $\Rightarrow$ Sattelpunkt

2) Nähe ein Minimum kein Eigenvektor klarmachen...
 $\Rightarrow$ Mehrere Schritte!

In dem Fall wird teilweise ein Vertrauensradius
 für die Stepsize vorgegeben, falls der Step zu groß
 $\Rightarrow$ Line search.

3) Die Schrittweite sollte den Bereich der
 Gültigkeit der Taylornäherung nicht verlassen.

Falls Hessian negative Eigenwert hat, hilft ein
 Shiftoperator

$$\Delta x' = - \frac{f_i}{\epsilon_i - \lambda}$$

$\leftarrow$ wenn es der kleinste Eigenwert ...

4) Es ist sehr aufwendig, die Hessian Matrix
 zu bestimmen.

In dem Fall gibt es iterative Methoden, die eine
 Näherung der Hesse'schen Matrix ermöglichen.

z.B. BFGS-Methode

$$\underline{H}_0 = Id \Rightarrow \text{erster Schritt} \Rightarrow \text{Steepest Descent} \\ \text{Line Search.}$$

$$\underline{H}_{k+1} = \underline{H}_k + \underline{U}_k + \underline{V}_k$$

Wobei $y = g_1 - g_{k+1}$ und $s_k = x_{k+1} - x_k$

mit
$$\underline{U}_{11} = \frac{y_1 \cdot y_k^T}{y_k \cdot s_k} \quad \underline{V}_{11} = - \frac{\underline{H} \cdot s_1 \cdot s_k \cdot \underline{H}}{s_k \cdot \underline{H} \cdot s_k}$$

Die Entwicklung kann ohne die Anwendung einer Matrix ausgerechnet werden. (Formel oder Herleitung)

$$\begin{aligned} \underline{H}_{k+1}^{-1} &= \underline{H}_{11}^{-1} + \frac{(s_1 \cdot y_k + y_k \cdot \underline{H}_{11}^{-1} \cdot y_1)(s_k - s_{k+1})}{(s_1^T \cdot y_1)^2} \\ &\quad - \frac{s_1 \cdot \underline{H}_{11}^{-1} \cdot y_k + y_k \cdot \underline{H}_{11}^{-1} \cdot s_1}{s_1 \cdot y_1} \end{aligned}$$

Wichtig: Hier muss keine Matrix gespeichert werden!

Meist eine gute Wahl, die Bedeutung der H Matrix zu verstehen.

Beispiel für Gradientenfreie Methode: Nelder-Mead

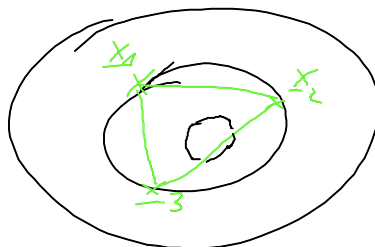
Manchmal ist die Bedeutung eines Gradienten zu aufwendig, dann z.B. Nelder-Mead verwenden.

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

Die Vektoren sind sortiert, d.h.

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

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



Der schlechteste Wert bei jeder Iteration ersetzt:

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

Das wird für verschiedene Werte von μ versucht.

$\mu = \mu_0, \mu = 2\mu_0, \frac{1}{2}\mu_0, -\frac{1}{2}\mu_0$ dies ergibt eine Verbesserung $\Rightarrow$ stop!

Entspricht Reflexion, Expansion, Kontraktion ...

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

Allgemeine Bemerkung zu Optimierung

1) Das Koordinatensystem spielt eine wichtige Rolle. Die richtige Koordinate kann die Konvergenz beschleunigen.

Beispiel:
$$F_{\text{Morse}} = [1 - e^{-x}]^2$$

Iteration in x , $y = e^{-x}$ und $z = e^x$

$\underbrace{\hspace{10em}}$ mittel schnell (5. Iteration)	$\underbrace{\hspace{10em}}$ schnell (2. Iteration)	$\underbrace{\hspace{10em}}$ langsam (7. Iteration)
--	---	---

Genauso spielt auch die richtige Barriere eine Rolle.

2.) Oft gibt es auch Nebenbedingungen!

1) Man addiert zu zu optimierender Funktion $C[f]$, eine Kostenfunktion $C_{NB}[f]$ hinzu, die Abweichen der Nebenbedingung mit hohem Funktionswert bestraft und optimiert $C[f] + C_{NB}[f]$

Nachteil: Nebenbedingung werden nur
approximativ eingehalten.

2) Lagrange Multiplikator Nebenbed.

$$L(x) = f(x) - \sum_i \lambda_i (g_i(x) - c_i)$$

λ_i wird zusätzlich optimiert und
 L minimiert. ~~offiziell~~ λ_i physikalisch relevant.

(Tatsächlich ist das System unterbestimmt
und i Dimensionen müsste entfernt werden,
ist aber zu kompliziert.)

3.) Projektionsmethode

IV.11 Zeitliche Dynamik: Schrödingergl., Blochgleichung, Dichtematrix Theorie

Bislang haben wir uns vorrangig Eigenwertprobleme
und stationäre Probleme angesehen.

Jetzt schauen wir uns zeitabhängige Probleme
an.

1. Beispiel: zeitabhängige Schrödingergl. in
Raum diskretisiert.

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

z.B. für unser Exziton Problem.

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

$$H_{el-L} = \delta(r_e - r_n) \underline{d} \cdot \underline{E}(r_n) + \dots$$

$$= \delta(r) \underline{d} \cdot \underline{E}$$

↑
rain lid Homogen Anregung

Bewegungsgl.

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

Delta Funktion, kann durch Zerlegen in Intervalle und Integrale selbst werden.

Damit ist es z.B. möglich, die zeitliche Dynamik bei optischer Anregung zu berechnen.

Vorteil: in Vergleich zu Eigenwertproblem nur ψ (nicht jede Zeitsolver) muss in Speicher gehalten werden. Zeitsolver löst Eigenwertproblem für relevant Zustände implizit.

Nachteil: Eigenvektoren noch nicht bekannt.

Kann leicht auf Nichtlinear Optik erweitert werden (erfordert Biexzitonen, Rephasierung z.B. Quantenprojektion und Jump-Operatoren).

2. Beispiel: Blochgleichungen (z.B. in Dichtematrixtheorie)

$$H = H_0 + H_{el-L}$$

$$H_0 = \sum_n \epsilon_n |n\rangle\langle n|$$

↑
Energie

Vielteilchen-zustände

$$H_{el-L} = \sum_{nm} d_{nm} \underline{E}(t) |n\rangle\langle m|$$

↑
Dipolmoment

2. B. Zweiwaben System



$$H_0 = \hbar \varepsilon_1 |1\rangle \langle 1| + \hbar \varepsilon_2 |2\rangle \langle 2|$$

$$H_{\text{ee-L}} = \hbar d_{12} \cdot \underline{E}(\underline{r}, t) |1\rangle \langle 2| + \hbar d_{21} \cdot \underline{E}(\underline{r}, t) |2\rangle \langle 1|$$

Verwende Liouville von Nambu:

$$\partial_t \rho = -\frac{i}{\hbar} [H, \rho] \quad P \rightarrow d_{12} \underbrace{\langle 1| \rho |2\rangle}_{\rho_{12}}$$

Blockgleichungen:

$$\partial_t \rho_{12} = -i(\varepsilon_1 - \varepsilon_2) \rho_{12} - i E \cdot d_{21} (\rho_{12} - \rho_{11})$$

$$\begin{aligned} \partial_t \rho_{11} &= -i E \cdot d_{21} \rho_{21} + i E \cdot d_{12} \rho_{12} \\ &= 2 \operatorname{Im}(E \cdot d_{21} \rho_{11}) \end{aligned}$$

Beispiel für Nichtlinear Gleichung in $\underline{E}$, aber auch das ist in Prinzip kein Problem.

3. Beispiel: Halbleiter Bloch Gleichung



Freie Bewegung

$$H_0 = \sum_k \varepsilon_{v,k} a_{v,k}^\dagger a_{v,k} + \sum_k \varepsilon_{c,k} a_{c,k}^\dagger a_{c,k}$$

$$H_{\text{eh}} = \sum_k \underline{d}_{11} \cdot \underline{E} a_{v,k}^\dagger a_{c,k} + \text{h.c.} + \text{Coulomb}$$

$$H_{\text{Coulomb}} = \sum_{\substack{l, l' \\ q \neq 0 \\ l \neq l'}} V_q a_{l, l+q}^\dagger a_{l', l-q}^\dagger a_{l', l} a_{l, l-q}$$



(Interband wie Interband Prozess)

$$P(t) = \sum_n P_n d_n \quad P_n = \langle a_{cl}^\dagger a_{vl} \rangle$$

$$\Rightarrow \text{Halbleiterblockglieder (im kohärenten Limit, Heisenberg)} \quad \begin{aligned} f_n^e &= \langle a_{cl}^\dagger a_{cl} \rangle \neq f_{ll}^h = \langle a_{vl}^\dagger a_{vl} \rangle \end{aligned}$$

$$\begin{aligned} \partial_t P_{lk} = & -\frac{i}{\hbar} (\epsilon_{v,l} - \epsilon_{c,l} + i\gamma_{vl}) P_{lk} + \frac{i}{\hbar} \left(\sum_{q \neq 0} V_q (f_{c,l+q} - f_{v,l-q}) \right) P_{lk} \\ & + \frac{i}{\hbar} \left(d_{vl} E(t) + \sum_{q \neq 0} V_q P_{l+q} \right) (n_{c,l} - n_{v,l}) \end{aligned}$$



Mittelmoment.

Stromterm 2. B.

$$n_1 n_2 (1 - n_3) (1 - n_4)$$

$\Rightarrow$ Mittelwert.