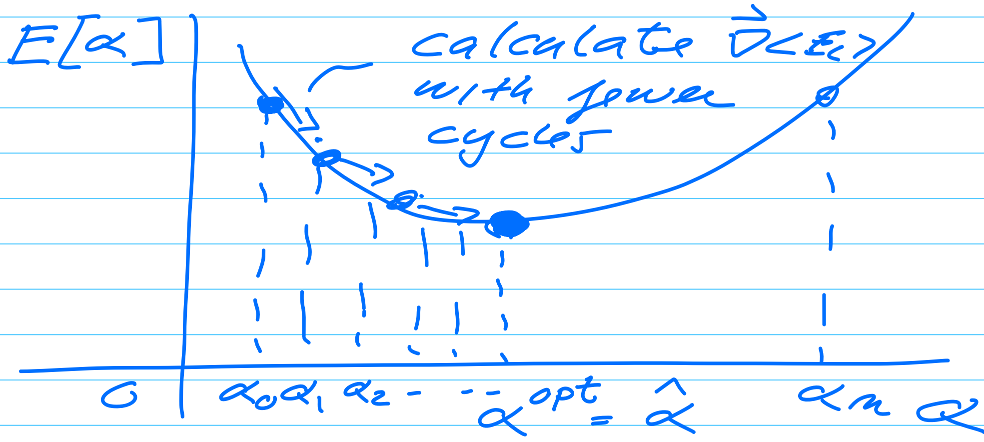
FYS 4411/9411 FEB 23, 2023

optimization



$$\mathbb{E}[E_L(\alpha)] = \langle E_L(\alpha)\rangle = \int_{x\in D} dx \times P(x;\alpha)E_L(x;\alpha)$$

$$\frac{d}{d\alpha}\langle E_L(\alpha)\rangle = 0$$

$$\hat{\alpha} = \arg\min_{\alpha\in\mathbb{R}^m}\langle E_L(\alpha)\rangle$$

(in general $\vec{\nabla}\langle E_L(\vec{\alpha})\rangle = 0$)

what do we have to code?

(i) $\vec{\nabla}\langle E_L(\alpha)\rangle$

(ii) Gradient descent method
- Newton-Raphson
- gradient descent (GD)
- stochastic GD

- steepest descent
- Conjugate GD
- Pseudo Newton methods

Project 1 without two-body interaction

1 particle in 1-Dim

$$\psi_T(x;\alpha) = e^{-1/2\,\alpha^2 x^2}$$

$$E_L(x;\alpha) = \frac{1}{\psi_T} H \psi_T$$

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2} k x^2$$

$$\hbar = m = c = e = \underline{1}$$

$$\omega_0 = \sqrt{k/m}$$

$$E_{n_x} = \hbar\omega_0 (n_x + 1/2)$$

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2} k x^2\right]\psi = E\psi$$

$$k = m \cdot \omega_0^2$$

Dimless $\rho = x/\alpha$

$$\left[-\frac{\hbar^2}{2m\alpha^2}\frac{d^2}{d\rho^2} + \frac{1}{2} m \omega_0^2 \alpha^2 \rho^2\right]\psi =$$

$$E\psi$$

$$\times \quad \frac{m\alpha^2}{\hbar^2}$$

$$\left[-\frac{1}{2}\frac{d^2}{d\xi^2} + \frac{1}{2}\underbrace{\frac{m^2\omega_0^2\alpha^4}{\hbar^2}}_{=1}\xi^2 \right]\psi$$

$$= E'\psi$$

$$\alpha = \sqrt{\frac{\hbar}{m\omega_0}}$$

$$[\alpha] = \text{length}$$

$$\frac{\hbar \cdot \hbar c^2}{\underbrace{mc^2}_{\text{energy}}\underbrace{\hbar\omega_0}_{\text{energy}}} \quad \text{energy}^2 \times \text{length}^2$$

$$\omega_0 = 1 = \hbar = m = c = e$$

$$E_{n_x} = n_x + 1/2$$

$$\psi_T(x;\alpha) = e^{-\frac{1}{2}\alpha^2 x^2}$$

$$P(x;\alpha) = \frac{e^{-\alpha^2 x^2}}{\int dx\, e^{-\alpha^2 x^2}}$$

$$E_L = \frac{1}{\psi_T}\hat{H}\psi_T$$

$$E_L(x;\alpha) = \frac{1}{2}\left[\alpha^2 + x^2(1-\alpha^4)\right]$$

$$\langle E_L(\alpha)\rangle = \int dx\, P(x;\alpha) E_L(x;\alpha)$$

$$= \frac{1}{4}\left[\alpha^2 + \frac{1}{\alpha^3}\right]$$

$$\frac{d\langle E_L\rangle}{d\alpha} = 0 = \frac{1}{2}\alpha - \frac{1}{2\alpha^3} = 0$$

$$\Rightarrow \underline{\alpha = 1}$$

$$\frac{d^2\langle E_L\rangle}{d\alpha^2} = \frac{1}{2} + \frac{3}{2\alpha^4} > 0$$

$\alpha > 0$, convex optimization.

our problem

$$\langle E_L(\vec{\alpha})\rangle = \int_{\vec{R}\in D} d\vec{R}\, P(\vec{R};\vec{\alpha}) E_L(\vec{R};\vec{\alpha})$$

$\vec{D}_{\vec{\alpha}}\langle E_L(\vec{\alpha})\rangle$, specialize to our project 1 case

1-Dim:

$$\frac{d\langle E_L\rangle}{d\alpha} = \frac{d}{d\alpha}\left[\int dx\, P(x;\alpha) E_L(x;\alpha)\right]$$

$$= \int dx \left[\frac{dP}{d\alpha} E_L + P \frac{dE_L}{d\alpha} \right]$$

$$P(x;\alpha) = \frac{e^{-\alpha^2 x^2}}{\underbrace{\int dx\, e^{-\alpha^2 x^2}}_{N}}$$

$$\int \frac{dP}{d\alpha} E_L \, dx$$

$$= \int dx \left[\left(\frac{-2\alpha^2 x\, e^{-\alpha^2 x^2}}{N} \right) E_L(x;\alpha) \right.$$

$$\left. - \frac{e^{-\alpha^2 x^2} E_L}{N} \int dx' \left(\frac{-2\alpha^2 x' e^{-\alpha^2 x'^2}}{N} \right) \right]$$

$P(x;\alpha)\, E_L(x;\alpha)$

$$\left\langle \left(d\psi_T / d\alpha \right) / \psi_T \right\rangle$$

$$-\alpha^2 x' = \left(\frac{d\psi_T}{d\alpha} \right) / \psi_T$$

$$= \frac{d\, e^{-\frac{1}{2}\alpha^2 x^2}}{d\alpha} = \alpha^2 x\, \cancel{e^{-\alpha^2 x^2/2}}$$

$$2\left[\left\langle \left(\frac{d\psi_T}{d\alpha}\Big/\psi_T\right) E_L \right\rangle \right.$$

$$\left. - \left\langle \frac{d\psi_T}{d\alpha}\Big/\psi_T \right\rangle \langle E_L \rangle \right]$$

3 - integrals

$$\left[d\psi_T/d\alpha\right]/\psi_T = \frac{d \ln \psi_T}{d\alpha}$$

$$\psi_T \sim \exp(f) \Rightarrow$$

$$df/d\alpha$$

Optimization methods (aka gradient methods)

$$\vec{\nabla}_{\vec{\alpha}} \langle E_L(\vec{\alpha}) \rangle = 0$$

Taylor-expand $\langle E_L(\vec{\alpha}) \rangle = \bar{E}(\hat{\alpha})$

$$\bar{E}(\hat{\alpha}) = \bar{E}(\alpha^{(n)}) + (\hat{\alpha} - \alpha^{(n)})^T g^{(n)}$$

(skip $\vec{\alpha} \Rightarrow \alpha$)

$$+ \frac{1}{2}(\hat{\alpha} - \alpha^{(n)})^T H^{(n)} (\hat{\alpha} - \alpha^{(n)})$$

$$+ O\left((\alpha - \alpha^{(n)})^3\right)$$

$$g^{(n)} = \vec{\nabla}_\alpha \left(\langle E_L(\alpha^{(n)}) \rangle \right)$$
$$= \vec{\nabla}_\alpha \bar{E}(\alpha^{(n)})$$

H is the Hessian matrix which contains a matrix elements the second derivative $\frac{\partial^2 \bar{E}(\alpha^{(n)})}{\partial \alpha_i \partial \alpha_j} = H_{ij}$

In our special case: one α

$$\bar{E}(\hat{\alpha}) = \bar{E}(\alpha^{(n)}) + (\hat{\alpha} - \alpha^{(n)}) \frac{d\bar{E}}{d\alpha}$$
$$+ \frac{1}{2} (\hat{\alpha} - \alpha^{(n)})^2 \frac{d^2 \bar{E}}{d\alpha^2} \Big|_{\alpha = \alpha^{(n)}}$$
$$+ O\left((\hat{\alpha} - \alpha^{(n)})^3 \right)$$

$$\bar{E}(\hat{\alpha}) \simeq \bar{E}(\alpha^{(n)}) + (\hat{\alpha} - \alpha^{(n)}) \frac{d\bar{E}}{d\alpha}$$
$$+ \frac{1}{2} (\hat{\alpha} - \alpha^{(n)})^2 \frac{d^2 \bar{E}}{d\alpha^2}$$

The general form is that of

$$f(x) = \overset{\text{constant}}{c} + x^T \cdot g + \frac{1}{2} x^T A x$$

$$\frac{\partial f}{\partial x} = 0 = A \cdot x + g \Rightarrow$$

$$A \cdot x = -g \Rightarrow$$

$$x = -A^{-1} g$$

$A = H$ (Hessian matrix)

$$x = \hat{\alpha} - \alpha^{(n)}$$

$$\hat{\alpha} - \alpha^{(n)} = -H^{-1}(\alpha^{(n)}) \cdot g(\alpha^{(n)})$$

$$\boxed{\hat{\alpha} = \alpha^{(n)} - H^{-1}(\alpha^{(n)})\, g(\alpha^{(n)})}$$

$\vec{\nabla}_{\alpha^{(n)}} E(\alpha^{(n)})$

Newton-Raphson
for $f(x) = 0 = \nabla_x E(x)$

in MC calculations, $H(\alpha^{(n)})$ involves multidimensional integrals which are evaluated stochastically.

Simplest method

$H(\alpha^{(n)}) \Rightarrow \gamma$ constant

$\gamma \in \{10^{-4}, 10^{-3}, 10^{-2}\}$

Gradient descent

$$\hat{\alpha} = \alpha^{(n)} - \gamma \vec{\nabla}_\alpha \bar{E}(\alpha^{(n)})$$

Quasi-Newton : Broyden's algo.

Gradient descent with Momentum

Stochastic gradient descent

Stochastic reconfiguration