FYS-STK 3155/4155, Sept 30

$$\beta^{(n+1)} = \beta^{(n)} - \gamma^{(n)} g(\beta^{(n)})$$

$$C(\beta^{(n+1)}) = C(\beta^{(n)}) + (\beta^{(n+1)} - \beta^{(n)})^T g(\beta^{(n)}) + \frac{1}{2}(\beta^{(n+1)} - \beta^{(n)})^T H(\beta^{(n)}) \times (\beta^{(n+1)} - \beta^{(n)}) + \ldots$$

$$C(\beta) = \frac{1}{n}(y - X\beta)^T (y - X\beta)$$

$$H = \frac{2}{n} X^T X$$

$$b = \beta^{(n+1)} - \beta^{(n)}$$

$$C(\beta^{(n+1)}) = C_0 + b^T g(\beta^{(n)}) + \frac{1}{2} b^T H b$$

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

$$\frac{\partial f(x)}{\partial x} = 0 = Ax - b$$

$$Ax = b$$

$$H = X^T X \in \mathbb{R}^{p \times p}$$

square & symmetric

positive definite matrix

$$\beta^{(n+1)} = \beta^{(n)} - \underbrace{H^{-1}(\beta^{(n)})}_{\gamma^{(n)}} g(\beta^{(n)})$$

- learning rate updates
  - linear update
  - constant
  - exponential update
  - momentum based

$$\beta^{(m+1)} = \beta^{(m)} + \gamma g(\beta^{(m)}) + \delta(\beta^{(m)} - \beta^{(m-1)})$$

- Adagrad (convex function)
- RMS prop (non-convex)
- Adam
- Full gradient calculation
- stochastic GD

Steepest descent

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

$$\frac{\partial f}{\partial x} = 0 = b - Ax = -g$$

$$\boxed{x_{k+1} = x_k + \alpha_k r_k}$$

r_{k+1} = residual

$$r_{k+1} = b - Ax_{k+1}$$

$$r_k = b - Ax_k$$

$$r_0 = b - Ax_0$$

$$r_{k+1}^T r_k = 0$$

$$r_{k+1} = b - Ax_{k+1}$$

$$= \underbrace{b - (Ax_k}_{r_k} + A\alpha_k r_k)$$

$$r_k^T r_{k+1} = 0 = r_k^T r_k + \alpha_k r_k^T A r_k$$

$$\Rightarrow \boxed{\alpha_k = \frac{r_k^T r_k}{r_k^T A r_k}}$$

$$r_k = b - Ax_k = -g_k$$

$$x_{k+1} = x_k + \alpha_k r_k$$

$$= x_k - \underbrace{\alpha_k}_{\gamma_k} g_k$$

$$\| \qquad \gamma_k$$

$$\gamma_k = \frac{g_k^T g_k}{g_k^T H g_k}$$

$$r_k = -g_k \qquad A = H$$

if $\quad H g_k = \lambda_k g_k$

$$\gamma_k = \frac{1}{\lambda_k}$$

AdaGrad $\quad \gamma_k \sim \frac{1}{g_k^T g_k}$

— Schedulers for γ_k

- constant γ_0
- linear

$$\gamma_k = (1-\alpha)\gamma_0 + \alpha \gamma_\tau$$

$$\gamma_\tau \sim \frac{1}{100}\gamma_0$$

$$\alpha = \frac{k}{\tau}$$

k — # iterations

τ — example in notebook

- $\gamma_k = \dfrac{\gamma_0}{1 + k\gamma_\tau}$

- exponential decay

$\gamma_k = \gamma_0 \exp(-k\gamma_\tau)$

- Adagrad

algorithm:
require initial γ_0
— " — β_0
adagrad with SGD
batches # epochs
$D = \{(x_0 y_0), (x_1 y_1) \ldots (x_{n-1}, y_{n-1})\}$

while stopping criterion not met

- compute gradient g_k
- Define $G = \sum_{i=1}^{k} g_i g_i^T$

- Define $\dfrac{\gamma_0}{\delta+\sqrt{G_{ii}}} = \gamma_K$

simple approach:
only diagonal elements
in $\sqrt{G_{ii}}$

- update

$$\beta_{K+1} = \beta_K - \begin{cases} 10^{-8} \sim 10^{-6} \\ \dfrac{\gamma_0}{\delta+\sqrt{G_{ii}}} \odot g_K \end{cases}$$

$$x \odot y = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \odot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 y_1 \\ x_2 y_2 \end{bmatrix}$$

end update,

RMSprop

require γ_0, β_0
decay rate ρ, $\delta \sim 10^{-8}$
while stopping criterion
not met

compute g_k

$$G_k = \rho G_{k-1} + (1-\rho)\sum_{i=0}^{k} g_i g_i^T$$

$$\gamma_k = \frac{\gamma_0}{\delta + \sqrt{G_k}}$$

diagonal terms only

$$\beta_{k+1} = \beta_k - \gamma_k \odot g$$

FYS4480, october 6, 2022

Hartree-Fock theory:

$$h^{HF} = \sum_{i=1}^{N} -\frac{\nabla_i^2}{2m} + \sum_{i=1}^{N} u_{ext}(x_i)$$

$$+ \sum_{j \le F} \langle j | u^{HF} | j \rangle$$

relation to FCI theory:

$$H_{ij} = \langle \Phi_i | H | \Phi_j \rangle$$

$$
\begin{array}{c|c}
 & \begin{array}{ccccc} 0p0h & 1p1h & 2p2h & \cdots & NpNh \end{array} \\
\begin{array}{c} 0p0h \\ 1p1h \\ 2p2h \\ \vdots \\ \vdots \\ NpNh \end{array} &
\begin{bmatrix}
x & x & x & 0 & \cdots & \\
x & x & x & x & 0 & \cdots \\
x & x & & & & \\
0 & x & & & & \\
\vdots & 0 & & & & \\
0 & \vdots & & & &
\end{bmatrix}
\end{array}
$$

Approximation to Full CI

Keep 0p0h and $\underbrace{1p1h}_{\text{single excitation}}$

CIS = CI with singles

Keep 0p0h, 1p1h, $\underbrace{2p2h}_{\text{doubles}}$

CISD

$\boxed{\text{CIS}}$

$$
H = \begin{array}{c|c}
 & \begin{array}{cc} 0p0h & 1p1h \end{array} \\
\begin{array}{c} 0p0h \\ 1p1h \end{array} &
\begin{bmatrix} x & x \\ x & x \end{bmatrix}
\end{array}
$$

$$\langle \Phi_0 | H | \Phi_0 \rangle = E_0^{\text{Ref.}}$$

$$\langle \Phi_0 | H | \Phi_i^a \rangle = \langle i | \hat{f} | a \rangle$$

$$= \langle i | \hat{h}_0 | a \rangle + \sum_{j \le F} \langle ij | \hat{v} | aj \rangle_{AS}$$

$$u^{\dagger} H u = D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$\langle i | \hat{f} | a \rangle = 0$$

solution of HF equations

- coordinate space HF
- variations of coefficient

<u>coordinate space HF</u>

$$\langle \delta \Phi_0 | H | \Phi_0 \rangle = 0$$

$$D = u_M^{\dagger} \cdots u_1^{\dagger} H u_1 u_2 \cdots u_M$$

$$\langle \Phi_0 | H | \Phi_0 \rangle =$$

$$\sum_{i=1}^{N} \int dx_i\, \varphi^*_{\alpha_i}(x_i) \{ \underbrace{\hat{t}_i + u_{ext}(x_i)}_{\hat{h}_0(x_i)} \}$$

$$\times \cdot \varphi_{\alpha_i}(x_i)$$

$$\left(\underbrace{\int dx_i\, \varphi^*_{\alpha_i} (\hat{h}_0 (x_i)] \varphi_{\alpha_i}(x_i)}_{\langle \alpha_i | \hat{h}_0 | \alpha_i \rangle} \right)$$

$$+ \frac{1}{2} \sum_{ij}^{N} \underbrace{\int dx_i\, dx_j\, (\varphi^*_{\alpha_i}(x_i) \varphi^*_{\alpha_j}(x_j)) \times v(|x_i - x_j|)\, \varphi_{\alpha_i}(x_i)\, \varphi_{\alpha_j}(x_j)}_{\langle \alpha_i \alpha_j | v | \alpha_i \alpha_j \rangle}$$

(Direct element)

$$- \frac{1}{2} \sum_{ij}^{N} \underbrace{\int dx_i\, dx_j\, (\varphi^*_{\alpha_i}(x_i) \varphi^*_{\alpha_j}(x_j) \times v(|x_i - x_j|)\, \varphi_{\alpha_i}(x_j)\, \varphi_{\alpha_j}(x_i))}_{\langle \alpha_i \alpha_j | \hat{v} | \alpha_j \alpha_i \rangle}$$

exchange term.

$$\delta \Phi_0 \qquad \Phi_0 = \frac{1}{\sqrt{N!}} \sum_P (-)^P \hat{P}\, \varphi_{\alpha_1}(x_1)\, \varphi_{\alpha_2}(x_2) \cdots \varphi_{\alpha_N}(x_N)$$

$$\langle \delta\Phi_0 | H | \Phi_0 \rangle + \langle \Phi_0 | H | \delta\Phi_0 \rangle$$

$$= 0$$

$$\langle \delta\Phi_0 | H | \Phi_0 \rangle = 0$$

$$= \sum_{i=1}^{N} \int dx_i \, \delta\varphi^*_{\alpha_i}(x_i) \, \hat{h}_0(x_i) \, \varphi_{\alpha_i}(x_i)$$

$$+ \frac{1}{2} \sum_{i,j} \int dx_i \, dx_j \; \delta\varphi^*_{\alpha_i}(x_i) \, \varphi^*_{\alpha_j}(x_j)$$

$$\times v(|x_i - x_j|) \, \varphi_{\alpha_i}(x_i) \, \varphi_{\alpha_j}(x_j)$$

$$+ \frac{1}{2} \sum_{i,j} \int dx_i \, dx_j \; \varphi^*_{\alpha_i}(x_i) \, \delta\varphi_{\alpha_j}(x_j)$$

$$\times v(|x_i - x_j|) \, \varphi_{\alpha_i}(x_i) \, \varphi_{\alpha_j}(x_j)$$

$$i \leftrightarrow j$$

$-$ Exchange term

$$\sum_{i} \int dx_i \, \delta\varphi^*_{\alpha_i}(x_i) \, \hat{h}_0(x_i) \, \varphi_{\alpha_i}(x_i)$$

$$+ \sum_{i,j} \int dx_i \, dx_j \, (\delta\varphi_{\alpha_i}(x_i))$$

$$\times \varphi_{\alpha_j}(x_j)\, v\,(\,(x_i - x_j)\,)\Big)$$

$$\times \Big(\varphi_{\alpha_i}(x_i)\,\varphi_{\alpha_j}(x_j) - \varphi_{\alpha_i}(x_j)\,\varphi_{\alpha_j}(x_i)\Big)$$

constraint

$$\int d\tau\; \Phi_0^* \Phi_0 = \underline{1}$$

Lagrangian multiplier λ

specific term $\delta\varphi^*_{\alpha_i}(x_i)$

$(\alpha_1\, \alpha_2 \,--\, \alpha_N)$

$$\Rightarrow \quad \sum_i \int dx_i\, \delta\varphi^*_{\alpha_i}(x_i)\, \underline{\hat{h}^{HF}\varphi_{\alpha_i}(x_i)}$$

$$- \sum_i \int dx_i\, \delta\varphi^*_{\alpha_i}(x_i)\, \underline{\varphi_{\alpha_i}(x_i)}$$

$$\times\ \underline{\lambda_i}$$

$$= 0$$

$$\boxed{\hat{h}^{HF}\varphi_{\alpha_i}(x_i) = \lambda_i\, \varphi_{\alpha_i}}$$

$$\hat{h}^{HF} \varphi_{\alpha_i}(x_i) = \hat{h}_0 \varphi_{\alpha_i}(x_i)$$

$$\left. \begin{aligned} &+ \sum_j \int dx_j \, \varphi_{\alpha_j}(x_j) \, v(|x_i - x_j|) \\ &\quad \times \varphi_{\alpha_i}(x_i) \, \varphi_{\alpha_j}(x_j) \end{aligned} \right\} \text{Hartree}$$

$$- \sum_j \int dx_j \, \varphi_{\alpha_j}(x_j) \, v(|x_i - x_j|) \times \varphi_{\alpha_i}(x_j) \, \varphi_{\alpha_j}(x_i)$$

Fock term.

Hartree term:

$$\hat{V}_i^H = \sum_j \int dx_j \, \varphi^*_{\alpha_j}(x_j) \, v(|x_i - x_j|) \times \varphi_{\alpha_j}(x_j)$$

$$\hat{V}_i^H \varphi_{\alpha_i}(x_i)$$

$$\hat{V}_i^F = \sum_j \int dx_j \left[\varphi^*_{\alpha_j}(x_j) \, v(|x_i - x_j|) \times \varphi_{\alpha_i}(x_j) \right] \varphi_{\alpha_j}(x_i)$$

Coefficient variation

Computational sp basis φ_λ

$$\hat{h}_0 \varphi_\lambda = \varepsilon_\lambda \varphi_\lambda \qquad \langle \varphi_\lambda | \varphi_\delta \rangle = \delta_{\lambda\delta}$$

$$\varphi_p^{HF} = \sum_\lambda C_{\lambda p} | \varphi_\lambda \rangle$$

$$C_{\lambda p} = \langle \varphi_p^{HF} | \varphi_\lambda \rangle$$

$$\Phi_0 = \frac{1}{\sqrt{N!}} \sum_P (-)^P \hat{P} \varphi_{\lambda_1} \varphi_{\lambda_2} \cdots \varphi_{\lambda_N}$$

$$\langle \Phi_0 | H | \Phi_0 \rangle = E_0^{Ref.}$$

$$\Phi_0^{HF} = \frac{1}{\sqrt{N!}} \sum_P (-)^P \hat{P} \varphi_{p_1} \varphi_{p_2} \cdots \varphi_{p_N}$$

$$\langle \Phi_0^{HF} | H | \Phi_0^{HF} \rangle \leq E_0^{Ref}$$

$$\langle \Phi_0^{HF} | H | \Phi_0^{HF} \rangle = E_0^{HF} \geq E$$