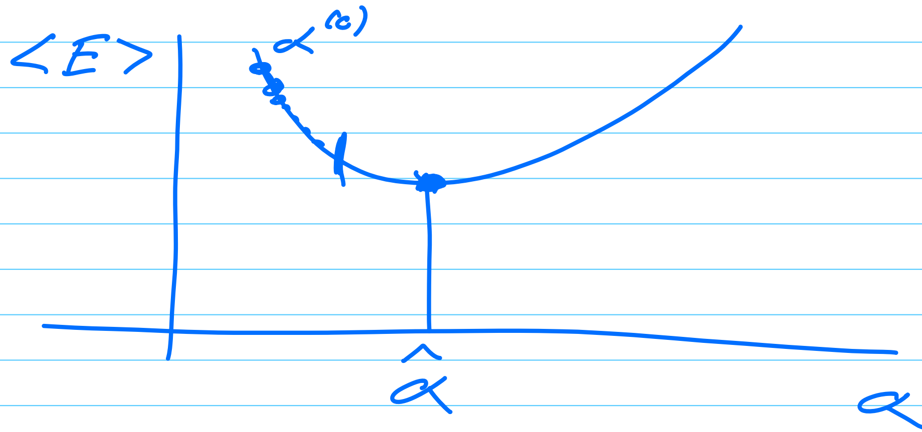
FYS4411/9411, March 9, 2023

One parameter α, optimal value $\hat{\alpha}$



$\langle E \rangle \rightarrow E(\alpha)$, Taylor expand around $\hat{\alpha}$, $\alpha^{(n)}$

$$E(\hat{\alpha}) = E(\alpha^{(n)}) + (\hat{\alpha}-\alpha^{(n)}) \times \left.\frac{dE}{d\alpha}\right|_{\alpha^{(n)}} + \frac{(\hat{\alpha}-\alpha^{(n)})^2}{2} \left.\frac{d^2E}{d\alpha^2}\right|_{\alpha^{(n)}} + O\left((\hat{\alpha}-\alpha^{(n)})^3\right)$$

$$E(\hat{\alpha}) = E(\alpha^{(n)}) + (\hat{\alpha}-\alpha^{(n)})\frac{dE}{d\alpha} + \frac{(\hat{\alpha}-\alpha^{(n)})^2}{2}\frac{d^2E}{d\alpha^2}$$

$$\frac{dE(\hat{\alpha})}{d\alpha} = 0$$

Simplify $E(\alpha) \to f(x)$

$$f(x) = C + x\frac{df}{dx} + \frac{1}{2}x^2\frac{d^2f}{dx^2}$$

$$\frac{df}{dx} = 0 \implies$$

$$x = -\frac{df}{dx} \Big/ \frac{d^2f}{dx^2}$$

if we have more than one α, $\frac{df}{dx} \to \nabla_x f = g$

$$\frac{d^2f}{dx^2} \to H$$

$$h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$f(x) = C + g^T x + \frac{1}{2}x^T H x$$

$$\frac{\partial f}{\partial x} = 0 = Hx + g \implies$$

$$x = -H^{-1}g \qquad g(x^{(n)})$$

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

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

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

$$g(\alpha^{(n)}) = \nabla_\alpha E(\alpha^{(n)})$$

$$\nabla_\alpha E(\alpha) = 2\left(\left\langle \frac{d \ln \psi(\alpha)}{d\alpha} E_L(\alpha) \right\rangle - \left\langle \frac{d \ln \psi(\alpha)}{d\alpha} \right\rangle \langle E_L(\alpha) \rangle \right)$$

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

replace H with a constant

$$\hat{\alpha} = \alpha^{(n)} - \gamma g \quad \Big| \text{ gradient descent}$$

$$E(\alpha^{(n)} - \gamma g(\alpha^{(n)})) =$$

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

$$+ O(\gamma^3)$$

optimal γ is obtained

$$\frac{dE}{d\gamma} = -g^T g + \gamma g^T H g \Rightarrow$$

$$\gamma = \frac{g^T g}{g^T H g}$$

$g^T H g > g^T g$, we may encounter problems with the Taylor expansion,

assume $\quad H g = \lambda g$

$$\gamma = \frac{1}{\lambda} = \frac{\cancel{g^T g}}{\lambda \cancel{g^T g}}$$

$$\gamma = \frac{1}{\lambda_{max}} \quad \text{smallest}$$

$$\gamma = \frac{1}{\lambda_{min}} \quad \text{largest.}$$

Simplest approach

iterative approach with various γ-values

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

Gradient descent.

<u>steepest descent</u>

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

$$\frac{\partial f}{\partial x} = 0 = Hx - b \Rightarrow$$

$$Hx = b$$

Define residual

$$r = b - Hx$$

when we have x $r = 0$

start with a guess x_0

$$r_0 = -Hx_0 + b$$

in general

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

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

$$r_{k+1} = b - H(x_k + \alpha_k r_k)$$

$$= (\underbrace{b - Hx_k}_{r_k}) - \alpha_k H r_k$$

$$r_{k+1} = r_k - \alpha_k H r_k$$

we want $r_{k+1} = 0$

$$r_k = \alpha_k H r_k \Rightarrow |r_k^T$$

$$\frac{r_k^T r_k}{r_k^T H r_k} = \alpha_k$$

$$\frac{\partial f(x)}{\partial x} = 0 = \underbrace{Hx - b} \Rightarrow$$

$$x_{k+1} = x_k - \alpha_k g(x_k)$$

Speed up GD

GD with momentum

$$m \frac{d^2x}{dt^2} + \mu \frac{dx}{dt} = -\nabla V(x) = F$$

discretize 2nd & 1st derivative

$$\frac{d^2x}{dt^2} \approx \frac{x(t+\Delta t) + x(t-\Delta t) - 2x(t)}{(\Delta t)^2}$$

$$\frac{dx}{dt} \approx \frac{x(t+\Delta t) - x(t)}{\Delta t}$$

$$x(t \pm \Delta t) = x_{t \pm \Delta t} \qquad x(t) = x_t$$

$$m\frac{\left(x_{t+\Delta t} + x_{t-\Delta t} - 2x_t\right)}{(\Delta t)^2} + \mu \frac{x_{t+\Delta t} - x_t}{\Delta t} = -\nabla V(x)$$

Define : $\Delta x_{t+\Delta t} = x_{t+\Delta t} - x_t$

$$\Delta x_t = x_t - x_{t-\Delta t}$$

$$\Delta x_{t+\Delta t} = \underbrace{-\frac{(\Delta t)^2}{m + \mu \Delta t}}_{\gamma} \nabla V(x) + \underbrace{\frac{m}{m + \mu \Delta t}}_{\delta} \Delta x_t$$

$$\Delta x_{t+\Delta t} = -\gamma \nabla V(x) + \delta \Delta x_t$$

$$x_{t+\Delta t} = x_t - \gamma \nabla V(x)$$

$$\left(\alpha^{(n+1)} = \alpha^{(n)} - \gamma\, g(\alpha^{(n)})\right)$$

$X_t \Rightarrow \alpha^{(n)}$ $\quad X_{t+\Delta t} \Rightarrow \alpha^{(n+1)}$

$DV(x) \Rightarrow g(\alpha^{(n)})$

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

δ can be interpreted as a memory parameter.

Convergence criterion

- fix max iterations
- stop if reached max iterations
- or stop $|\alpha^{(n+1)} - \alpha^{(n)}| \leq \varepsilon$