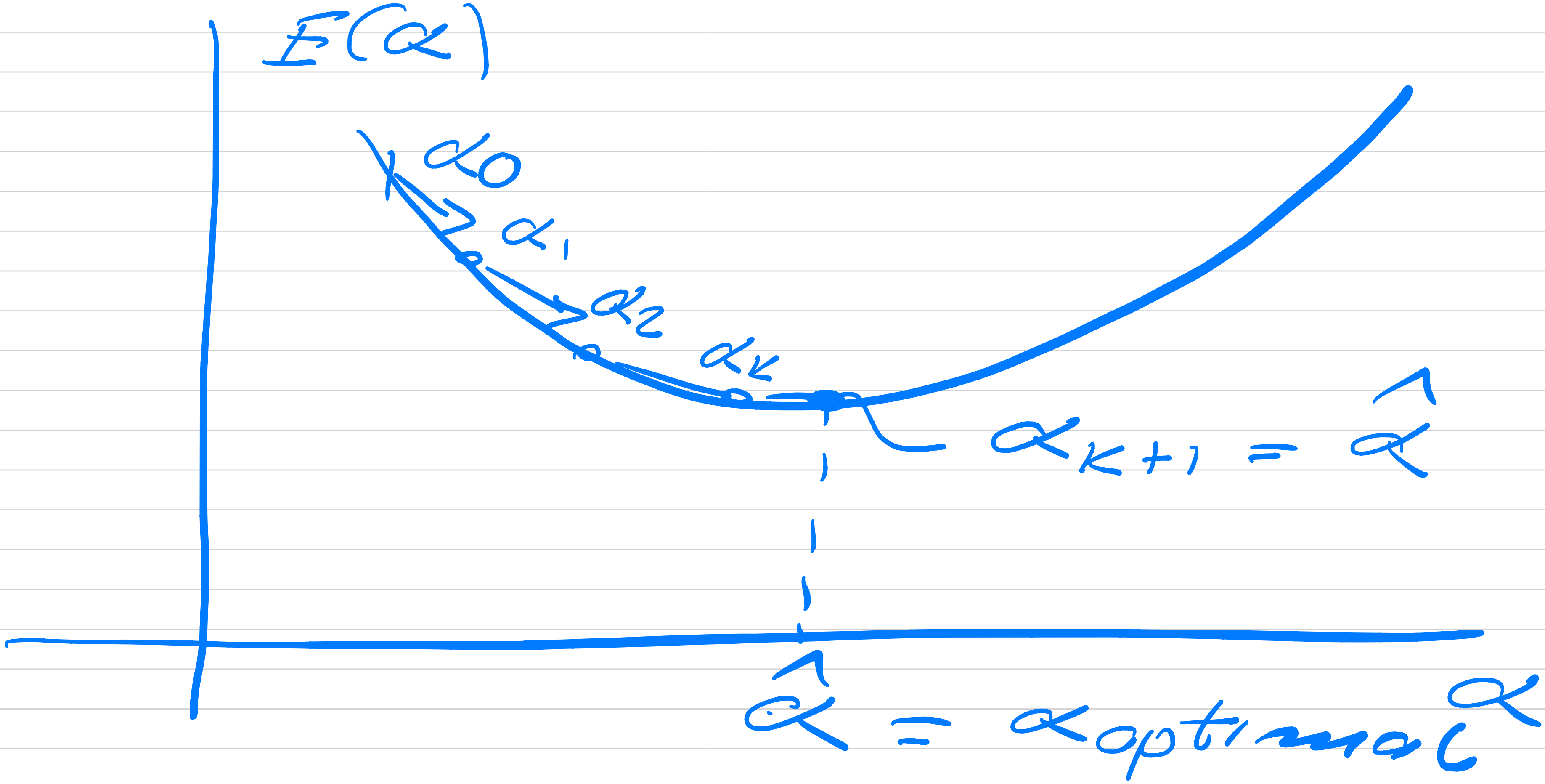
FYS4411/9411, lecture February 28, 2025

FYS4411/9411 Feb 28, 2025

$$\mathbb{E}[E_L(\hat{\alpha})] = \langle E_L(\hat{\alpha}) \rangle = E(\hat{\alpha})$$



$$E(\hat{\alpha}) = E(\alpha_k) + (\nabla E(\alpha_k))^T(\hat{\alpha}-\alpha_k)$$

$\hat{\alpha} = \alpha_{k+1}$

$$\left(\frac{dE}{d\alpha}\bigg|_{\alpha=\alpha_k}\right)(\hat{\alpha}-\alpha_k)$$

$$+\frac{1}{2}(\hat{\alpha}-\alpha_k)^T\nabla^2 E(\alpha_k)(\hat{\alpha}-\alpha_k)$$

$$+ O\left(\|\hat{\alpha}-\alpha_k\|_2^{3/2}\right) + \ldots$$

Truncate at the level of $\nabla^2 E$

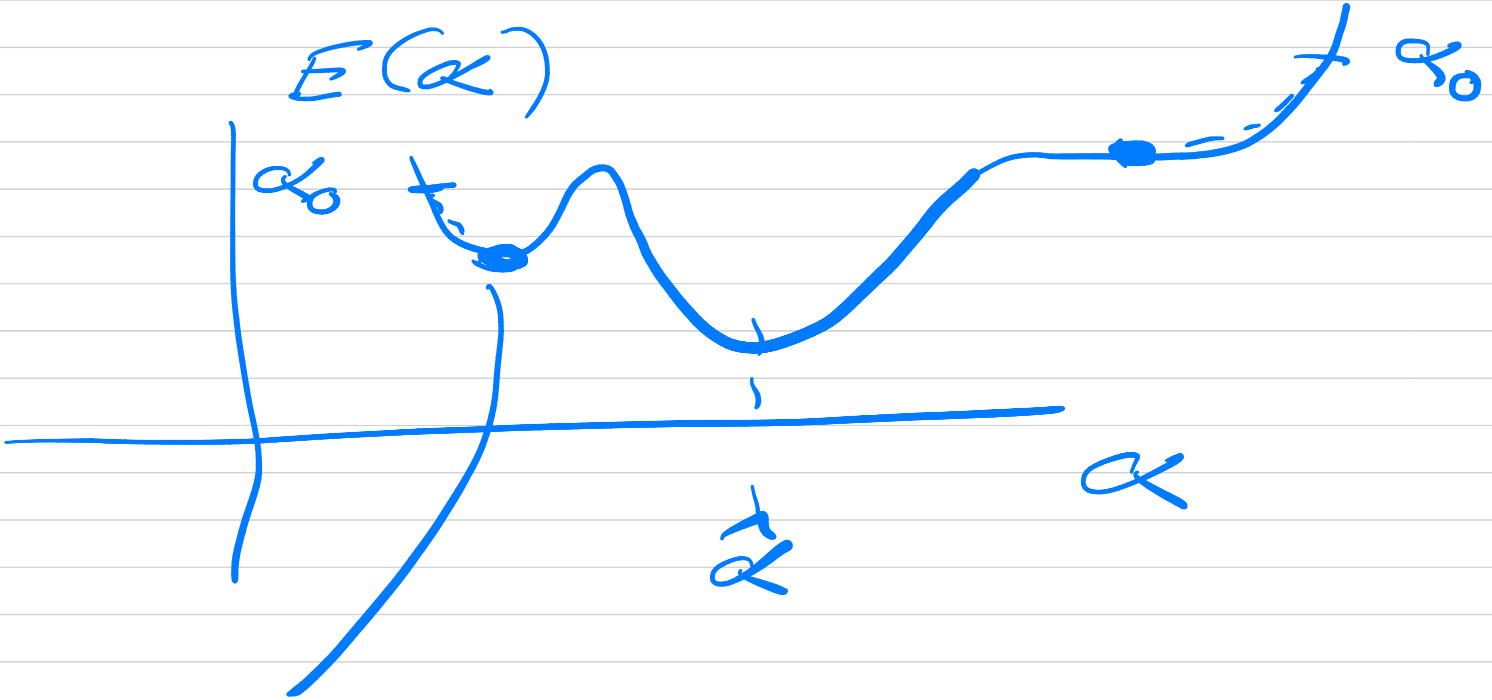
$$\alpha_{k+1} = \hat{\alpha} = \alpha_k - \underbrace{\left(\nabla^2 E(\alpha_k)\right)}_{\text{Hessian matrix}}{}^{-1}\nabla E(\alpha_k)$$

Hessian matrix

Problems

- we need $\nabla^2 E(\alpha_k)$
- sensitive to initial condition

α_0



$\nabla E(\alpha) = 0$

- Delicate dependence on the evaluation of $\nabla E(\alpha)$

First simple case: plain gradient descent (GD)

$$\alpha_{k+1} = \alpha_t - \gamma_k \nabla E(\alpha_k) \sim g(\alpha_k)$$

γ_k = learning rate

- k-dependent value
  * in steepest GD

$$\gamma_k = \frac{g^T(\alpha_k)\, g(\alpha_k)}{g^T(\alpha_k)\, \nabla^2 E\, g(\alpha_k)}$$

- constant $\gamma_k \Rightarrow \gamma$

$$\gamma = [10^{-4}, 10^{-3}, \dots 10^{-1}]$$

- optimizers
  * Momentum
  * RMSprop
  * Adagrad
  * ADAM
- conjugate GD
- Quasi-Newton methods -
  * Broyden's method
  * BFGS (scipy)
  * Powell

estimate of Hessian matrix

algorithm

- initialize γ (+ method for update)
- Fix # of iterations
- need initial guess α_0

while (# iteration $\leq$ Max OR
$\nabla E > \varepsilon$)

calculate $\boxed{\nabla E(\alpha_i)}$ Need this

update

$i = i+1$

$\alpha_{i+1} = \alpha_i - \gamma \nabla E(\alpha_i)$

end while

replace with scipy.optimize

$\nabla E(\alpha)$

simple example $\quad \underbrace{e^{-\frac{1}{2}\alpha^2 x^2}}_{\text{in general complex}} =$

$\psi_T(x;\alpha)$

$$E(\alpha) = \frac{\int dx \, \psi_T^*(x;\alpha) H \psi_T(x;\alpha)}{\int dx \, |\psi_T(x;\alpha)|^2}$$

one variable α

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

$$2\int dx\, \frac{d\psi_\alpha}{d\alpha}\, \psi_\alpha^* H \psi_\alpha$$

$$\cancel{\left[\int dx\, \psi_\alpha^2\right]^2}$$

$$\psi_T(x;\alpha) \to \psi_\alpha$$

$$-2\left[\frac{\int dx\, \psi_\alpha^* H \psi_\alpha}{\int dx\, \psi_\alpha^2}\right]$$

$$\times\left[\frac{\int dx\, \psi_\alpha \frac{d\psi_\alpha}{d\alpha}}{\int dx\, \psi_\alpha^2}\right]$$

$$P_\alpha(x) = \frac{|\psi_\alpha(x)|^2}{\int dx\, |\psi_\alpha(x)|^2}$$

2nd - term (in last term)

$$\frac{\int dx\, \psi_\alpha \frac{d\psi_\alpha}{d\alpha} \cdot \frac{\psi_\alpha}{\psi_\alpha}}{\int dx\, |\psi_\alpha(x)|^2}$$

$$= \int dx\, P_\alpha(x) \cdot \frac{1}{\psi_\alpha} \frac{d\psi_\alpha}{d\alpha}$$

$$\frac{1}{\psi_\alpha}\frac{d\psi_\alpha}{d\alpha} = \frac{d \ln \psi_\alpha}{d\alpha}$$

in most applications

$$\psi_\alpha \sim \exp(\alpha, x, \dots)$$

in project 1 $\psi_\alpha \sim e^{-\frac{1}{2}\alpha^2 x^2}$

$$\ln \psi_\alpha = -\frac{1}{2}\alpha^2 x^2$$

$$\frac{dE}{d\alpha} = 2\left[\left\langle \frac{d\ln\Psi_\alpha}{d\alpha} E_L \right\rangle - \left\langle \frac{d\ln\Psi_\alpha}{d\alpha} \right\rangle \langle E_L \rangle \right]$$

Three integrals to evaluate using MC-sampling