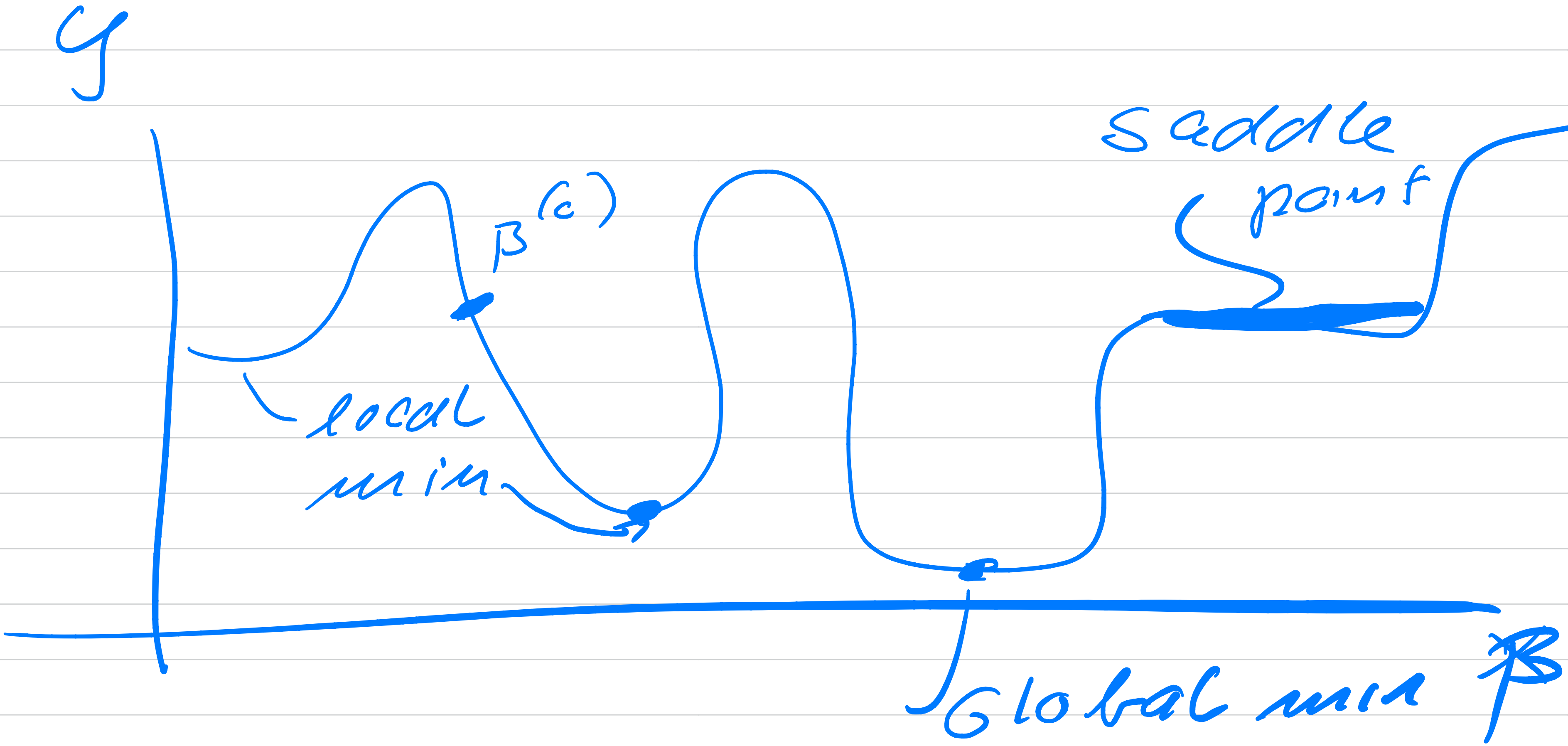
Lecture Fys-stk3155/4155, October 5, 2023

$$\hat{\beta} = \beta^{(n+1)} = \beta^{(n)} - H^{-1}(\beta^{(n)})\, g(\beta^{(n)})$$

$$\approx \beta^{(n)} - \gamma^{(n)} g(\beta^{(n)})$$



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

$$\left(\frac{1}{2} \beta^T H \beta - g^T \beta \right)$$

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

$$Ax = b$$

Example

$$f(x_1, x_2) = \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \overbrace{\begin{bmatrix} 2 & 1 \\ 1 & 20 \end{bmatrix}}^{H} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= x_1^2 + x_1 x_2 + 10 x_2^2 - 5x_1 - 3x_2$$

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 0 = 2x_1 + x_2 - 5$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = 0 \quad x_1 + 20x_2 - 3$$

$$f(x_1^{(n)}, x_2^{(n)})$$

$$x_i^{(n+1)} = x_i^{(n)} - \gamma^{(n)} \frac{\partial f(x_1, x_2)}{\partial x_i}$$

How do we set γ

- constant γ_0 all the way
- exponential decay

$$\gamma^{(k)} = \gamma_0 \exp(-k\gamma^{(k-1)}\tau)$$

$-k\gamma\tau$

k, γ, τ are parameters

- $$\gamma^{(k)} = \frac{\gamma_0}{1+k\gamma_\tau}$$

γ_τ, k, γ_0 are now new parameters

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

k = number of iterations

- linear

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

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

γ_τ is a constant

γ_0 –

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

α is a parameter,

- Adagrad
- RMSprop
- Adam

irrespective of

- gradient descent (GD)
- stochastic GD
  - or/and any of the iterative updates in the previous page

you would set up a grid of values

$$\gamma = \{10^{-5}, 10^{-4}, 10^{-3}, \ldots 10^{-1}\}$$

Steepest descent

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

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

Define a residual

$$r = b - Ax$$

Solution $r = 0$

start with a guess x_0

$$r_0 = -Ax_0 + b \qquad x_0 = 0$$

$$r_0 = b$$

$$r_k = b - Ax_k$$

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

$$r_{k+1} = b - A(\underbrace{x_k + \alpha r_k}_{x_{k+1}})$$

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

want $r_{k+1} = 0$

$$r_k = \alpha_k A r_k$$

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

learning rate

gradient