# Seminar 8. How to accelerate gradient descent?

Alexandr Katrutsa

# Seminar 8

$f(x) = |x|$

$x^* = 0$

$$\boxed{0 \in \partial f(x^*)}$$

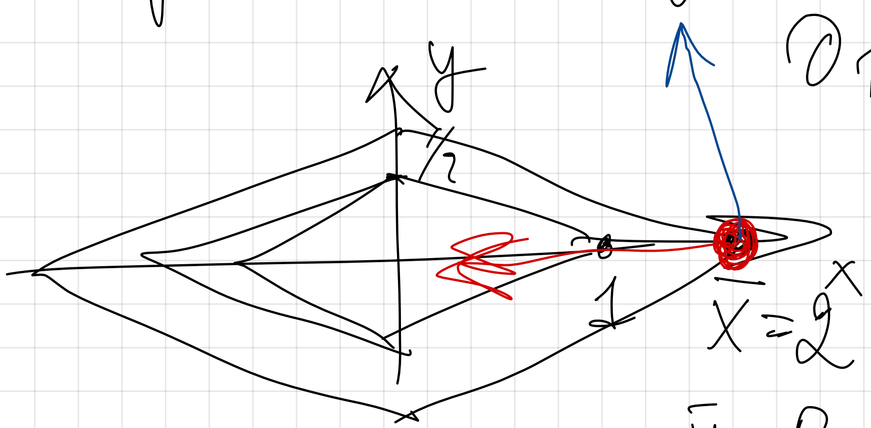
$x_1, x_2, \ldots \to 0 \in \partial f(x_k) \quad \underline{x^* = x_k}$

$x \neq 0 \quad \partial f(x) = \{ f'(x) \}$

$0 \notin \partial f(x_k), \; x_k \neq 0.$

$f(x, y) = |x| + 2|y|$

$\partial f(\bar{x}, \bar{y}) = \begin{pmatrix} 1 \\ [-2; 2] \end{pmatrix}$




$\bar{x} = 2$

$\bar{y} = 0$

$-g: \begin{pmatrix} -1 \\ [-2; 2] \end{pmatrix}$

# Main steps to get the optimal parameters for Heavy-Ball method

$$\begin{bmatrix} x_{k+1} - x^* \\ x_k - x^* \end{bmatrix} = A_k \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \end{bmatrix}$$

$\rho(A_k)$ — ?

$$A_k = \left[\begin{array}{c|c} (1+\beta_k)I - \beta_k A_k(x) & -\beta_k I \\ \hline I & O \end{array}\right]$$

$$A_k \longrightarrow \hat{A}_k = \left[\begin{array}{c|c} \ddots & \ddots \\ \hline \ddots & O \end{array}\right]$$

$A_k(x) = f''(z)$ – based on the theorem about the mean value. $\Rightarrow A_k(x)$ is symmetric $\Rightarrow \exists\ U$ – orthogonal:

$$A_k(x) = U \underbrace{\Lambda}_{\text{diagonal}} U^T$$

After that:

$$\begin{bmatrix} U^T & 0 \\ 0 & U^T \end{bmatrix} \cdot \left[\begin{array}{c|c} (1+\beta_n)I - \alpha_n A_n(x) & -\beta_n I \\ \hline I & 0 \end{array}\right] \cdot$$

$$\cdot \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} = \left[\begin{array}{c|c} \ddots & \ddots \\ \hline \ddots & 0 \end{array}\right]$$

Orthogonal transformation does not change the $\rho$-spectral radius.

Now, we can swap the rows and colums of such matrix. Consider $4 \times 4$ example:

$$\left[\begin{array}{cc|cc} a & 0 & b & 0 \\ 0 & a & 0 & b \\ \hline c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \end{array}\right] \rightarrow \begin{pmatrix} a & b & 0 & 0 \\ 0 & 0 & a & b \\ c & 0 & 0 & 0 \\ 0 & 0 & c & 0 \end{pmatrix} \rightarrow$$

$$\rightarrow \left(\begin{array}{cc|cc} a & b & 0 & 0 \\ c & 0 & 0 & 0 \\ \hline 0 & 0 & a & b \\ 0 & 0 & c & 0 \end{array}\right) = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \Rightarrow$$

$\Rightarrow$ reduce the problem to minimization of maximum eigenvalue of matrices $T_1, T_2, \dots$. But matrices $T_i$ are $2 \times 2$ and their spectra can be computed analytically. From such formulas follows the expressions for $\alpha^*$ and $\beta^*$.