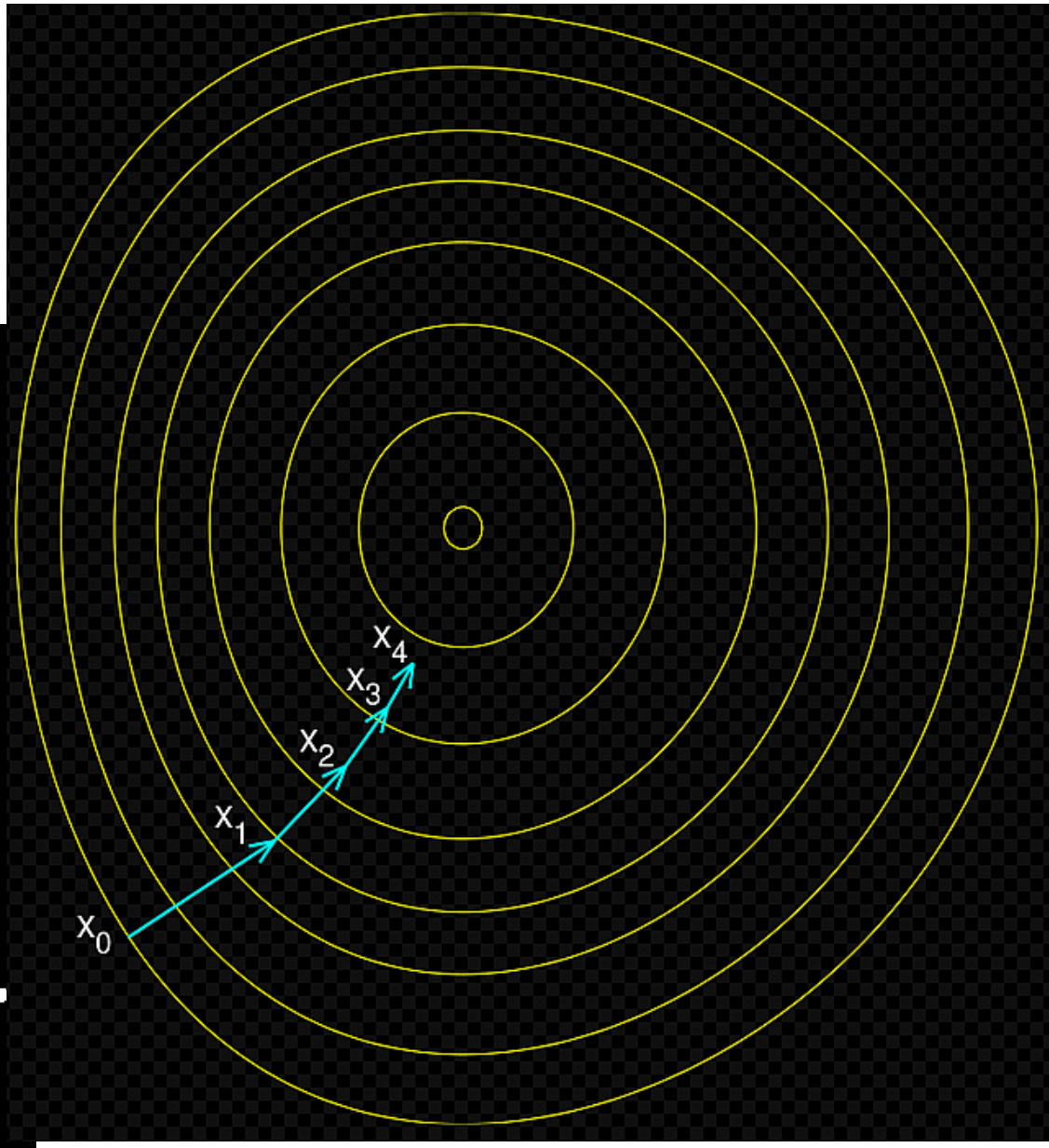
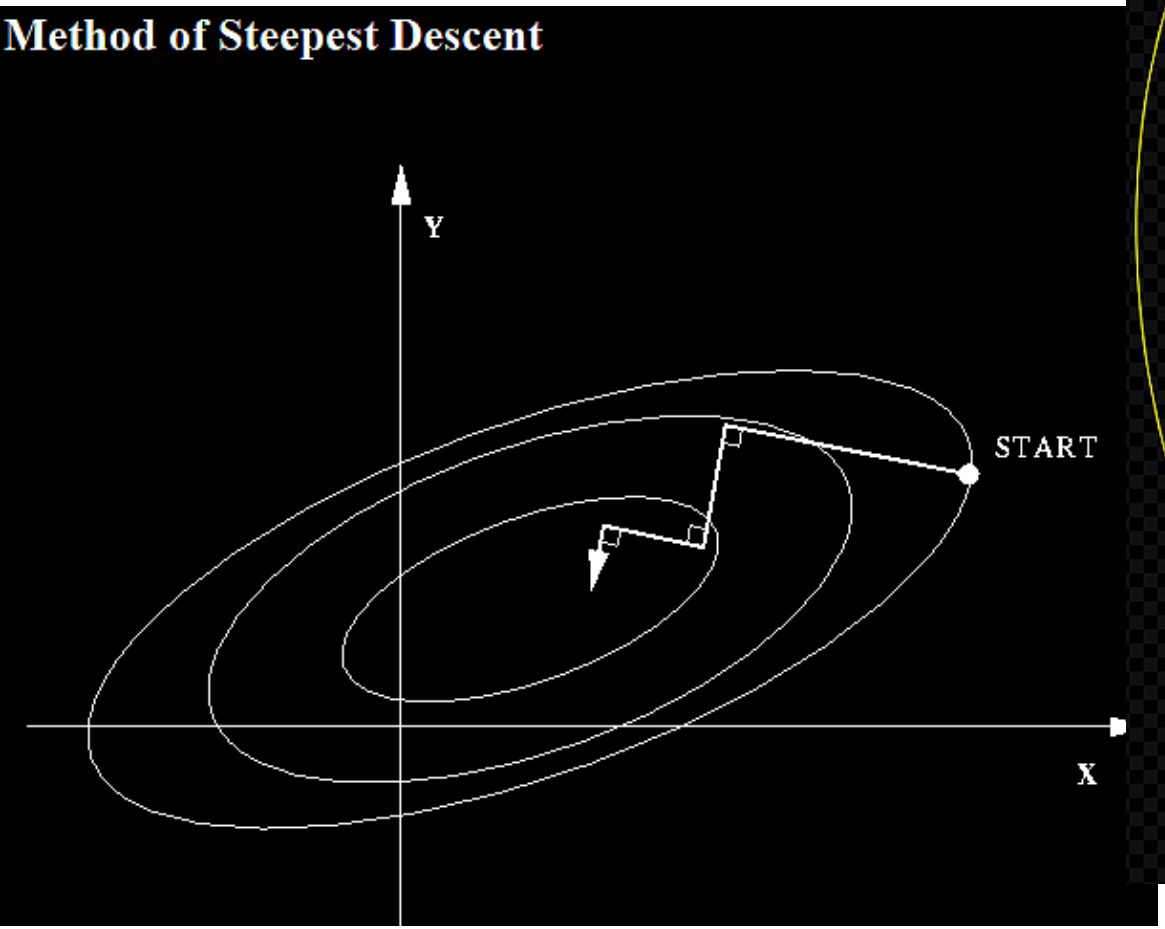
SD

vs

GD

Gradient descent





$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma_n \nabla F(\mathbf{x}_n),\ n \geq 0.$$

$$\gamma_n = \frac{\left|(\mathbf{x}_n - \mathbf{x}_{n-1})^T \left[\nabla F(\mathbf{x}_n) - \nabla F(\mathbf{x}_{n-1})\right]\right|}{\left\|\nabla F(\mathbf{x}_n) - \nabla F(\mathbf{x}_{n-1})\right\|^2}$$

Gradient descent

THE METHOD OF STEEPEST DESCENT

The method of steepest descent is a gradient algorithm where the step size α_k is chosen to achieve the maximum amount of decrease of the objective function at each individual step. Specifically, α_k is chosen to minimize $\phi_k(\alpha) \triangleq f(x^{(k)} - \alpha \nabla f(x^{(k)}))$. In other words,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x^{(k)} - \alpha \nabla f(x^{(k)})).$$

To summarize, the steepest descent algorithm proceeds as follows: at each step, starting from the point $x^{(k)}$ we conduct a line search in the direction $-\nabla f(x^{(k)})$ until a minimizer, $x^{(k+1)}$, is found. A typical sequence resulting from the method of steepest