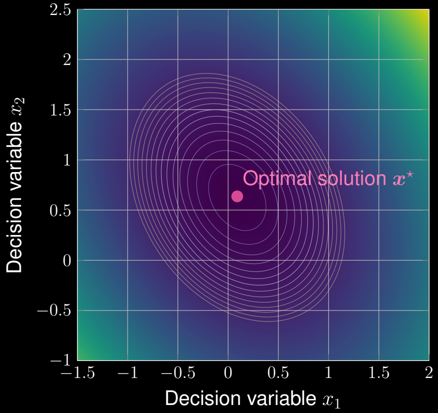
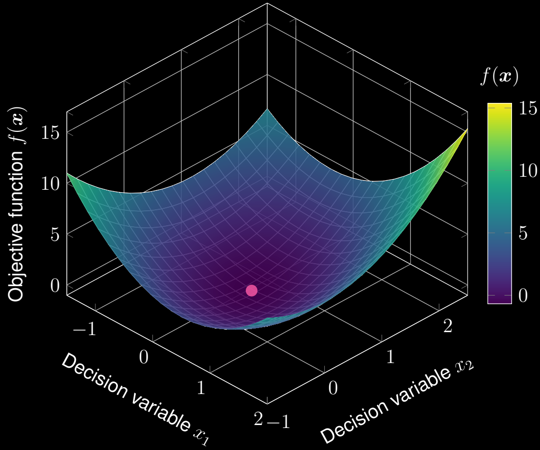
Quadratic Optimization

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} \qquad \text{with} \qquad \boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$





Steepest Gradient Descent for Quadratic Optimization

For any symmetric positive definite matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$:

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

with an initial point $\boldsymbol{x}_0$.

Steepest Gradient Descent for Quadratic Optimization

For any symmetric positive definite matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$:

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

with an initial point $\boldsymbol{x}_0$.

❶ Gradient (first-order derivative of $f(\boldsymbol{x}_k)$):

$$\boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}$$

Steepest Gradient Descent for Quadratic Optimization

For any symmetric positive definite matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$:

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

with an initial point $\boldsymbol{x}_0$.

❶ Gradient (first-order derivative of $f(\boldsymbol{x}_k)$):

$$\boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}$$

❷ Decision vector $\boldsymbol{x}$ is updated by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \underbrace{\alpha_k}_{\text{step size (unknown)}} \boldsymbol{g}_k$$

Steepest Gradient Descent

Quadratic optimization:

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

Rewrite the objective function:

$$\Rightarrow \quad f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k)$$

Steepest Gradient Descent

Quadratic optimization:

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

Rewrite the objective function:

$$\begin{aligned} \Rightarrow \quad f(\boldsymbol{x}_{k+1}) &= f(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) \\ &= \frac{1}{2}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k)^\top \boldsymbol{A}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) - \boldsymbol{b}^\top(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) \end{aligned}$$

Steepest Gradient Descent

Quadratic optimization:

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

Rewrite the objective function:

$$\begin{aligned}
\Rightarrow \quad f(\boldsymbol{x}_{k+1}) &= f(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) \\
&= \frac{1}{2}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k)^\top \boldsymbol{A}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) - \boldsymbol{b}^\top(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) \\
&= \frac{1}{2}\alpha_k^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha_k \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{x}_k + \alpha_k \boldsymbol{b}^\top \boldsymbol{g}_k + \underbrace{\frac{1}{2}\boldsymbol{x}_k^\top \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}^\top \boldsymbol{x}_k}_{\text{constant}}
\end{aligned}$$

Steepest Gradient Descent

Quadratic optimization:

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

Rewrite the objective function:

$$\begin{aligned}
\Rightarrow \quad f(\boldsymbol{x}_{k+1}) &= f(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) \\
&= \frac{1}{2}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k)^\top \boldsymbol{A}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) - \boldsymbol{b}^\top(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) \\
&= \frac{1}{2}\alpha_k^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha_k \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{x}_k + \alpha_k \boldsymbol{b}^\top \boldsymbol{g}_k + \underbrace{\frac{1}{2}\boldsymbol{x}_k^\top \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}^\top \boldsymbol{x}_k}_{\text{constant}} \\
&= \frac{1}{2}\alpha_k^2 \underbrace{\boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k}_{\geq 0} - \alpha_k \boldsymbol{g}_k^\top \underbrace{(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b})}_{=\boldsymbol{g}_k} + \text{constant}
\end{aligned}$$

Steepest Gradient Descent

Optimize step size α_k:

$$\min_{\alpha_k} \quad p(\alpha_k) = \frac{1}{2}\alpha_k^2 \underbrace{\boldsymbol{g}_k^\top \boldsymbol{A} \boldsymbol{g}_k}_{\geq 0} - \alpha_k \underbrace{\boldsymbol{g}_k^\top \boldsymbol{g}_k}_{\geq 0}$$

Steepest Gradient Descent

Optimize step size α_k:

$$\min_{\alpha_k} \quad p(\alpha_k) = \frac{1}{2}\alpha_k^2 \underbrace{\boldsymbol{g}_k^\top \boldsymbol{A} \boldsymbol{g}_k}_{\geq 0} - \alpha_k \underbrace{\boldsymbol{g}_k^\top \boldsymbol{g}_k}_{\geq 0}$$

❸ The optimal α_k:

$$\frac{\mathrm{d}p(\alpha_k)}{\mathrm{d}\alpha_k} = 0$$ (first-order derivative is zero)

Steepest Gradient Descent

Optimize step size α_k:

$$\min_{\alpha_k} \quad p(\alpha_k) = \frac{1}{2}\alpha_k^2 \underbrace{\boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k}_{\geq 0} - \alpha_k \underbrace{\boldsymbol{g}_k^\top \boldsymbol{g}_k}_{\geq 0}$$

❸ The optimal α_k:

$$\frac{\mathrm{d}p(\alpha_k)}{\mathrm{d}\alpha_k} = 0 \quad \text{(first-order derivative is zero)}$$

$$\Rightarrow \quad \alpha_k \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \boldsymbol{g}_k^\top \boldsymbol{g}_k = 0$$

Steepest Gradient Descent

Optimize step size α_k:

$$\min_{\alpha_k} \quad p(\alpha_k) = \frac{1}{2}\alpha_k^2 \underbrace{\boldsymbol{g}_k^\top \boldsymbol{A} \boldsymbol{g}_k}_{\geq 0} - \alpha_k \underbrace{\boldsymbol{g}_k^\top \boldsymbol{g}_k}_{\geq 0}$$

❸ The optimal α_k:

$$\frac{\mathrm{d}p(\alpha_k)}{\mathrm{d}\alpha_k} = 0 \quad \text{(first-order derivative is zero)}$$

$$\Rightarrow \quad \alpha_k \boldsymbol{g}_k^\top \boldsymbol{A} \boldsymbol{g}_k - \boldsymbol{g}_k^\top \boldsymbol{g}_k = 0$$

$$\Rightarrow \quad \alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{g}_k^\top \boldsymbol{A} \boldsymbol{g}_k}$$

Steepest Gradient Descent

Quadratic optimization:

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

△ The update equations:

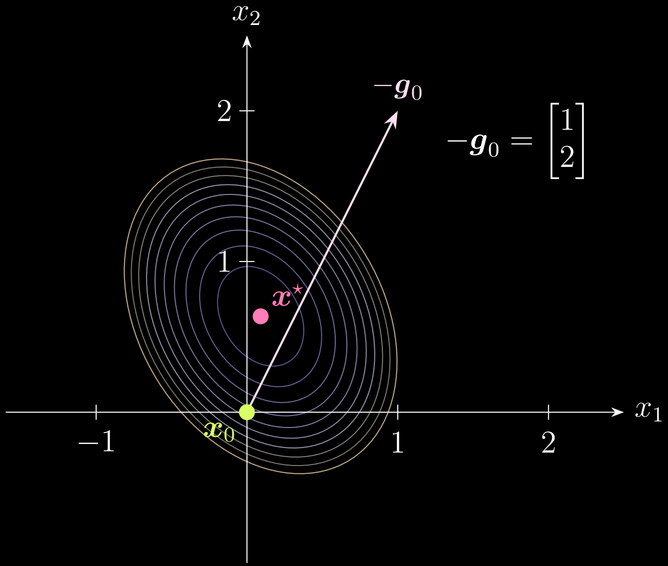
$$\boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b} \qquad \text{(gradient)}$$

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k} \qquad \text{(step size)}$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k \qquad \text{(decision vector)}$$

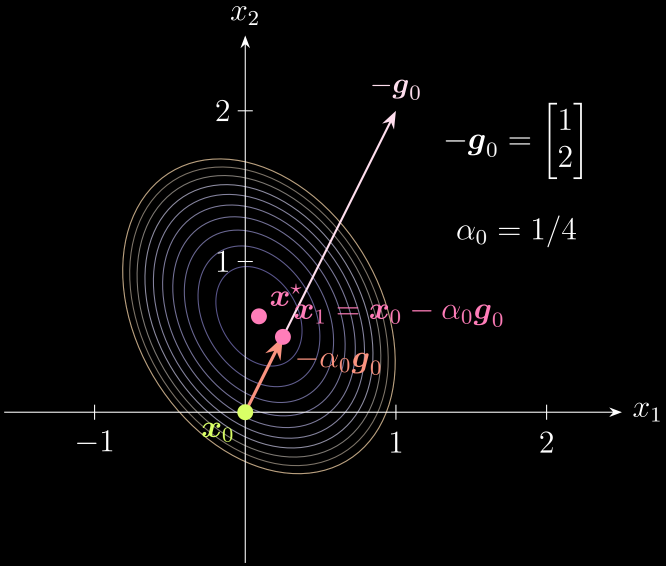
with an initial point $\boldsymbol{x}_0$.

Steepest Gradient Descent for Quadratic Optimization



$$-\boldsymbol{g}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Steepest Gradient Descent for Quadratic Optimization

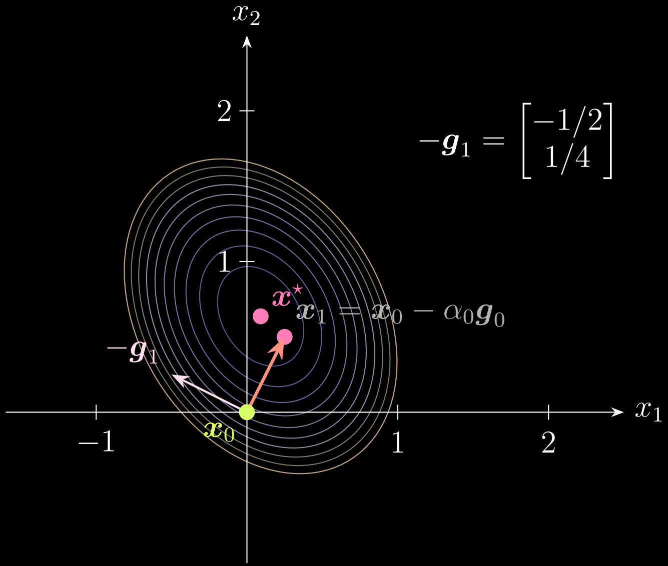


$$-\boldsymbol{g}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

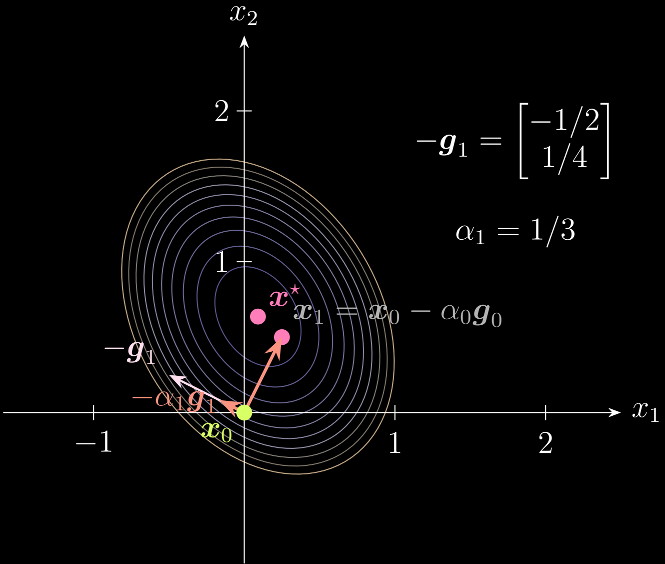
$$\alpha_0 = 1/4$$

Steepest Gradient Descent for Quadratic Optimization

$$-\boldsymbol{g}_1 = \begin{bmatrix} -1/2 \\ 1/4 \end{bmatrix}$$



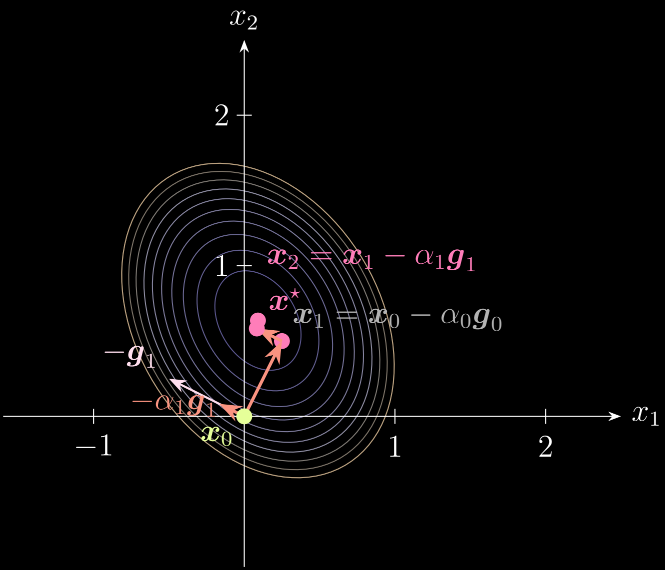
Steepest Gradient Descent for Quadratic Optimization



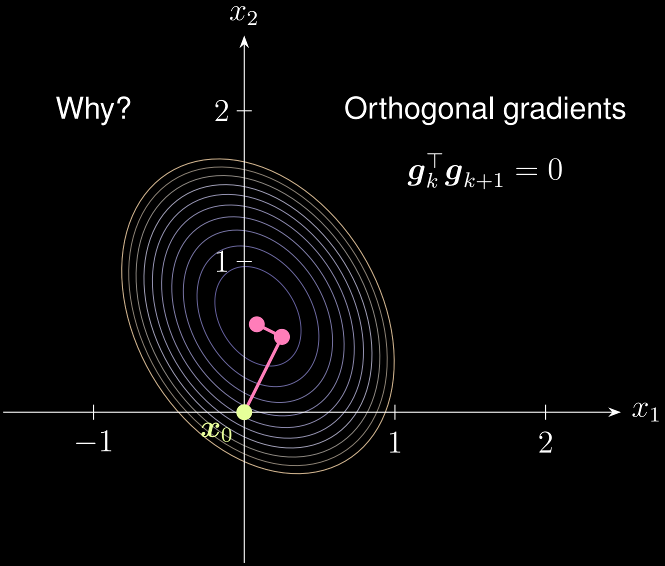
$$-\boldsymbol{g}_1 = \begin{bmatrix} -1/2 \\ 1/4 \end{bmatrix}$$

$$\alpha_1 = 1/3$$

Steepest Gradient Descent for Quadratic Optimization



Steepest Gradient Descent for Quadratic Optimization



Why?

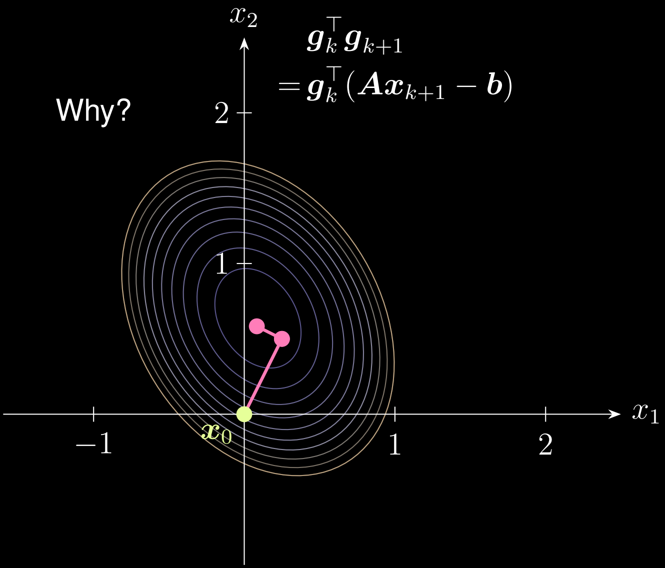
Orthogonal gradients

$$\boldsymbol{g}_k^\top \boldsymbol{g}_{k+1} = 0$$

Steepest Gradient Descent for Quadratic Optimization

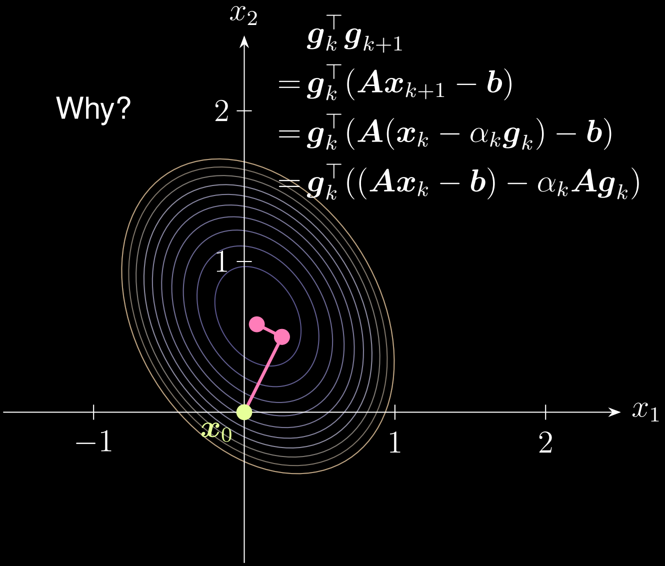
$$\boldsymbol{g}_k^\top \boldsymbol{g}_{k+1}$$
$$= \boldsymbol{g}_k^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b})$$

Why?



Steepest Gradient Descent for Quadratic Optimization

Why?

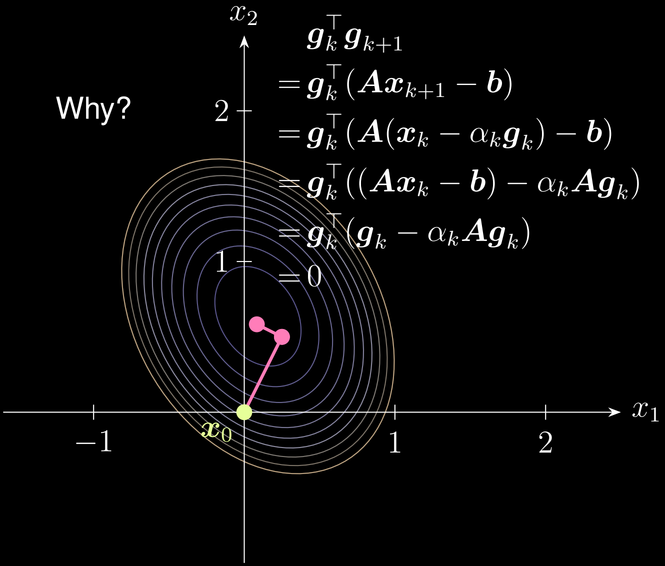


$$
\begin{aligned}
&\boldsymbol{g}_k^\top \boldsymbol{g}_{k+1} \\
&= \boldsymbol{g}_k^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}) \\
&= \boldsymbol{g}_k^\top (\boldsymbol{A}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) - \boldsymbol{b}) \\
&= \boldsymbol{g}_k^\top ((\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) - \alpha_k \boldsymbol{A}\boldsymbol{g}_k)
\end{aligned}
$$

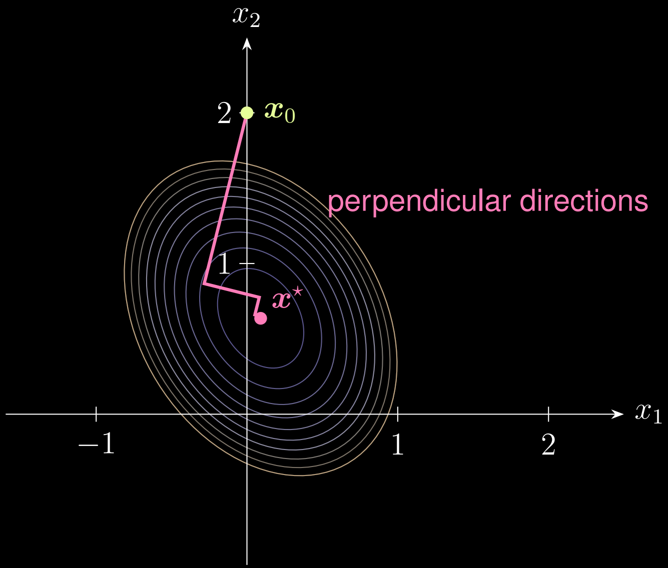
Steepest Gradient Descent for Quadratic Optimization

Why?

$$
\begin{aligned}
&\boldsymbol{g}_k^\top \boldsymbol{g}_{k+1} \\
&= \boldsymbol{g}_k^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}) \\
&= \boldsymbol{g}_k^\top (\boldsymbol{A}(\boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k) - \boldsymbol{b}) \\
&= \boldsymbol{g}_k^\top ((\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) - \alpha_k \boldsymbol{A}\boldsymbol{g}_k) \\
&= \boldsymbol{g}_k^\top (\boldsymbol{g}_k - \alpha_k \boldsymbol{A}\boldsymbol{g}_k) \\
&= 0
\end{aligned}
$$

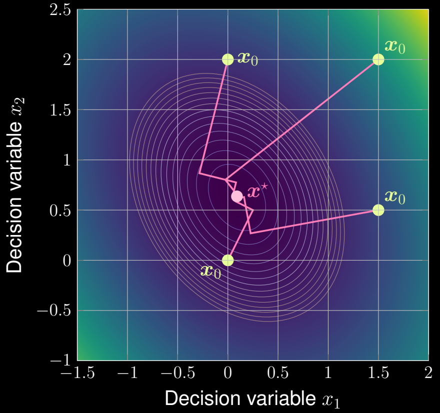


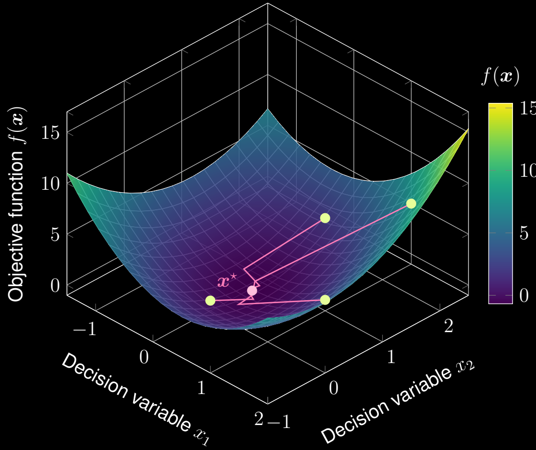
Steepest Gradient Descent for Quadratic Optimization



Steepest Gradient Descent for Quadratic Optimization

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} \qquad \text{with} \qquad \boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$





Thanks for your attention!

About me:

Homepage: https://xinychen.github.io

How to reach me: chenxy346@gmail.com