Applied Numerical Methods for Civil Engineering

CGN 3405 - 0002

Week 13: Optimization techniques: Part II

Xinyu Chen

Assistant Professor

University of Central Florida

Linear & Quadratic Programming

More material:

- Linear programming slides:

  https://xinychen.github.io/youtube/linprog.pdf

- Linear programming explained:

  https://www.youtube.com/watch?v=Wt8oSh26BY8

Derivatives of $f(x)$ and $f(x,y)$

A quick revisit!

- **Derivative.** Given a scalar function $f(x)$ of the single variable x, the derivative is defined by

$$\frac{\mathrm{d}\, f(x)}{\mathrm{d}\, x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$

Derivatives of $f(x)$ and $f(x,y)$

A quick revisit!

- **Derivative.** Given a scalar function $f(x)$ of the single variable x, the derivative is defined by

$$\frac{\mathrm{d}\, f(x)}{\mathrm{d}\, x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$

- **Partial derivatives.** Given a scalar function $f(x,y)$ of two variables x, y, the partial derivatives are defined by

$$\begin{cases} \dfrac{\partial f(x,y)}{\partial x} = \lim\limits_{\Delta x \to 0} \dfrac{f(x+\Delta x, y)}{\Delta x} \\ \dfrac{\partial f(x,y)}{\partial y} = \lim\limits_{\Delta y \to 0} \dfrac{f(x, y+\Delta y)}{\Delta y} \end{cases}$$

Derivative of $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$

Given $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$, write down the derivative $\dfrac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}}$:

- The function $f(\boldsymbol{x})$ can be written as

$$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2$$

- The partial derivatives of $f(x_1, x_2, \ldots, x_n)$ with respect to $x_1, x_2, \ldots, x_n$ are

Derivative of $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$

Given $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$, write down the derivative $\dfrac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}}$:

- The function $f(\boldsymbol{x})$ can be written as

$$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2$$

- The partial derivatives of $f(x_1, x_2, \ldots, x_n)$ with respect to $x_1, x_2, \ldots, x_n$ are

$$\begin{aligned} \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_1} &= 2x_1 \\ \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_2} &= 2x_2 \\ &\vdots \\ \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_n} &= 2x_n \end{aligned} \quad \Rightarrow \quad \frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_1} \\ \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2\boldsymbol{x}$$

Derivative of $f(x) = x^\top A x$

For any symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, suppose $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$, write down the derivative $\dfrac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}}$:

- The function $f(\boldsymbol{x})$ can be written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j$$

- The partial derivatives of $f(\boldsymbol{x})$:

Derivative of $f(x) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$

For any symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, suppose $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$, write down the derivative $\dfrac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\,\boldsymbol{x}}$:

- The function $f(\boldsymbol{x})$ can be written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j$$

- The partial derivatives of $f(\boldsymbol{x})$:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_h} = \sum_{i=1}^{n} A_{ih}x_h + \sum_{j=1}^{n} x_h A_{hj} \quad h = 1, 2, \ldots, n$$

- If $\boldsymbol{A}$ is symmetric, we have

$$\frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\,\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^\top \boldsymbol{x} = 2\boldsymbol{A}\boldsymbol{x}$$

Derivative of $f(x) = \boldsymbol{b}^\top x$

For any vector $\boldsymbol{b} \in \mathbb{R}^n$, suppose $f(\boldsymbol{x}) = \boldsymbol{b}^\top \boldsymbol{x}$, write down the derivative $\frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}}$:

- The function $f(\boldsymbol{x})$ can be written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} b_i x_i$$

- The partial derivatives of $f(\boldsymbol{x})$:

Derivative of $f(x) = \boldsymbol{b}^\top \boldsymbol{x}$

For any vector $\boldsymbol{b} \in \mathbb{R}^n$, suppose $f(\boldsymbol{x}) = \boldsymbol{b}^\top \boldsymbol{x}$, write down the derivative $\dfrac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}}$:

- The function $f(\boldsymbol{x})$ can be written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} b_i x_i$$

- The partial derivatives of $f(\boldsymbol{x})$:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_h} = b_h \quad h = 1, 2, \ldots, n$$

- So, we have

$$\frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}} = \boldsymbol{b}$$

Optimization over Quadratic Form

- Function of quadratic form:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} + c$$

  with symmetric and positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and known $\boldsymbol{b} \in \mathbb{R}^n, c \in \mathbb{R}$.

- Minimizing $f(\boldsymbol{x})$:

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

- Derivative of $f(\boldsymbol{x})$:

$$\frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\,\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

- Let the derivative be zero, it gives us the optimal values:

Optimization over Quadratic Form

- Function of quadratic form:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} + c$$

  with symmetric and positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and known $\boldsymbol{b} \in \mathbb{R}^n, c \in \mathbb{R}$.

- Minimizing $f(\boldsymbol{x})$:

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

- Derivative of $f(\boldsymbol{x})$:

$$\frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

- Let the derivative be zero, it gives us the optimal values:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \quad \Rightarrow \quad \boldsymbol{x}^\star = \boldsymbol{A}^{-1}\boldsymbol{b}$$

Linear Systems

- A system of linear equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

- Rewrite the problem:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}}_{\boldsymbol{A} \in \mathbb{R}^{n \times n}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\boldsymbol{x} \in \mathbb{R}^n} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}}_{\boldsymbol{b} \in \mathbb{R}^n}$$

- Matrix form $\boldsymbol{Ax} = \boldsymbol{b}$

Update Equations

How to solve $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$?

- Essential idea: **Gradient descent** solves linear system iteratively.
- $\boldsymbol{x}$ is updated by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \underbrace{\alpha}_{\text{step size}} \boldsymbol{g}_k$$

  with initial points $\boldsymbol{x}_0$ and known constant α
- $\boldsymbol{g}_k$ is the gradient:

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

- If the residual is $\boldsymbol{r}_k = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_k$, then

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \underbrace{(-\alpha(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}))}_{\text{step size} \times \text{negative gradient}}$$

$$\Rightarrow \quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \underbrace{\alpha\boldsymbol{r}_k}_{\text{step size} \times \text{residual}}$$

Gradient Descent

- Update equations:

$$\begin{aligned} \boldsymbol{g}_k &= \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b} && \text{(gradient)} \\ \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k - \alpha \boldsymbol{g}_k && \text{(variables)} \end{aligned}$$

- Problem statement:

$$\begin{aligned} 4x_1 + x_2 &= 1 \\ x_1 + 3x_2 &= 2 \end{aligned} \qquad \text{with} \qquad \boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Suppose we have initial points $x_1 = 0, x_2 = 0$ (i.e., $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for the initialization) and step size $\alpha = \frac{1}{4}$.

- Exact solution: $x_1 = \frac{1}{11} \approx 0.09, x_2 = \frac{7}{11} \approx 0.63$

A Simple Linear System

Given $\boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and $\alpha = \frac{1}{4}$, we have

- **First iteration**:

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

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - \alpha \boldsymbol{g}_0 = \begin{bmatrix} 1/4 \\ 1/2 \end{bmatrix}$$

- **Second iteration**:

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

$$\boldsymbol{x}_2 = \boldsymbol{x}_1 - \alpha \boldsymbol{g}_1 = \begin{bmatrix} 1/8 \\ 9/16 \end{bmatrix}$$

- **Third iteration**:

$$\boldsymbol{g}_2 = \boldsymbol{A}\boldsymbol{x}_2 - \boldsymbol{b} = \begin{bmatrix} 1/16 \\ -3/16 \end{bmatrix}$$

$$\boldsymbol{x}_3 = \boldsymbol{x}_2 - \alpha \boldsymbol{g}_2 = \begin{bmatrix} 7/64 \\ 39/64 \end{bmatrix} \approx \begin{bmatrix} 0.109 \\ 0.609 \end{bmatrix}$$

Quick Summary

Today's Class:

- Derivatives of $f(\boldsymbol{x})$
- Gradient descent for minimizing

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

- Update equations:

$$\begin{aligned} \boldsymbol{g}_k &= \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b} \qquad \text{(gradient)} \\ \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k - \alpha \boldsymbol{g}_k \quad \text{(variables)} \end{aligned}$$

- Solving linear systems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$

Steepest Gradient Descent

How to solve $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$?

- Essential idea: **Gradient descent** solves linear system iteratively.
- $\boldsymbol{x}$ is updated by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha \boldsymbol{g}_k$$

  with initial points $\boldsymbol{x}_0$ and unknown constant α
- $\boldsymbol{g}_k$ is the gradient:

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

Optimal Step Size α

- Objective function:

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

- The optimal step size α related to $\boldsymbol{x}_k$ and $\boldsymbol{g}_k$:

Optimal Step Size α

- Objective function:

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

- The optimal step size α related to $\boldsymbol{x}_k$ and $\boldsymbol{g}_k$:

$$\begin{aligned} \Rightarrow \quad p(\alpha) &= \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{x}_k + \alpha \boldsymbol{g}_k^\top \boldsymbol{b} + c_1 \\ &= \frac{1}{2}\alpha^2 \underbrace{\boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k}_{\text{scalar}} - \alpha \boldsymbol{g}_k^\top \underbrace{(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b})}_{=\boldsymbol{g}_k} + c_1 \\ &= \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha \boldsymbol{g}_k^\top \boldsymbol{g}_k + c_1 \end{aligned}$$

with the positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$.

Positive Definite Matrix

Extension from $y = ax^2$ to $y = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$:

If $\boldsymbol{A}$ is a **positive definite matrix**, then it always holds that $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$ for any $\boldsymbol{x} \neq \boldsymbol{0}$.

- **Example**: Is $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ a positive definite matrix?
- Solution: For any nonzero vector $\boldsymbol{x} = (x_1, x_2)^\top$, we have
  - matrix-vector multiplication:

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

  - inner product:

$$\boldsymbol{x}^\top (\boldsymbol{A}\boldsymbol{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2 > 0$$

So $\boldsymbol{A}$ is a positive definite matrix.

Optimal Step Size α

- Objective function:

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} + c \\ \Rightarrow \quad f(\boldsymbol{x}_{k+1}) &= f(\boldsymbol{x}_k - \alpha \boldsymbol{g}_k) \\ &= \frac{1}{2}(\boldsymbol{x}_k - \alpha \boldsymbol{g}_k)^\top \boldsymbol{A}(\boldsymbol{x}_k - \alpha \boldsymbol{g}_k) - \boldsymbol{b}^\top (\boldsymbol{x}_k - \alpha \boldsymbol{g}_k) + c \\ \Rightarrow \quad p(\alpha) &= \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha \boldsymbol{g}_k^\top \boldsymbol{g}_k + c_1 \end{aligned}$$

- Optimizing step size α:

$$\min_{\alpha} \quad \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha \boldsymbol{g}_k^\top \boldsymbol{g}_k$$

- Thus, we have an optimal α such that

$$\alpha = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k}$$

by letting $\dfrac{\mathrm{d}\, p(\alpha)}{\mathrm{d}\, \alpha} = 0$.

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

- Recall that

$$
\begin{aligned}
p(\alpha) &= \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{x}_k + \alpha \boldsymbol{g}_k^\top \boldsymbol{b} + c \\
&= \frac{1}{2}\alpha^2 \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \alpha \boldsymbol{g}_k^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) + c
\end{aligned}
$$

- First-order derivative of $p(\alpha)$:

$$
\begin{aligned}
\frac{\mathrm{d}\, p(\alpha)}{\mathrm{d}\, \alpha} &= \alpha \boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k - \boldsymbol{g}_k^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) \\
&= \boldsymbol{g}_k^\top (\alpha \boldsymbol{A}\boldsymbol{g}_k - \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{b}) \\
&= \boldsymbol{g}_k^\top (\boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}_k - \alpha \boldsymbol{g}_k)) \\
&= \boldsymbol{g}_k^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_{k+1}) \\
&= -\boldsymbol{g}_k^\top \boldsymbol{g}_{k+1} = 0
\end{aligned}
$$

- So, gradients $\boldsymbol{g}_k$ and $\boldsymbol{g}_{k+1}$ are orthogonal.

Steepest Gradient Descent

- Update equations:

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

where α_k is the optimal step size for each iteration.

A Simple Linear System

- Problem statement:

$$\begin{aligned} 4x_1 + x_2 &= 1 \\ x_1 + 3x_2 &= 2 \end{aligned} \qquad \text{with} \qquad \boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

  Suppose we have initial points $x_1 = 0, x_2 = 0$ (i.e., $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for the initialization).

- Exact solution: $x_1 = \dfrac{1}{11}, x_2 = \dfrac{7}{11}$

A Simple Linear System

Given $\boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, we have

- **First iteration**:

$$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b} = \begin{bmatrix} -1 \\ -2 \end{bmatrix} \quad \alpha_0 = \frac{\boldsymbol{g}_0^\top \boldsymbol{g}_0}{\boldsymbol{g}_0^\top \boldsymbol{A}\boldsymbol{g}_0} = \frac{1}{4} \quad \boldsymbol{x}_1 = \boldsymbol{x}_0 - \alpha_0 \boldsymbol{g}_0 = \begin{bmatrix} 1/4 \\ 1/2 \end{bmatrix}$$

- **Second iteration**:

$$\boldsymbol{g}_1 = \boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b} = \begin{bmatrix} 1/2 \\ -1/4 \end{bmatrix} \quad \alpha_1 = \frac{\boldsymbol{g}_1^\top \boldsymbol{g}_1}{\boldsymbol{g}_1^\top \boldsymbol{A}\boldsymbol{g}_1} = \frac{1}{3} \quad \boldsymbol{x}_2 = \boldsymbol{x}_1 - \alpha_1 \boldsymbol{g}_1 = \begin{bmatrix} 1/12 \\ 7/12 \end{bmatrix}$$

A Simple Linear System

Given $\boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, we have

- **First iteration**:

$$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b} = \begin{bmatrix} -1 \\ -2 \end{bmatrix} \quad \alpha_0 = \frac{\boldsymbol{g}_0^\top \boldsymbol{g}_0}{\boldsymbol{g}_0^\top \boldsymbol{A}\boldsymbol{g}_0} = \frac{1}{4} \quad \boldsymbol{x}_1 = \boldsymbol{x}_0 - \alpha_0 \boldsymbol{g}_0 = \begin{bmatrix} 1/4 \\ 1/2 \end{bmatrix}$$

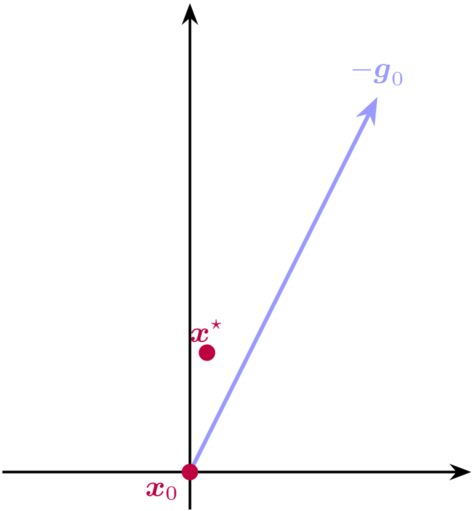
- **Second iteration**:

$$\boldsymbol{g}_1 = \boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b} = \begin{bmatrix} 1/2 \\ -1/4 \end{bmatrix} \quad \alpha_1 = \frac{\boldsymbol{g}_1^\top \boldsymbol{g}_1}{\boldsymbol{g}_1^\top \boldsymbol{A}\boldsymbol{g}_1} = \frac{1}{3} \quad \boldsymbol{x}_2 = \boldsymbol{x}_1 - \alpha_1 \boldsymbol{g}_1 = \begin{bmatrix} 1/12 \\ 7/12 \end{bmatrix}$$

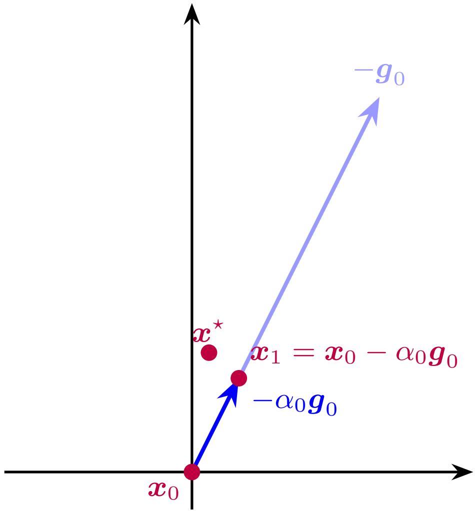
Check the **orthogonality**?

$$\boldsymbol{g}_1^\top \boldsymbol{g}_0 = \begin{bmatrix} 1/2 & -1/4 \end{bmatrix} \begin{bmatrix} -1 \\ -2 \end{bmatrix} = 0$$

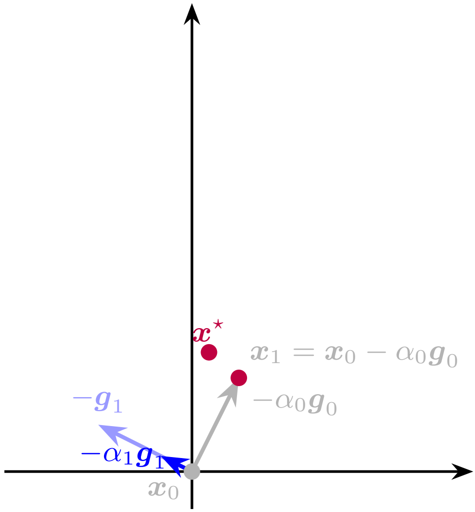
Geometric Interpretation: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k$



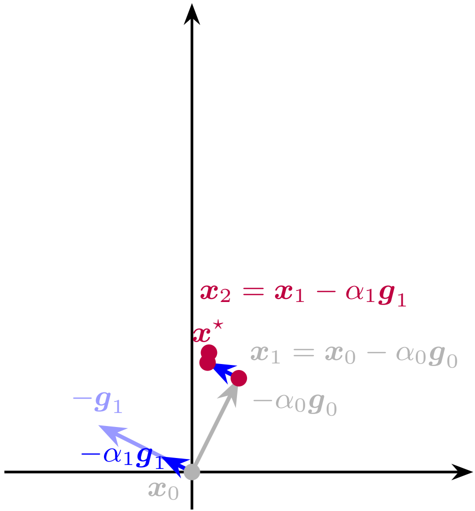
Geometric Interpretation: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k$



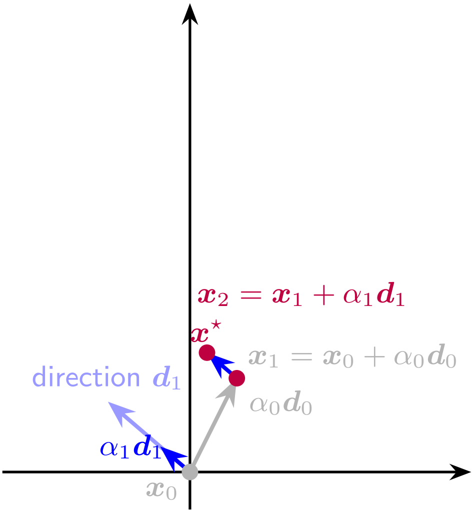
Geometric Interpretation: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k$



Geometric Interpretation: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k$



Revisit Conjugate Gradient



Quick Summary

Today's Class:

- Steepest gradient descent with optimal step size α
- How to write down the formula of α?
- How to prove the orthogonal gradients?
- Solving linear systems $\boldsymbol{Ax} = \boldsymbol{b}$

Hessian Matrix

Function $f(\boldsymbol{x}) = x_1^2 + x_2^2$:

- The first-order derivatives:

$$\boldsymbol{g} = \nabla f(\boldsymbol{x}) = \frac{\mathrm{d}\, f(\boldsymbol{x})}{\mathrm{d}\, \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

- The second-order derivatives:

$$\boldsymbol{H} = \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

This is the Hessian matrix!

Hessian Matrix

How about $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$?

- **Gradient**: First-order derivatives

$$\boldsymbol{g} = \nabla f(\boldsymbol{x}) = 2\boldsymbol{x}$$

- **Hessian matrix**: Second-order derivatives

$$\boldsymbol{H} = \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{I}$$

Hessian Matrix

- For any $\boldsymbol{x} \in \mathbb{R}^n$, if all second-order derivatives of $f(\boldsymbol{x})$ exist, then the Hessian matrix $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is defined as

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

  equivalently, we have

$$\boldsymbol{H} = \begin{bmatrix} | & | & & | \\ \frac{\partial \boldsymbol{g}}{\partial x_1} & \frac{\partial \boldsymbol{g}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{g}}{\partial x_n} \\ | & | & & | \end{bmatrix}$$

- The entry of the ith row and the jth column:

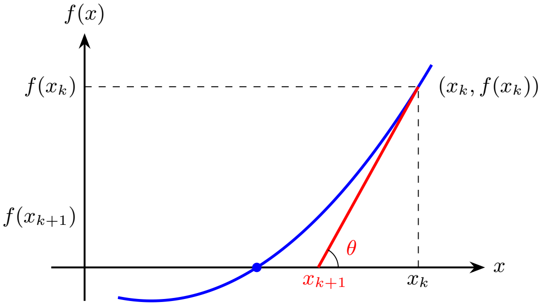
$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Newton-Raphson Method

Slope $f'(x_k)$:

- Derivative of function $f(x)$ evaluated at x_k
- Definition of the slope of a line:

$$\text{slope} = \frac{\text{change in } y}{\text{change in } x} = \frac{\Delta y}{\Delta x} = \frac{0 - f(x_k)}{x_{k+1} - x_k}$$



Newton-Raphson Method

We start with the equation:

$$f'(x_k) = \frac{-f(x_k)}{x_{k+1} - x_k}$$

1. Multiply both sides by $(x_{k+1} - x_k)$:

$$f'(x_k) \cdot (x_{k+1} - x_k) = -f(x_k)$$

2. Divide both sides by $f'(x_k)$:

$$x_{k+1} - x_k = -\frac{f(x_k)}{f'(x_k)}$$

3. Add x_k to both sides to isolate x_{k+1}:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Single-Variable vs. Multi-Variable Function

- Solving $f(x) = 0$:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

- Minimizing $f(\boldsymbol{x})$:

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x})$$

- Solving $\nabla f(\boldsymbol{x}) = \mathbf{0}$:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k - \underbrace{\left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1}}_{\text{inverse of Hessian matrix}} \nabla f(\boldsymbol{x}_k) \\ &= \boldsymbol{x}_k - \boldsymbol{H}_k^{-1} \boldsymbol{g}_k \end{aligned}$$

Newton's Method

- Update equation:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}_k^{-1}\boldsymbol{g}_k$$

  for solving

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

  where

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \quad \text{(gradient)} \qquad\qquad \boldsymbol{H} = \boldsymbol{A} \quad \text{(Hessian matrix)}$$

- Why $\boldsymbol{H} = \boldsymbol{A}$? Hint: $g_i = \sum_{j=1}^{n} A_{ij}x_j - b_i$, $i = 1, 2, \ldots, n$

A Simple Linear System

- Problem statement:

$$\begin{aligned} 4x_1 + x_2 &= 1 \\ x_1 + 3x_2 &= 2 \end{aligned} \quad \text{with} \quad \boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

  Suppose we have initial points $x_1 = 0, x_2 = 0$ (i.e., $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for the initialization).
- Exact solution: $x_1 = \frac{1}{11}, x_2 = \frac{7}{11}$

A Simple Linear System

- Inverse of Hessian matrix:

$$\boldsymbol{H}_k^{-1} = \boldsymbol{A}^{-1} = \frac{\mathrm{adj}(\boldsymbol{A})}{\det(\boldsymbol{A})} = \frac{1}{11}\begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}$$

- **First iteration**:

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

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - \boldsymbol{H}_0^{-1}\boldsymbol{g}_0 = \begin{bmatrix} 1/11 \\ 7/11 \end{bmatrix}$$

where

$$\boldsymbol{H}_0^{-1}\boldsymbol{g}_0 = \begin{bmatrix} 3/11 & -1/11 \\ -1/11 & 4/11 \end{bmatrix}\begin{bmatrix} -1 \\ -2 \end{bmatrix} = \begin{bmatrix} -3/11 + 2/11 \\ 1/11 - 8/11 \end{bmatrix} = \begin{bmatrix} -1/11 \\ -7/11 \end{bmatrix}$$

Exact solution!

Gradient Descent vs. Newton's Method

Solving the optimization

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} \qquad \Rightarrow \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- Gradient descent (slowest):

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha \boldsymbol{g}_k \qquad \boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}$$

- Steepest gradient descent (faster):

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k \qquad \boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b} \qquad \alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{g}_k^\top \boldsymbol{A}\boldsymbol{g}_k}$$

- Conjugate gradient (faster):

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \qquad \boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k \boldsymbol{A}\boldsymbol{x}_k \qquad \boldsymbol{d}_{k+1} = \boldsymbol{r}_{k+1} + \beta_k \boldsymbol{d}_k$$

- Newton's method (fastest):

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}_k^{-1} \boldsymbol{g}_k \qquad \boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b} \qquad \boldsymbol{H}_k = \boldsymbol{A}$$

Quick Summary

Today's Class:

- Definition of Hessian matrix
- Newton's method in optimization