The Gradient

Quick Reminders

- It is important to remember that $x \in \mathbb{R}^n$ is a vector:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{or} \qquad x = [x_1, x_2, \ldots, x_n]^T$$

  This true for any vector: x is not special.

- We use $x^T y = \sum_{i=1}^{n} x_i y_i$ to denote the Cartesian (dot) product.

- The norm (length) of a vector p is $\|p\|_2 = \sqrt{p^T p}$

Recall the Definition of a Local Minimizer

Definition: Local minimizer

A point x^* is a local minimizer if there is a neighborhood $\mathcal{N}$ of x^* such that $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}$.

The Gradient of a Function

Definition: Gradient

Consider $f : \mathbb{R}^n \to \mathbb{R}$. We say f is differentiable at x if there exists $g \in \mathbb{R}^n$ such that

$$\lim_{\epsilon \to 0} \frac{f(x + \epsilon p) - f(x)}{\epsilon \|p\|_2} - \frac{g^T p}{\|p\|_2} = 0,$$

for all $p \in \mathbb{R}^n$ with $p \neq 0$.

The Gradient of a Function (cont.)

The Gradient of a Function (cont.)

If g exists, we denote it by ∇f, and its components are given by the partial derivatives of f:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$