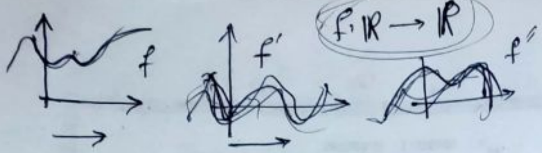
Mathematics for AI

Lecture 19

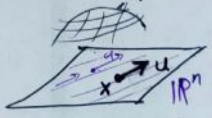
Second Derivatives, Multilinear maps and Tensors, the Hessian Matrix, Quadratic Functions, Quadratic Approximation, Taylor Series

Second derivative in multivariable functions

$f: \mathbb{R} \to \mathbb{R}$



$f: \mathbb{R}^n \to \mathbb{R}$



$$D[u]f(x) = D[u]f\big|_x = (D[u]f)(x)$$

$$= \frac{d}{d\alpha} f(x+\alpha u) = \nabla_x^T u$$

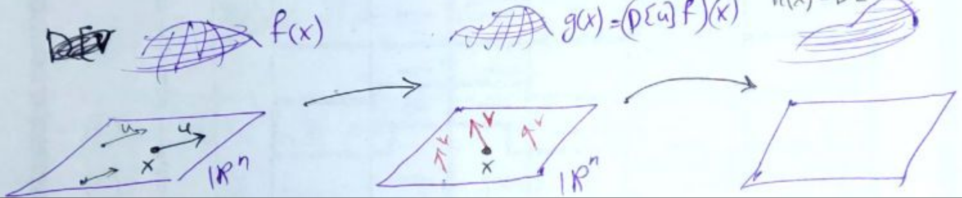
differentiable

u: fixed

$D[u]f: \mathbb{R}^n \to \mathbb{R}$

$$(\underbrace{D[u]f}_{g})(x) = \frac{d}{d\alpha} f(x+\alpha u)$$

Take a second directional derivative, this time in direction $\vec{v}$.



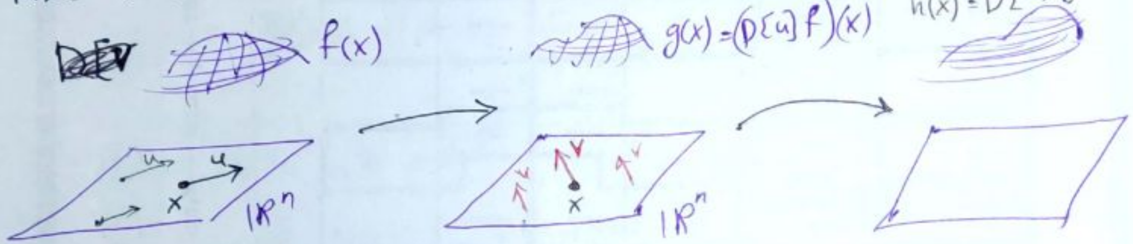
Second derivative in multivariable functions

K. N. Toosi
University of Technology

$D[u]f : \mathbb{R}^n \to \mathbb{R}$

$$\underbrace{(D[u]f)}_{g}(x) = \frac{d}{d\alpha} f(x+\alpha u)$$

Take a second directional derivative, this time in direction $\vec{v}$.



$$(D[v]\,g)(x) = D[v]\big(D[u]f\big)(x) \doteq D^2[v,u]$$

if f is twice differentiable $\boxed{\text{linear in both } u, v}$

$D^2[v,u]$: bilinear in (v,u)

Bilinear maps

Let $\ell : \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}$ be bilinear

$$\begin{cases} \ell(\alpha x + \beta x', y) = \alpha\, \ell(x,y) + \beta\, \ell(x', y) \\ \ell(x, \alpha y + \beta y') = \alpha\, \ell(x,y) + \beta\, \ell(x, y') \end{cases}$$

$\ell(x,y)$ $\quad x \in \mathbb{R}^m,\ y \in \mathbb{R}^n$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ | \\ x_m \end{bmatrix} = x_1 \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{e_1} + x_2 \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{e_2} \cdots + \cdots + x_m \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{e_m}$$

$$x = \sum x_i \cdot \vec{e}_i$$

$$\ell\Big(\sum_{i=1}^{m} x_i \cdot e_i,\ \sum_{j=1}^{n} y_j \cdot e'_j\Big)$$

$$= \sum_{i=1}^{m} x_i \cdot \ell\Big(e_i,\ \sum_{j=1}^{n} y_j \cdot e'_j\Big)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} y_j x_i \cdot \underbrace{\ell(e_i, e'_j)}_{a_{ij}} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i \cdot y_j\ a_{ij} = x^T A y$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \end{bmatrix}$$

Tensors of order 1 and 2

$f(x)$ $\quad f: \mathbb{R}^n \rightarrow \mathbb{R}$, f linear $\Rightarrow \exists a$ $f(x) = a^T x$

$a \in \mathbb{R}^n$

f: is a tensor of degree 1 $\quad a \in \mathbb{R}^n$

(rank) $\quad$ بردار

order $\quad$ a 1D array $a[i]$

$$\boxed{f(x) = a^T x = \sum_{i=1}^{m} a_i x_i}$$

$f(x,y): \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ $\quad$ bilinear

$$\Longrightarrow \exists A \in \mathbb{R}^{m \times n} \quad f(x,y) = x^T A y$$

f: is a degree 2 tensor

A

$$A_{ij} = a_{ij} = f(e_i, e_j)$$

A: 2D array $A[i;j]$

$$\boxed{f(x,y) = x^T A y = \sum_{i=1}^{m} \sum_{j=1}^{n} a\, x_i\, y_i\, a_{ij}}$$

Trilinear functions, tensors of order 3

K. N. Toosi
University of Technology

$f(x,y,z): \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p$ trilinear

f is a tensor of rank 3
degree

A: 3D array

$a_{ijk} = A[i,j,k]$

$f(x,y,z) = \sum_{i,j,k}\sum\sum x_i \, y_j \, z_k \, a_{ijk}$

Multilinear functions, tensors of order p

$f(x_1, x_2, \dots, x_p)$: $\quad f: \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \dots \times \mathbb{R}^{n_p} \longrightarrow \mathbb{R}$

$x_i \in \mathbb{R}^{n_i}$ $\quad$ Multilinear

$\Longrightarrow$ f is a tensor of rank p

f: can be represented by a p-dimentional array

The Hessian Matrix

K. N. Toosi
University of Technology

$D^2[u,v]f = D[v]D[u]f$ is a bilinear function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$\Rightarrow D^2[u,v]\Big|_x = D[v]\big(D[u]f\big)\Big|_x = v^T H u \qquad H = H(x) \in \mathbb{R}^{n\times n}$$

Hessian Matrix

The Hessian Matrix

K. N. Toosi
University of Technology

$D^2[u,v]f = D[v]D[u]f$ is a bilinear function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$\Rightarrow D^2[u,v]\Big|_x = D[v]\big(D[u]f\big)\Big|_x = v^T H u \qquad H = H(x) \in \mathbb{R}^{n \times n}$$

Hessian Matrix

The Hessian Matrix

K. N. Toosi
University of Technology

$$(D[u]f)(x) = \nabla_f^T u = \left[\frac{\partial f}{\partial x_1}\ \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n}\right] u = \boxed{\sum \frac{\partial f}{\partial x_i} u_i}$$

MA 19 III

$$D[v](D[u]f) = \nabla_{D[u]f}^T v = \sum_{i=1}^{n} \left(\nabla_{\frac{\partial f}{\partial x_i}} u_i\right)^T v$$

$$= \sum_{i=1}^{n} \cancel{\sum_{j=1}^{n}} u_i \, \nabla_{\frac{\partial f}{\partial x_i}}^T v = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i \left(\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) v_j\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} u_i \cdot v_j \boxed{\frac{\partial^2 f}{\partial x_j \partial x_i}}$$

$$= \sum \sum u_i \cdot v_j \; h_{ji}$$

$$\left[\frac{\partial^2 f}{\partial x^2} \ \ \frac{\partial^2 f}{\partial x \partial x} \cdots \frac{\partial f}{\partial x}\right] = v^T H u$$

The Hessian Matrix

$$\mathbf{H}_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\[2.2ex] \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\[2.2ex] \vdots & \vdots & \ddots & \vdots \\[2.2ex] \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

https://en.wikipedia.org/wiki/Hessian_matrix

The Hessian Matrix

$$C^0 \supseteq C^1 \supseteq C^2 \cdots \supseteq C^\infty$$

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

symmetric

$$f \in C^2 \Longrightarrow H^T = H \qquad \frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

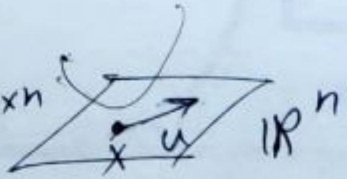
↳ twice continuously differentiable

Directional Curvature

Directional curvature

$$C[u] = D[u]D[u]f(x) = u^T H u$$

$$H = H(x) \in \mathbb{R}^{n \times n}$$



Calculating Hessian

K. N. Toosi
University of Technology

MA19

How to derive the Hessian matrix

1- find $\frac{\partial^2 f}{\partial x_j \partial x_i}$ for $i = 1 - n$, $j = 1 - n$ and arrang in a matrix

2- (a) find $D[u]f = g(x, u)$

(b) find $D[v]g = D[v]D[u]f(x)$

(c) write in the form $v^T H u$

Quadratic forms and Quadratic functions

K. N. Toosi
University of Technology

Quadratic form

$$f(x,y) = f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = ax^2 + by^2 + cxy$$

$$x = \begin{bmatrix} x \\ y \end{bmatrix} \qquad = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$x^T A x$$

Quadratic function

$$f(x,y) = ax^2 + by^2 + cxy + dx + ey + g$$

$$= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + g$$

$$x^T A x + h^T x + g$$

A symmetric

Quadratic forms and Quadratic functions

Quadratic form $f(x)$ $\quad f: \mathbb{R}^n \longrightarrow \mathbb{R}$

$$f(x) = x^T A x \quad A \text{ symmetric}$$

Quadratic function $f(x) = x^T A x + b^T x + c$

$A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$

symmetric

$$\nabla_f = 2Ax + \vec{b} \in \mathbb{R}^n$$

$$H_f = 2A \in \mathbb{R}^{n \times n}$$

A positive definite

Linear vs Quadratic Approximation

K. N. Toosi
University of Technology

MA 19 (V)

$$f(x) = \underline{x^T A x} + \boxed{b^T x + c}$$

$$A = 0_{n\times n} \Rightarrow f: \text{affine}$$

Approximate f around x_0

1: linearization $f(x) = f(x_0) + \nabla|_{x_0}^T (x - x_0)$

2: ~~pr~~ approximate with a quadratic (better)

$$f(x) = f(x_0) + \nabla|_{x_0}^T (x - x_0) + \frac{1}{2}(x - x_0)^T H|_{x_0} (x - x_0)$$

$$f(x) = f(x_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - x_{0i}) + \frac{1}{2} \sum_i \sum_j (x_i - x_{0i})(x_j - x_{0j})$$

Taylor Series

K. N. Toosi
University of Technology

$$f(x) = f(x_0) + \sum \frac{\partial f}{\partial x_i}(x_i - x_{0i}) + \frac{1}{2}\sum\sum (x_i - x_{0i})(x_j - x_{0j}) \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$+ \frac{1}{3!}\sum_{i}\sum_{j}\sum_{k=1}^{n} (x_i - x_{0i})(x_j - x_{0j})(x_k - x_{0k}) \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}$$

$$+ \frac{1}{4!}\sum\sum\sum\sum$$

$$\vdots$$

Exercise: Calculate Hessian

K. N. Toosi
University of Technology

$$f(x) = x^T A x + b^T x + c = \sum_i \sum_j x_i x_j a_{ij} + \sum_i b_i x_i + c$$

A symmetric

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

$$D[u]f = 2\, u^T A x + b^T u$$

$$D[v](D[u]f) = 2\, D[v](u^T A x) + \cancelto{0}{D[v](b^T u)}$$

$$= 2 \frac{d}{d\alpha} u^T A (x + \alpha v)\Big|_{\alpha=0} = 2\, u^T A v = 2\, v^T A u$$

A symmetric

$$= v^T \underbrace{(2A)}_{\text{Hessian}} u$$

$$f(x) = x^T x\, x^T A x$$

Exercise: Calculate Hessian

Calculate Hessian for

$f(x) = x^T x x^T A x$