Module – 2.3

CONCEPT FROM MULTI-VARIATE CALCULUS:

AN OVERVIEW

S. Lakshmivarahan

School of Computer Science

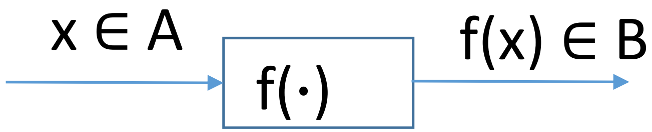
University of Oklahoma

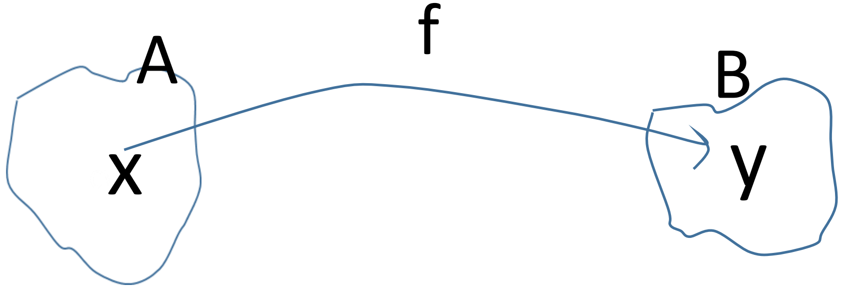
Norman, Ok – 73069, USA

varahan@ou.edu

FUNCTIONS

- f: A -> B, A - Domain, B – range
- f is defined for all members of the domain and by definition it is single-valued, that is, $f(x) \in B$ is unique for $x \in A$





- f is 1-1 (injective) if $f(x) \neq f(y)$ for $x \neq y$ ($|A| \leq |B|$)
- f is onto (surjective) if $B = \{f(x) \mid x \in A\}$ ($|A| \geq |B|$)
- f is 1-1 and onto (bijective) if f is both injective and surjective

<u>Examples of functions:</u> $f(x) = |x|$, x^2, sinx, e^x

TYPES OF FUNCTIONS

1. f is a scalar valued function of a scalar: f: R -> R
   - Examples: f(x) = $x\log_2 x$, 2^x, e^x
2. f is a scalar valued function of a vector: f: R^n -> R
   - This is also called a functional
   - Examples:
     - f(x) = ||x||, x^TAx
     - f(x) = <a, x> for a fixed a ∈ R^n
3. f is a vector valued function of a vector: f: R^n -> R^m
   - f(x) = $(f_1(x), f_2(x), \ldots f_m(x))^T$
   - Examples: n = 3, m = 2, x = $(x_1, x_2, x_3)^T$

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} x_1^2 + x_2^2 + x_3^2 \\ x_1x_2x_3 \end{pmatrix}$$

4. c[a, b] – set of all continuous functions defined on [a, b]

   c^k[a, b] – set of all functions with continuous derivative of order up to k.

THE GRADIENT

- Let $f: R^n \to R$. Let $x, z \in R^n$
- $f(x)$ is differentiable at x if and only if there exists a vector $u \in R^n$ such that

$$f(x + z) - f(x) = \langle u, z \rangle + HOT(z)$$

HOT(z) = higher order term in z

$$\lim_{||z|| \to 0} \frac{\mathrm{HOT(z)}}{||z||} = 0$$

- The vector $u \in R^n$ defined above is called the Gradient of $f(x)$ with respect to x
- Gradient is denoted by $\nabla_x f(x)$ and

$$\nabla_x f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots \frac{\partial f}{\partial x_n} \right)^T$$

is a vector of partial derivation of $f(x)$

PROPERTIES OF GRADIENT OPERATOR ∇

- Let f, g: $R^n \rightarrow R$
- $\nabla_x(f + g) = \nabla_x f + \nabla_x g$ – Additive
- $\nabla_x(cf) = c\nabla_x f(x)$ – Homogeneous
- $\nabla_x(fg) = f(x)\, \nabla_x g + (\nabla_x f(x))g(x)$ – product rule
- Directional derivative of f at x in the direction $z \in R^n$:

$$f'(x, z) = \langle \nabla_x f(x), z \rangle = ||\nabla_x f||\,||z||\cos\theta$$

where θ is the angle between $\nabla_x f$ and z

- A differentiable function changes at a maximum rate when $z = \nabla_x f(x)$ by Cauchy-Schwarz inequality – (Module 2)
- Let $x(t) = (x_1(t), x_2(t), \dots x_n(t))^T$, then

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t} + \dots + \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t} + \dots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t}$$

is called the total derivative of f with respect to t by chain rule

THE HESSIAN MATRIX

- Let f: R^n -> R
- The Hessian matrix, denoted by ∇_x^2f is an nxn matrix of second-order partial derivatives

$$\nabla_x^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] \in R^{nxn}$$

- Hessian f is naturally a symmetric matrix, since

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

THE JACOBIAN MATRIX

- Let f: $R^n \to R^m$, $x \in R^n$, $f(x) = (f_1(x), f_2(x), \dots f_m(x))^T$
- The Jacobian of f denoted by $D_x(f)$ is an mxn matrix

$$D_x(f) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f_i}{\partial x_j}\right] \in R^{mxn}$$

- Notice that the rows of $D_x(f)$ are the transpose of the gradient of f_i, $1 \le i \le m$

EXAMPLES

1. Let a, x ∈ R^n $f(x) = a^T x = \sum_{i=1}^{n} a_i x_i$

   Then $\nabla_x f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a$

2. Let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ $f(x) = x^T A x$

   $f(x) = ax_1^2 + 2bx_1x_2 + cx_2^2$

   $\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2ax_1 + 2bx_2 \\ 2bx_1 + 2cx_2 \end{bmatrix} = 2\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2Ax$

3. Let $f(x) = \frac{1}{2} x^T A x - b^T x$. Then

   $\nabla_x f(x) = Ax - b$

EXAMPLES

4. Let $h(x) = (h_1(x), h_2(x), \dots h_m(x))^T$. Let $f(x) = a^T h(x) = h^T(x)a$

   where $a \in R^m$, $x \in R^n$

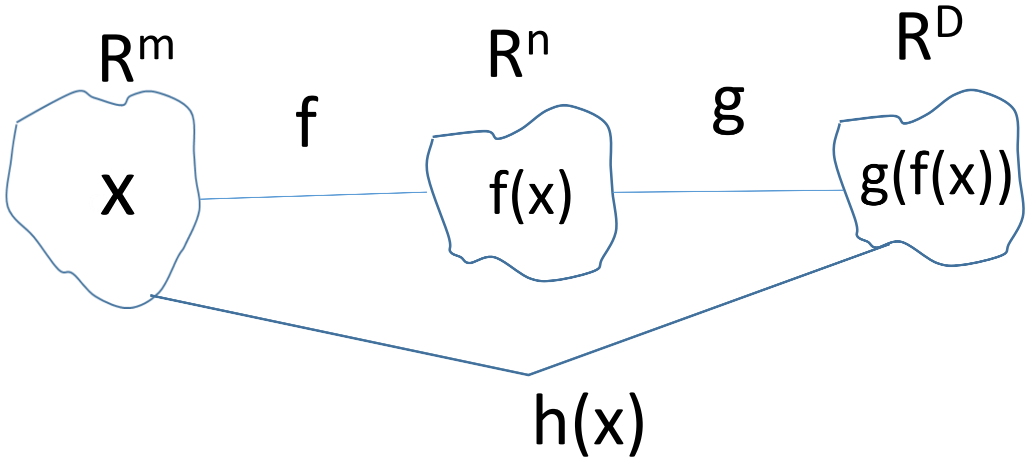
   Then, $\nabla_x f(x) = D_x^T(h)a$, $D_x(h) \in R^{mxn}$ - Jacobian of h

5. Let $h(x) = (h_1(x), h_2(x), \dots h_m(x))^T$, $A \in R^{mxn}$. Let $f(x) = h^T(x)Ah(x)$

   $\nabla_x f(x) = 2D_x^T(h)Ax$

6. $h(x) = g(f(x)) = g \circ f(x)$

   Then $D_x(h) = D_x(g)D_x(f)$



TAYLOR SERIES EXPANSION: f: R -> R

- Let x, z ∈ R

$$f(x + z) = f(x) + \frac{df}{dx} z + \frac{1}{2}\frac{d^2 f}{dx^2} z^2 + \ldots + \frac{1}{k!}\frac{d^k f}{dx^k} z^k + \ldots$$

- This an infinite series. By truncating at the k^n degree term in z, we get k^n order approximation

- We would be often interested in first and second order expansion

TAYLOR SERIES EXPANSION: f: R^n -> R

- $f(x + z) \approx f(x) + [\nabla_x f(x)]^T z + \frac{1}{2} z^T \nabla_x^2 f(x) z$

- Since $[\nabla_x f(x)]^T = D_x(f)$

- $f(x + z) \approx f(x) + D_x(f) z + \frac{1}{2} z^T \nabla_x^2 f(x) z$

TAYLOR SERIES EXPANSION: f: R^n -> R^m

- $f(x) = (f_1(x), f_2(x), \dots f_m(x))^T$; $x, z \in R^n$
- $f(x+z) \approx f(x) + D_x(f)z + \frac{1}{2} D_x^2(f, z)$

where $D_x(f) = \left[\frac{\partial f_i}{\partial x_j}\right] \in R^{mxn}$ Jacobian matrix

and

$$D_x^2(f, z) = \begin{bmatrix} z^T \nabla_x^2 f_1(x) z \\ z^T \nabla_x^2 f_2(x) z \\ \vdots \\ z^T \nabla_x^2 f_m(x) z \end{bmatrix}$$

with $\nabla_x^2 f_k(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] \in R^{nxn}$ the Hessian of $f_k(x)$

FIRST AND SECOND VARIATION: f: R^n -> R

- Let $\delta x = (\delta x_1, \delta x_2, \ldots \delta x_n)^T$ be a small increment or perturbation of x
- Let $\Delta f(x)$ be the resulting change f(x) induced by increment in x
- By Taylor Series expansion

$$f(x + \delta x) \approx f(x) + [\nabla_x f\,(x)]^T \delta x + \frac{1}{2}(\delta x)^T [\nabla_x^2 f(x)] \delta x$$

$$\approx f(x) + \delta f + \delta^{(2)} f(x)$$

where $\delta f = [\nabla_x f\,(x)]^T \delta x = \langle \nabla_x f\,(x), \delta x \rangle$ is called the <u>first variation</u> of f(x) and $\delta^{(2)} f(x) = \frac{1}{2}(\delta x)^T [\nabla_x^2 f(x)] \delta x = \frac{1}{2}\langle \delta x, \nabla_x^2 f(x) \delta x \rangle$ is called the <u>second variation</u> of f

FIRST VARIATION: f: $R^n \rightarrow R^m$

- Let $f(x) = (f_1(x), f_2(x), \ldots f_m(x))^T$ $x \in R^n$
- The first variation δf is a vector in R^m given by

$$\delta f = \begin{bmatrix} \delta f_1 \\ \delta f_2 \\ \vdots \\ \delta f_m \end{bmatrix} = \begin{bmatrix} <\nabla_x f_1, \delta x> \\ <\nabla_x f_2, \delta x> \\ \vdots \\ <\nabla_x f_m, \delta x> \end{bmatrix} = D_x(f)\delta x$$

EXAMPLES

1. $f(x) = \langle a, x \rangle \Rightarrow \delta f = \langle a, \delta x \rangle = \langle \delta x, a \rangle$

2. $f(x) = \frac{1}{2}\langle x, Ax \rangle \Rightarrow \delta f = \langle Ax, \delta x \rangle$, A is symmetric

3. $f(x) = (z - Hx)^T(z - Hx) \Rightarrow \delta f = \langle H^T(Hx - z), \delta x \rangle$

EXERCISE

4.1 Let $x = (x_1, x_2)^T$, h: $R^2 \rightarrow R^2$ given

$$h(x) = \begin{pmatrix} h_1(x) \\ h_2(x) \end{pmatrix} = \begin{pmatrix} e^{x_1} + e^{x_2} \\ x_2^2 + x_2^2 \end{pmatrix}$$

Let $x_c = (1, 1)^T$. Compute the second – order Taylor approximation of h(x) around x_c

4.2) Compute the first variation of

1) $(Z - Hx)^T W (z - Hx)$

2) $(x - x_b)^T B^{-1} (x - x_b)$

4.3) Verify $\nabla_x f(x) = D_x^T(h)a$ when $f(x) = a^T h(x)$

4.4) Verify $D_x(h) = D_x(g) D_x(f)$ when $h(x) = g(f(x)) = g \circ f(x)$

4.5) Compute the gradient and Hessian of

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2}(Z - Hx)^T R^{-1} (Z - Hx)$$

REFERENCES

1. T. M. Apostol (1957) Mathematical Analysis, *Addison-Wesley*