1 Linear Approximation

The linear approximation of $f(x)$ at a point a is the linear function:

$$L(x) = f(a) + f'(a)(x - a).$$

The linear approximation of $f(x, y)$ at (a, b) is the linear function:

$$L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

the linearization can be written more compactly using the **gradient** as:

$$L(\mathbf{\tilde{x}}) = f(\mathbf{\tilde{a}}) + \nabla f(\mathbf{\tilde{a}}) \cdot (\mathbf{\tilde{x}} - \mathbf{\tilde{a}}).$$

1.1 Chain Rule

The single variable chain rule is given by:

$$\frac{dy}{dx} = \frac{df}{du} \cdot \frac{du}{dx}$$

For a multivariate function, the chain rule is:

$$\frac{\partial y}{\partial x_i} = \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial x_i} + \frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial x_i} + \frac{\partial f}{\partial w} \cdot \frac{\partial w}{\partial x_i} + \cdots$$

The matrix form of the multivariate chain rule is expressed using Jacobian matrices:

$$D(\mathbf{y})(\mathbf{x}) = Df(\mathbf{u})(\mathbf{x}) \cdot Dg(\mathbf{x})$$

where $Df(\mathbf{u})$ and $Dg(\mathbf{x})$ are the Jacobian matrices of f and g respectively.

1.2 Tangent line, tangent plane & tangent hyperplane

The equation of the tangent line to the curve $y = f(x)$ at the point $(a, f(a))$ is:

$$y = f(a) + f'(a)(x - a)$$

The equation of the tangent plane to the surface $z = f(x, y)$ at the point $(a, b, f(a, b))$ is:

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

Here, f_x and f_y denote the partial derivatives of f with respect to x and y, respectively.

For a function $f(x_1, x_2, \ldots, x_n)$ at a point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, the tangent hyperplane is given by:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

where $\nabla f(\mathbf{a})$ is the gradient of f at $\mathbf{a}$ and $\mathbf{x}$ is the vector of variables.

1.3 Derivatives of Vector Functions

Let $F : \mathbb{R}^n \to \mathbb{R}^m, F(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix}$ be a vector function of the variables x $= (x_1, ..., x_n)$.

Recall that the derivative of the vector function F with respect to the vector of variables $\tilde{x}$ is defined as

$$\frac{\partial \tilde{F}}{\partial \tilde{x}} = J_F(\tilde{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\tilde{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\tilde{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\tilde{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\tilde{x}) \end{bmatrix}$$

The second derivative of the function $f : \mathbb{R}^n \to \mathbb{R}$ (here m $= 1$) is given as

$$\frac{\partial^2 f}{\partial \tilde{x}^2} = \frac{\partial}{\partial \tilde{x}} \left(\frac{\partial f}{\partial \tilde{x}} \right)^T$$

.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex on D, if

$$f(t\tilde{x} + (1 - t)\tilde{y}) \leq tf(\tilde{x}) + (1 - t)f(\tilde{y})$$

for all $\tilde{x}, \tilde{y} \in D$ and for all $t \in [0, 1]$. The function f is concave on D, if the function $-f$ is convex on D.

1.4 Rules for Differentiating Vector Functions

1. $\frac{\partial \tilde{x}}{\partial \tilde{x}} = I_n$
2. If $A \in \mathbb{R}^{m \times n}$, then $\frac{\partial A\tilde{x}}{\partial \tilde{x}} = A$.
3. If $\tilde{a} \in \mathbb{R}^n$, then $\frac{\partial \tilde{a}^T \tilde{x}}{\partial \tilde{x}} = \tilde{a}^T$.
4. If $A \in \mathbb{R}^{n \times n}$, then $\frac{\partial (\tilde{x}^T A \tilde{x})}{\partial \tilde{x}} = \tilde{x}^T (A + A^T)$.
5. If $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix, then $\frac{\partial (\tilde{x}^T A \tilde{x})}{\partial \tilde{x}} = 2\tilde{x}^T A$.
6. $\frac{\partial \|\tilde{x}\|^2}{\partial \tilde{x}} = 2\tilde{x}^T$.
7. If $\tilde{z} = \tilde{z}(\tilde{x})$ and $\tilde{y} = \tilde{y}(\tilde{x})$, then $\frac{\partial (\tilde{y}^T \tilde{z})}{\partial \tilde{x}} = \tilde{y}^T \frac{\partial \tilde{z}}{\partial \tilde{x}} + \tilde{z}^T \frac{\partial \tilde{y}}{\partial \tilde{x}}$.
8. If $G : D_G \subseteq \mathbb{R}^m \to \mathbb{R}^n$ and $F : D_F \subseteq \mathbb{R}^n \to \mathbb{R}^p$ and $H = F \circ G$, then $\frac{\partial H}{\partial \tilde{x}} = \frac{\partial F}{\partial G}(\tilde{G}(\tilde{x}))\frac{\partial G}{\partial \tilde{x}}$.

1.5 Linear Approximation of Vector Functions

$$f(X(p_1, p_2, \ldots, p_m), S(p_1, p_2, \ldots, p_m)) \approx k_n(p_1, p_2, \ldots, p_m)$$

$$X = \begin{bmatrix} X_1(p_1, \ldots, p_m) \\ \vdots \\ X_n(p_1, \ldots, p_m) \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_n \end{bmatrix}, \quad p = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix}$$

$$f(X(p_0 + \delta)) \approx f(X(p_0)) + df$$

$$df = \frac{\partial f}{\partial X}\frac{\partial X}{\partial p}\delta = J_f \delta, \quad \text{take expansion of } \frac{\partial X}{\partial p} \text{ around } p_0$$

$$dX = \begin{bmatrix} \frac{\partial X_1}{\partial p_1} & \frac{\partial X_1}{\partial p_2} & \cdots & \frac{\partial X_1}{\partial p_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial X_n}{\partial p_1} & \frac{\partial X_n}{\partial p_2} & \cdots & \frac{\partial X_n}{\partial p_m} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} \frac{\partial X_1}{\partial p_j}\delta_j \\ \vdots \\ \sum_{j=1}^{m} \frac{\partial X_n}{\partial p_j}\delta_j \end{bmatrix}$$

$$J_f = \frac{\partial f}{\partial X}\frac{\partial X}{\partial p}$$

$$f(X(p_0 + \delta)) \approx f(X(p_0)) + J_f \delta$$