Calculus Prerequisites for Optimization

1 Derivatives and Gradients

Overview

For a single-variable function $f(x)$, the derivative $f'(x)$ or $\frac{df}{dx}$ measures the rate of change. In optimization, critical points where $f'(x) = 0$ are candidates for optima. For a multi-variable function $f(x_1, x_2, \ldots, x_n)$, the gradient is a vector of partial derivatives:

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

A necessary condition for optimality is $\nabla f(x) = 0$.

Example

Consider the function:

$$f(x_1, x_2) = x_1^2 + 2x_2^2 + x_1 x_2$$

- **Compute the gradient ∇f:**

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2, \quad \frac{\partial f}{\partial x_2} = 4x_2 + x_1$$

$$\nabla f = \begin{bmatrix} 2x_1 + x_2 \\ 4x_2 + x_1 \end{bmatrix}$$

- **Find points where $\nabla f = 0$:**

$$2x_1 + x_2 = 0 \quad (1)$$

$$x_1 + 4x_2 = 0 \quad (2)$$

From (1), $x_2 = -2x_1$. Substitute into (2):

$$x_1 + 4(-2x_1) = x_1 - 8x_1 = -7x_1 = 0 \implies x_1 = 0$$

$$x_2 = -2 \cdot 0 = 0$$

Critical point: $(x_1, x_2) = (0, 0)$.

2 Hessian

Overview

The Hessian is a matrix of second-order partial derivatives:

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{i,j=1,\ldots,n}$$

In optimization, the Hessian is used for second-order conditions:

- If $\nabla^2 f(x)$ is positive definite (all eigenvalues positive), x is a local minimum.
- If negative definite, x is a local maximum.

Example

For the function $f(x_1, x_2) = x_1^2 + 2x_2^2 + x_1 x_2$:

- **Compute the Hessian $\nabla^2 f$:**

$$\frac{\partial^2 f}{\partial x_1^2} = 2, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 1, \quad \frac{\partial^2 f}{\partial x_2 \partial x_1} = 1, \quad \frac{\partial^2 f}{\partial x_2^2} = 4$$

$$\nabla^2 f = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

- **Check if $(0, 0)$ is a local minimum:** Compute eigenvalues of the Hessian:

$$\det(\nabla^2 f - \lambda I) = \det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 4-\lambda \end{bmatrix} = (2-\lambda)(4-\lambda) - 1 = \lambda^2 - 6\lambda + 7 = 0$$

$$\lambda = \frac{6 \pm \sqrt{8}}{2} = 3 \pm \sqrt{2} \quad (\lambda_1 \approx 4.414, \quad \lambda_2 \approx 1.586)$$

Both eigenvalues are positive, so $\nabla^2 f$ is positive definite, and $(0, 0)$ is a local minimum.

3 Chain Rule and Product Rule

Overview

- **Chain Rule:** For a composite function $f(g(x))$:

$$\frac{d}{dx} f(g(x)) = f'(g(x)) \cdot g'(x)$$

- **Product Rule:** For $f(x) = u(x)v(x)$:

$$f'(x) = u'(x)v(x) + u(x)v'(x)$$

These rules are essential for computing gradients and Hessians of complex functions.

Example

Consider the function:

$$f(x) = \sin(x^2 + 1)$$

Compute $f'(x)$:

- Let $u = x^2 + 1$, so $f(x) = \sin(u)$.
- By the chain rule: $f'(x) = \frac{df}{du} \cdot \frac{du}{dx}$.

$$\frac{df}{du} = \cos(u) = \cos(x^2 + 1), \quad \frac{du}{dx} = 2x$$

$$f'(x) = \cos(x^2 + 1) \cdot 2x$$

4 Taylor Series

Overview

The Taylor series approximates a function. For a single-variable function $f(x)$ around x_0:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \cdots$$

For multi-variable functions:

$$f(x) \approx f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T\nabla^2 f(x_0)(x - x_0) + \cdots$$

This is used in Newton's method and convergence analysis.

Example

Approximate $f(x) = e^x$ around $x = 0$ up to second order:

$$f(x) \approx f(0) + f'(0)x + \frac{1}{2}f''(0)x^2$$

- $f(x) = e^x$:

$$f(0) = e^0 = 1, \quad f'(x) = e^x \implies f'(0) = 1, \quad f''(x) = e^x \implies f''(0) = 1$$

$$f(x) \approx 1 + x + \frac{1}{2}x^2$$

5 Optimality Conditions

Overview

- **First-Order Condition:** For a point x^* to be optimal:

$$\nabla f(x^*) = 0$$

- **Second-Order Condition:**
  - For a local minimum: $\nabla^2 f(x^*)$ must be positive definite.
  - For a local maximum: $\nabla^2 f(x^*)$ must be negative definite.

Example

For $f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2$, approximate around $(0, 0)$ up to second order (Taylor series):

$$f(x) \approx f(0) + \nabla f(0)^T x + \frac{1}{2} x^T \nabla^2 f(0) x$$

- $f(0, 0) = 0$
- Gradient: $\nabla f = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix} \implies \nabla f(0, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
- Hessian: $\nabla^2 f = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$$f(x_1, x_2) \approx 0 + 0 + \frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \frac{1}{2}(2x_1^2 + x_1x_2 + x_1x_2 + 2x_2^2) = x_1^2 + x_1x_2 + x_2^2$$

(Since the function is quadratic, the approximation is exact.)

6 Final Notes

These calculus concepts are essential for optimization algorithms in *Numerical Optimization* by Nocedal and Wright:

- Gradients and Hessians define optimality conditions.
- Chain and product rules help compute derivatives of complex functions.
- Taylor series are used in Newton's method and convergence analysis.