Some Calculus Prerequisites

Andrew Nobel

August, 2020

The Gamma Function and Stirling's Approximation

Gamma function: Defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$ for $\alpha > 0$

(1) $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$

(2) $\Gamma(\alpha + 1) = \alpha\, \Gamma(\alpha)$

(3) $\Gamma(n) = (n-1)!$

(4) $\log \Gamma(\alpha)$ is convex.

***Stirling's Approximation**

(1) $\Gamma(\alpha + 1) \sim \sqrt{2\pi\alpha} \left(\frac{\alpha}{e}\right)^{\alpha}$

(2) $n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^{n} \delta_n$ where $e^{1/(12n+1)} \leq \delta_n \leq e^{1/(12n)}$

Taylor's Theorem in One Dimension

Fact: If $g : \mathbb{R} \to \mathbb{R}$ is k-times differentiable with derivatives $g^{(1)}, \ldots, g^{(k)}$ then for each $x, h \in \mathbb{R}$

$$g(x+h) \ = \ \sum_{j=0}^{k-1} \frac{g^{(j)}(x)}{j!} h^j \ + \ \frac{g^{(k)}(\tilde{x})}{k!} h^k$$

where $g^{(0)} = g$ and $\tilde{x}$ lies between x and $x+h$. If $g^{(k)}$ is continuous then

$$g(x+h) \ = \ \sum_{j=0}^{k-1} \frac{g^{(j)}(x)}{j!} h^j \ + \ o(h^k)$$

Idea: Taylor's theorem gives a local polynomial approximation of g in a neighborhood of the point x

Taylor's Theorem in One Dimension

Equivalent form of the theorem: under the same conditions, for all $x, y \in \mathbb{R}$

$$g(y) \;=\; \sum_{j=0}^{k-1} \frac{g^{(j)}(x)}{j!}(y-x)^j \;+\; \frac{g^{(k)}(\tilde{x})}{k!}(y-x)^k$$

where $\tilde{x}$ lies between y and x.

Note

- Other versions of the theorem handle cases where g is defined on an interval in $\mathbb{R}$
- Taylor's theorem is a consequence of the mean value theorem from calculus

Inequalities from Calculus

Problem: Given differentiable functions $f, g : I \to \mathbb{R}$ defined on interval $I \subseteq \mathbb{R}$, we wish to show that $f(x) \geq g(x)$ for each $x \in I$. One solution

- Consider the difference $h(x) := f(x) - g(x)$
- Find $c \in I$, possibly an endpoint of I, such that $h(c) \geq 0$ and

$$h'(x) \text{ is } \begin{cases} \geq 0 & \text{if } x \geq c \\ \leq 0 & \text{if } x \leq c \end{cases}$$

Alternatively, expand h in Taylor series around a point $c \in I$ (usually where $h^{(0)}(c) = \cdots = h^{(k)}(c) = 0$) and then examine the remainder term

Also: Direct application of Taylor's theorem and examination of the remainder can yield polynomial upper and lower bounds on complex functions

Inequalities from Calculus: Examples

1. For all x, $1 + x \;\le\; e^x$
2. For $x \ge 0$, $\log(1+x) \;\le\; x - x^2/2 + x^3/2$
3. For $0 \le x < 1$, $\log(1-x) \;\le\; -x - x^2/2$
4. For $x \ge 0$,
$$(1+x)\log(1+x) - x \;\ge\; \frac{x^2}{2 + 2x/3}$$
5. For $x \ge 0$,
$$(x-1)^2 \;\le\; \left(\frac{4+2x}{3}\right)(x\log x - x + 1)$$

Multivariate Differentiation: Total Derivative

Definition: A function $f : \mathbb{R}^d \to \mathbb{R}^k$ is *differentiable* at $x \in \mathbb{R}^d$ if there exists a matrix $A \in \mathbb{R}^{k \times d}$ such that

$$\lim_{h \to 0} \frac{||f(x+h) - f(x) - Ah||}{||h||} \; = \; 0$$

which can be written in the equivalent form

$$f(x+h) \; = \; f(x) + Ah + o(||h||)$$

The (unique) matrix A satisfying these conditions is called the *total derivative* of f at x, and denoted by $Df(x)$ or $\dot{f}(x)$

Total Derivatives

First Examples: Consider a function $f : \mathbb{R}^d \to \mathbb{R}^k$

- If $d = k = 1$ then $Df(x) = f'(x)$ coincides with ordinary derivative
- If $f(x) = c$ is constant then $Df(x) = \mathbf{0}$ is the $k \times d$ zero matrix
- If $f(x) = Bx$ is linear then $Df(x) = B$
- If $f(x) = x^t V x$ where $V \in \mathbb{R}^{d \times d}$ is symmetric then $DF(x) = 2x^t V$

Chain Rule: If $f : \mathbb{R}^d \to \mathbb{R}^k$ is differentiable at x and $g : \mathbb{R}^k \to \mathbb{R}^l$ is differentiable at $f(x)$, then $g \circ f$ is differentiable at x and

$$D(g \circ f)(x) \; = \; Dg(f(x)) \, Df(x)$$

Jacobians

Note that $f : \mathbb{R}^d \to \mathbb{R}^k$ can be written $f = (f_1, \ldots, f_k)$ where $f_i : \mathbb{R}^d \to \mathbb{R}$

Definition: The *Jacobian* of f at x is the $k \times d$ matrix of partial derivatives

$$J_f(x) \;=\; \left[\frac{\partial f_i}{\partial x_j}(x) : 1 \leq i \leq k, 1 \leq j \leq d \right]$$

Fact: Jacobians and Total Derivatives

(a) If f is differentiable at x then $J_f(x)$ exists and is equal to $Df(x)$

(b) If the Jacobian J_f exists and is continuous at x then f is differentiable at x and $J_f(x) = Df(x)$

Gradients and Hessians

Definition. Let $f : \mathbb{R}^d \to \mathbb{R}$

- The *gradient* of f at x is the $d \times 1$ vector of partial derivatives

$$\nabla f(x) \;=\; \left(\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_d}(x) \right)^t$$

  Note that $\nabla f(x) = Df(x)^t$ when the derivative exists

- The *Hessian* of f at x is the $d \times d$ matrix of second partial derivatives

$$\nabla^2 f(x) \;=\; \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x) : 1 \leq i, j \leq d \right]$$

  If the second partials are continuous, then $\nabla^2 f(x)$ is symmetric

Multivariate Taylor's Theorem I

Fact: If $f : \mathbb{R}^d \to \mathbb{R}^k$ has continuous partial derivatives $\partial f_i / \partial x_j$ at each point in $\mathbb{R}^d$ then for every $x, h \in \mathbb{R}^d$

$$f(x+h) \;=\; f(x) \;+\; \langle \nabla f(\tilde{x}), h \rangle$$

where $\tilde{x} = x + \alpha h$ for some $\alpha \in [0,1]$. In particular, we have

$$f(x+h) \;=\; f(x) \;+\; \langle \nabla f(\tilde{x}), h \rangle \;+\; o(||h||)$$

Multivariate Taylor's Theorem II

Fact: If $f : \mathbb{R}^d \to \mathbb{R}$ has continuous second partial derivatives $\partial^2 f / \partial x_i \partial x_j$ at each point in $\mathbb{R}^d$ then for every $x, h \in \mathbb{R}^d$

$$f(x+h) \;=\; f(x) \;+\; \langle \nabla f(x), h \rangle \;+\; \frac{1}{2} h^t \nabla^2 f(\tilde{x}) h$$

where $\tilde{x} = x + \alpha h$ for some $\alpha \in [0,1]$. In particular, we have

$$f(x+h) \;=\; f(x) \;+\; \langle \nabla f(x), h \rangle \;+\; \frac{1}{2} h^t \nabla^2 f(x) h \;+\; o(||h||^2)$$