Mathematics 250: Lecture 16

Second-order Approximations

Dan Sloughter

Furman University

February 26, 2016

The single-variable case

- Recall (the second derivative test): If x_0 is a critical point of $f : \mathbb{R} \to \mathbb{R}$, then
  - $f''(x_0) < 0$ implies f has a local maximum at x_0,
  - $f''(x_0) > 0$ implies f has a local minimum at x_0,
  - $f''(x_0) = 0$ gives us no information about the behavior of f at x_0.
- If f'' is continuous at x_0, then we may use the following result to prove the second derivative test.
  - Suppose f is twice differentiable on an open interval (a, b).
  - Then for any points x_0 and x in (a, b), there exists a point x^* between x_0 and x such that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x^*)(x - x_0)^2.$$

The mean-value theorem

- Single-variable mean-value theorem
  - Suppose f is differentiable on (a, b) and continuous on $[a, b]$.
  - Then there exists a point c in (a, b) for which $f(b) = f(a) + f'(c)(b - a)$.
- Several-variable case:
  - Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at each point on the line segment from X to X_0.
  - Then there exists a point X^* on the line segment between X_0 and X^* such that

$$f(X) = f(X_0) + \nabla f(X^*) \cdot (X - X_0).$$

Proof

- Let $K(t) = X_0 + t(X - X_0)$, $0 \leq t \leq 1$.
- Let $g(t) = f(K(t))$.
- Then $g : \mathbb{R} \to \mathbb{R}$ is continuous on $[0, 1]$ and differentiable on $(0, 1)$.
- Hence, by the one-variable mean-value theorem, there exists a point t^* in $(0, 1)$ for which

$$g'(t^*) = \frac{g(1) - g(0)}{1 - 0} = f(X) - f(X_0).$$

- Now, by the chain rule,

$$g'(t^*) = \nabla f(K(t^*)) \cdot K'(t^*) = \nabla f(K(t^*)) \cdot (X - X_0).$$

- Hence, with $X^* = K(t^*) = X_0 + t^*(X - X_0)$,

$$f(X) = f(X_0) + \nabla f(X^*) \cdot (X - X_0).$$

Theorem

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $\nabla f(X) = \mathbf{0}$ for all X in $B(X_0; r)$, $r > 0$.
- Then there exists a scalar c such that $f(X) = c$ for all X in $B(X_0; r)$.
- Proof:
  - Let $c = f(X_0)$.
  - For any point X in $B(X_0; r)$, there exists a point X^* in $B(X_0; r)$ for which

$$f(X) = f(X_0) + \nabla f(X^*) \cdot (X - X_0).$$

  - But $\nabla f(X^*) = \mathbf{0}$, so $f(X) = f(X_0)$.

Second derivatives

- We say $f : \mathbb{R}^n \to \mathbb{R}$ is *twice differentiable* at a point X_0 if each first partial derivative of f is differentiable at X_0.
- Technical points:
  - If f is twice differentiable, then the mixed second-order partial derivatives of f are equal (even if they are not continuous).
  - If the second-order partial derivatives of f are continuous, then f is twice differentiable.

The Hessian

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at X_0.
- Let $[f''(X_0)]$ denote the $n \times n$ matrix with ij-th entry $\frac{\partial^2 f}{\partial x_j \partial x_i}(X_0)$.
- That is,

$$[f''(X_0)] = \begin{bmatrix} f_{11}(X_0) & f_{12}(X_)) & \cdots & f_{1n}(X_0) \\ f_{21}(X_0) & f_{22}(X_0) & \cdots & f_{2n}(X_0) \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1}(X_0) & f_{n2(X_0)} & \cdots & f_{nn}(X_0) \end{bmatrix}$$

- Note: The rows of $[f''(X_0)]$ are the gradients of the partial derivatives of f.
- We call $[f''(X_0)]$ the *Hessian* of f at X_0.
- Note: If f is twice differentiable, then $[f''(X_0)]$ is a symmetric matrix.

Example

- Suppose $f(x, y, z) = 4xy^2z + x^2z^3$.
- Then

$$\nabla f(x, y, z) = (4y^2z + 2xz^3, 8xyz, 4xy^2 + 3x^2z^2).$$

- So

$$[f''(x, y, z)] = \begin{bmatrix} 2z^3 & 8yz & 4y^2 + 6xz^2 \\ 8yz & 8xz & 8xy \\ 4y^2 + 6xz^2 & 8xy & 6x^2z \end{bmatrix}.$$

- Hence, for example,

$$[f''(1, 1, 1)] = \begin{bmatrix} 2 & 8 & 10 \\ 8 & 8 & 8 \\ 10 & 8 & 6 \end{bmatrix}.$$

Second-order approximations

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at each point on the line segment from X_0 to X.
- Let $K(t) = X_0 + t(X - X_0)$, $0 \leq t \leq 1$.
- Let $g(t) = f(K(t))$.
- By Taylor's theorem, there exists a point t^* in $(0, 1)$ such that

$$\begin{aligned} g(1) &= g(0) + g'(0)(1-0) + \frac{1}{2}g''(t^*)(1-0)^2 \\ &= g(0) + g'(0) + \frac{1}{2}g''(t^*). \end{aligned}$$

- Now $g(1) = f(X)$ and $g'(0) = \nabla f(X_0) \cdot (X - X_0)$, so we have

$$f(X) = f(X_0) + \nabla f(X_0) \cdot (X - X_0) + \frac{1}{2}g''(t^*).$$

Second-order approximations (cont'd)

- Now $g'(t) = \nabla f(K(t)) \cdot K'(t)$.
- So

$$\begin{aligned} g''(t) &= \nabla f(K(t)) \cdot K''(t) + K'(t) \cdot \frac{d}{dt}\nabla f(K(t)) = (X - X_0) \cdot \frac{d}{dt}\nabla f(K(t)) \\ &= (X - X_0) \cdot \frac{d}{dt}(f_1(K(t)), f_2(K(t)), \ldots, f_n(K(t)) \\ &= (X - X_0) \cdot (\nabla f_1(K(t)) \cdot K'(t), \cdots, \nabla f_n(K(t)) \cdot K'(t)) \\ &= (X - X_0) \cdot (\nabla f_1(K(t)) \cdot (X - X_0), \cdots, \nabla f_n(K(t)) \cdot (X - X_0)) \\ &= [X - X_0]^T [f''(K(t))][X - X_0]. \end{aligned}$$

- Hence, if we let $X^* = K(t^*)$, we have

$$f(X) = f(X_0) + \nabla f(X_0) \cdot (X - X_0) + \frac{1}{2}[X - X_0]^T [f''(X^*)][X - X_0].$$

Taylor's theorem

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at each point on the line segment from X_0 to X.
- Then there exists a point X^* on the line segment between X_0 and X such that

$$f(X) = f(X_0) + \nabla f(X_0) \cdot (X - X_0) + \frac{1}{2}[X - X_0]^T [f''(X^*)][X - X_0].$$

The second derivative

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at X_0.
- We call the function $f''(X_0) : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f''(X_0)(X) = [X]^T [f''(X_0)][X]$$

  the *second derivative* of f at X_0.
- Moreover, we call

$$\begin{aligned} p(X) &= f(X_0) + f'(X_0)(X - X_0) + \frac{1}{2}f''(X_0)(X - X_0) \\ &= f(X_0) + \nabla f(X_0) \cdot (X - X_0) + \frac{1}{2}[X - X_0]^T [f''(X_0)][X - X_0] \end{aligned}$$

  the *second-order Taylor polynomial* for f at X_0.

Example

- Suppose $f(x,y,z) = 4xy^2z + x^2z^3$.
- Then

$$\begin{aligned} f''(1,1,1)(x,y,z) &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 2 & 8 & 10 \\ 8 & 8 & 8 \\ 10 & 8 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 2x + 8y + 10z \\ 8x + 8y + 8z \\ 10x + 8y + 6z \end{bmatrix} \\ &= 2x^2 + 8y^2 + 6z^2 + 16xy + 20xz + 16yz. \end{aligned}$$

Example (cont'd)

- Now $\nabla f(1,1,1) = (6,8,7)$.
- So the second-order Taylor polynomial for f at $(1,1,1)$ is

$$\begin{aligned} p(x,y,z) = 5 &+ 6(x-1) + 8(y-1) + 7(z-1) \\ &+ (x-1)^2 + 4(y-1)^2 + 3(z-1)^2 \\ &+ 8(x-1)(y-1) + 10(x-1)(z-1) + 8(y-1)(z-1). \end{aligned}$$

Example (cont'd)

- The following wxMaxima command will find this Taylor polynomial:

```
taylor(4*x*y^2*z + x^2*z^3,[x,y,z],[1,1,1],[2,2,2]);
```