The Hessian, Taylor's Theorem, Extrema, Lagrange Multipliers and Quadratic Forms*

Jonathan Lee

March 10, 2010

An online version of these notes is posted on my website at
http://math.stanford.edu/~jlee/math51/.

Differentials and Taylor's Theorem

- given a k-times differentiable function $f : \mathbb{R} \to \mathbb{R}$, we can define its k-th order *Taylor polynomial* at the point p to be

$$p_k(x) = \sum_{i=0}^{k} \frac{f^{(i)}(p)}{i!} \cdot (x-p)^i$$

- *Taylor's theorem* provides us an error estimate: if $f : \mathbb{R} \to \mathbb{R}$ is a $(k+1)$-times differentiable function, then there exists some number ξ between p and x such that

$$R_k(x,a) = f(x) - p_k(x) = \frac{f^{(k+1)}(\xi)}{(k+1)!} \cdot (x-a)^{k+1}$$

- given a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we can define its first-order *Taylor polynomial* at the point $p = (p_1, \ldots, p_n)$ to be

$$\begin{aligned} p_1(x) &= f(p) + Df(p) \cdot (x-p) \\ &= f(p) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot (x_i - p_i)\,; \end{aligned}$$

  note that this gives the equation for the tangent plane at the point p, which is used to compute *linear approximations*

*Alternatively, *All you ever wanted to know about vector calculus*

- similarly, if f is twice-differentiable, we can define its second-order *Taylor polynomial* at the point p to be

$$\begin{aligned} p_2(x) &= f(p) + Df(p) \cdot (x-p) + \frac{1}{2}\left[(x-p)^T \cdot Hf(p) \cdot (x-p)\right] \\ &= f(p) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot (x_i - p_i) + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i x_j}(p) \cdot (x_i - p_i)(x_j - p_j)\,, \end{aligned}$$

  where for notational sanity, we define the *Hessian matrix* to be the $n \times n$ matrix whose (i,j)-th entry is $\frac{\partial^2 f}{\partial x_i x_j}$

  – find the first- and second-order Taylor polynomials for

the function	at the point
$f(x,y) = 1/(x^2+y^2+1)$	$a = (0,0)$
$f(x,y) = 1/(x^2+y^2+1)$	$a = (1,-1)$
$f(x,y) = e^{2x}\cos(3y)$	$a = (0,\pi)$

- as before, *Taylor's theorem* provides an error estimate; the same formula holds, with the change that ξ is taken to be a point on the line segment connecting p and x

Quadratic Forms

- given an $n \times n$ symmetric matrix, we can define a *quadratic form* $Q : \mathbb{R}^n \to \mathbb{R}$ such that $x \mapsto x^T Ax$; this occurs as the quadratic term in the second-order Taylor polynomials described above

- a quadratic form is defined to be *positive definite, positive semi-definite, negative definite*, or *negative semi-definite* if the values it takes are positive, non-negative, negative or non-positive, respectively; if none of these holds, the form is defined to be *indefinite*

- *Remark:* by definition, if a quadratic form is positive definite, then it is also positive semi-definite

- *Proposition:* given a quadratic form $Q : \mathbb{R}^2 \to \mathbb{R}$ defined by $x \mapsto x^T Ax$, we can recognize its type according to the following table (to remember it, use the fact that similar matrices have the same determinant and trace, and assume all matrices are diagonalizable):

type of quadratic form Q	eigenvalues of A	A's determinant	A's trace
positive definite	all positive	positive	positive
positive semi-definite	all non-negative	zero	positive
negative definite	all negative	positive	negative
negative semi-definite	all non-positive	zero	negative
indefinite	one positive, one negative	negative	irrelevant
degenerate	both zero	zero	zero

Extrema of Functions

- given a function $f : \mathbb{R}^n \to \mathbb{R}$, know what its *extrema*, both *global* and *local*, are defined to be; this (and the following) works analogously to the single-variable case
- define a point p to be a *critical point* of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ if $Df(p) = 0$
- *Theorem:* local extrema of differentiable functions must be critical points
- *Theorem:* let p be a $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function; then

if $Hf(p)$ is	then p is a ... of f
positive definite	local minimum
negative definite	local maximum
neither of the above but still invertible	saddle point

  - find the point on the plane $3x - 4y - z = 24$ closest to the origin
  - determine the absolute extrema of
  $$f(x, y) = x^2 + xy + y^2 - 6y$$
  on the rectangle given by $x \in [-3, 3]$ and $y \in [0, 5]$
  - determine the absolute extrema of
  $$f(x, y, z) = \exp(1 - x^2 - y^2 + 2y - z^2 - 4z)$$
  on the ball
  $$\{(x, y, z) \in \mathbb{R}^3 \, : \, x^2 + y^2 - 2y + z^2 + 4z \leq 0\}$$

- by Heine-Borel, say that a subset X of $\mathbb{R}^n$ is *compact* if it is closed and bounded
- *Extreme Value Theorem:* any continuous $\mathbb{R}$-valued function on a compact topological space attains global minima and maxima

Lagrange Multipliers

- supposing that we have a continuously differentiable function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is defined to be the set of solutions to $g_1 = g_2 = \cdots = g_k = 0$ for some continuously differentiable functions $g_1, \ldots, g_k : \mathbb{R}^n \to \mathbb{R}$, then we can determine the critical points of f by:
  - solving the system of linear equations (in the variables $x, \lambda_1, \ldots, \lambda_k$)
  $$Df(x) = \lambda_1 Dg_1(x) + \cdots + \lambda_k Dg_k(x) \quad \text{and} \quad g_1(x) = g_2(x) = \cdots = g_k(x) = 0$$
  using elimination, cross-multiplication or other convenient methods; each solution x will be a critical point of f

- determining the points x where the functions $Dg_1, \ldots, Dg_k$ are linearly dependent, which in the case $k = 1$ (and $Dg = Dg_1$) amounts simply to finding those x such that $Dg(x) = 0$; only some of these points will be critical points and thus they all need to be inspected individually

- as a result, we now have three methods for determining the critical points of a continuously differentiable function $f : S \to \mathbb{R}$, for some specified set S:
  - if S is a curve (which has dimension 1), a surface (which has dimension 2), or in general, some subset of dimension n, then attempt to parametrize S by a function $g : \mathbb{R}^n \to S$ and subsequently compute the critical points of $f \circ g : \mathbb{R}^n \to \mathbb{R}$ — this method can quite often be unnecessarily brutal, requiring many error-prone calculations, which you may illustrate to yourself by finding the closest point on a line to a given point
  - find some geometric interpretation of the problem, draw some picture making it clear where the extrema should occur, and be creative computing the coordinates of such extrema — for example, in the case of finding the closest point on a given plane to a given point, this amounts to determining the intersection of the plane with the unique line perpendicular to the plane that crosses the specified point
  - use Lagrange multipliers

- good, wholesome, enriching entertainment:
  - find the largest possible sphere, centered around the origin, that can be inscribed inside the ellipsoid $3x^2 + 2y^2 + z^2 = 6$
  - the intersection of the planes $x - 2y + 3z = 8$ and $2z - y = 3$ forms a line; find the point on this line closest to the point $(2, 5, -1)$
  - the intersection of the paraboloid $z = x^2 + y^2$ with the plane $x + y + 2z = 2$ forms an ellipse; determine the highest and lowest (with respect to the z-coordinate) points on it