MATH2640 Introduction to Optimisation

2. Taylor's theorem, Gradients, Hessians, Extrema, Quadratic Forms: summary

2(A) Taylor's theorem in several variables.

(i) in 1 dimension, Taylor's theorem is $f(x+\delta x) = f(x) + \delta x f_x + \frac{1}{2!}(\delta x)^2 f_{xx} + \cdots$

(ii) In n dimensions, it is $f(\mathbf{X}+\delta\mathbf{x}) = f(\mathbf{X}) + \delta\mathbf{x}\cdot\nabla f + \frac{1}{2!}\delta\mathbf{x}\cdot H\delta\mathbf{x} + \cdots$
where $\delta\mathbf{x} = (\delta x_1, \delta x_2, \cdots \delta x_n)$, $\mathbf{X} = (X_1, X_2, \cdots X_n)$, and H is the Hessian matrix with components $H_{ij} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$

(iii) The Taylor series formula is derived by considering $f(\mathbf{X}+\lambda\delta\mathbf{x})$ as a function of λ, and expanding about $\lambda = 0$ using the Taylor expansion formula for a function of one variable, λ, together with the chain rule. Then putting $\lambda = 1$ gives the above formula.

(iv) In the two variable case, the Taylor series formula is

$$f(X+h, Y+k) = f(X,Y) + h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y} + \frac{1}{2}\left(h^2\frac{\partial^2 f}{\partial x^2} + 2hk\frac{\partial^2 f}{\partial x\partial y} + k^2\frac{\partial^2 f}{\partial y^2}\right) + \cdots .$$

The linear term is $hf_x + kf_y$, and the quadratic term is

$$\frac{1}{2}\begin{pmatrix} h & k \end{pmatrix}\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}\begin{pmatrix} h \\ k \end{pmatrix}, \quad \text{where} \quad H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

is called the Hessian matrix. Because $f_{xy} = f_{yx}$ by Young's theorem, the Hessian matrix is symmetric. In three dimensions, the Hessian matrix is

$$\begin{pmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{pmatrix}$$

which is also symmetric because $f_{xy} = f_{yx}$, $f_{xz} = f_{zx}$, and $f_{yz} = f_{zy}$.

2(B) Gradient vectors, normals and the tangent plane.

(i) $\nabla f(x,y) = (f_x, f_y)$ is in the direction in which f increases most rapidly. The unit vector in this direction is $\dfrac{\nabla f(x,y)}{|\nabla f(x,y)|} = \hat{\mathbf{n}}$. $-\nabla f(x,y)$ is in the direction of the rate of fastest decrease. Since $df = f_x dx + f_y dy$ if (dx, dy) is perpendicular to ∇f, so then $(dx, dy)\cdot\nabla f = 0$, then $df = 0$, i.e. if we move in the direction perpendicular to ∇f, f is constant. The direction perpendicular to ∇f is therefore the direction of the contours of constant f.

(ii) For implicitly defined functions, e.g. $g(x,y,z) = 0$, $\nabla g = (g_x, g_y, g_z)$ is in the direction of the normal to the surface $g(x,y,z) = 0$. The unit normal is $\dfrac{\nabla g(x,y,z)}{|\nabla g(x,y,z)|} = \hat{\mathbf{n}}$. This can be evaluated at any point (x_0, y_0, z_0) lying on the surface $g(x,y,z) = 0$.

(iii) The tangent plane to the surface $g(x,y,z) = 0$ at the point $\mathbf{x_0} = (x_0, y_0, z_0)$ is

$$(\mathbf{x} - \mathbf{x_0})\cdot\nabla g = 0,$$

where ∇g is evaluated at (x_0, y_0, z_0). If a surface is given by $z = f(x,y)$ then the tangent plane at a point where $f_x = 0$ and $f_y = 0$ is of the form $z =$ constant.

2(C) First order conditions (FOC)

(i) Local maxima and minima of $f(x,y)$ occur where the tangent plane is horizontal, so $f_x = 0$ and $f_y = 0$ there. In n dimensions,

$$\frac{\partial f}{\partial x_i} = 0, \quad i = 1, 2, \cdots, n, \quad \text{at} \quad \mathbf{x} = \mathbf{x_0},$$

if $\mathbf{x_o}$ is a local min or max. These n conditions give n equations for the n unknowns x_1, $x_2, \cdots x_n$. We solve these equations, which are called the first order conditions (FOC), to find the critical points where maxima and minima occur. The name first order conditions arises from the fact that they come from the linear terms in the Taylor expansion. There may be more than one solution of the FOC, i.e. more than one critical point.

(ii) Not all critical points are maxima or minimima. For example saddle points can occur as well. To classify each critical point we must look at the quadratic terms, i.e. look at the Hessian matrix.

(iii) Sometimes, all the quadratic terms may vanish, i.e. every element in H is zero. If this happens, we have to examine the cubic terms in the Taylor expansion to classify the critical point. In general this is complicated, but it is practical for some simple special cases.

2(D) Quadratic forms

(i) When the FOC are satisfied, the behaviour near the critical point is given by

$$f(\mathbf{X}+\delta\mathbf{x}) = f(\mathbf{X}) + \frac{1}{2!}\delta\mathbf{x}\cdot H\delta\mathbf{x} + \cdots .$$

The expression $\delta\mathbf{x}\cdot H\delta\mathbf{x}$ is called a quadratic form. To study quadratic forms we usually consider just $\mathbf{x}\cdot A\mathbf{x} = Q(\mathbf{x})$ where A is a symmetric matrix.

In the 2-dimensional case,

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad \text{and so} \quad Q = ax^2 + 2bxy + cy^2$$

is the quadratic form. In the 3-dimensional case

$$A = \begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}, \quad \text{and so} \quad Q = ax^2 + by^2 + cz^2 + 2hxy + 2gxz + 2fyz.$$

In any number of dimensions, quadratic forms fall into five classes.

(PD) If $Q > 0$ for all non-zero $\mathbf{x}$, Q is positive definite.

(ND) If $Q < 0$ for all non-zero $\mathbf{x}$, Q is negative definite.

(I) If Q can take either positive or negative values, Q is indefinite.

(PSD) If $Q \geq 0$ for all non-zero $\mathbf{x}$, then Q is positive semidefinite. Note there may be some nontrivial $\mathbf{x}$ which makes $Q = 0$.

(NSD) If $Q \leq 0$ for all non-zero $\mathbf{x}$, then Q is negative semidefinite. Note there may be some nontrivial $\mathbf{x}$ which makes $Q = 0$.

Condition PD implies condition PSD, and ND implies NSD, but not the other way round.

If A is the Hessian matrix, then case (PD) is sufficient to give a local minimum, since the quadratic terms cause f to increase as we move away from $\mathbf{X}$. Similarly, case (ND) is sufficient to give a local maximum. Case (I) is not a maximum or a minimum. In 2-dimensions case (I) corresponds to a saddle point. In case (PSD), if we are at a minimum, PSD must hold, but it may be that PSD holds but its not a local minimum. Similarly, at a maximum, NSD must hold but NSD is not sufficient to prove a maximum. We say PD and ND are sufficient conditions to establish a min, max respectively, and PSD and NSD are necessary conditions for there to be a min, max respectively.

(ii) To classify the quadratic form we first look at the *leading principal minors*, the LPMs. For a 3×3 symmetric matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

There are 3 LPM's, LPM_1, LPM_2 and LPM_3 defined by

$$LPM_1 = a_{11}, \quad LPM_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2,$$

$$LPM_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}^2) + a_{12}(a_{23}a_{13} - a_{12}a_{33}) + a_{13}(a_{12}a_{23} - a_{22}a_{13}),$$

so the vertical lines mean take the determinant. In the $n\times n$ case, we define n LPM's by extending in the obvious way. Assuming that the determinant $\det(A) \neq 0$, then the LPMs determine the signature of the matrix: *i)* if all $LPM_i > 0$, $i = 1, \ldots, n$, the matrix A is positive definite (PD); *ii)* if $LPM_1 < 0$, and the signs of the subsequent LPM_i alternate, (i.e., all the odd LPM_i are negative and the even LPM_i are positive), then A is (ND); *iii)* if there is any other sign pattern, A is indefinite.

(iii) However, if $LPM_n = \det(A) = 0$ for the $n \times n$ case the situation is more complicated, and A may be semi-definite but not definite. In this case, we have to find all the *principal minors*. A principal minor of order $n-k$ is found by deleting k columns and the same k rows and finding the determinant of what's left. If $k = 0$ there is just principal minor of order n, LPM_n, which is the determinant of A. If $k = 1$, there are n principal minors of order $n - 1$, since we can delete the row and column in n different ways. If **all** the principal minors are ≥ 0 A is positive semi-definite (PSD). If **all** the principal minors of odd order are ≤ 0 and **all** the principal minors of even order are ≥ 0, then A is negative semi-definite (NSD). Otherwise its indefinite.

(iv) Another way to classify quadratic forms is to find the *eigenvalues* of the matrix A. Since A is a symmetric matrix, all the eigenvalues are real. If all the eigenvalues are positive, A is positive definite. If all the eigenvalues are negative, A is ND. If some eigenvalues are zero and the rest positive, A is PSD. If some are zero and the rest negative, A is NSD. If any two eigenvalues have opposite sign, A is indefinite. To find the eigenvalues, we must solve the equation for λ given by

$$A = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{12} & a_{22} - \lambda & a_{23} \\ a_{13} & a_{23} & a_{33} - \lambda \end{vmatrix} = 0$$

in the 3×3 case, with the obvious generalization in the $n \times n$ case. In the 3×3 case this means solving a cubic equation, which will have 3 roots giving 3 eigenvalues. Solving cubic equations (or higher order equations) can be hard, but there are many computer packages which will find the eigenvalues of an $n \times n$ symmetric matrix very quickly.

FWN September 20, 2019