# Introduction to Optimisation Summary Notes:

January 19, 2020

## 1 Non-optimisation related differentiation.

There is simple derivatives of functions of multiple variables that I will assume by this point you know ($\frac{\partial}{\partial x}$) and things from other modules (i.e. Contour plots and planes).

### 1.1 *Total Derivatives:*

the total derivative of a function $z = z(x(t), y(t))$ with respect to t is the derivative:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

if instead $z = z(x(s,t), y(s,t))$ then the chain rules for several variables are:

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$
$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial s}\frac{\partial s}{\partial t}$$

### 1.2 *Gradient vectors and directional derivatives:*

If we have a function $f = f(x_1, x_2, \ldots, x_n)$ then the gradient vector is:

$$\text{grad} f = \nabla f = (f_{x_1}, f_{x_2}, \ldots, f_{x_n})$$

and the directional derivative in the direction $\mathbf{u}$ is:

$$D_{\mathbf{u}} f = \mathbf{u} \cdot \nabla f$$

the directional derivative is then greatest when $\nabla f$ is parallel to $\mathbf{u}$.

## 1.3 *Differentials:*

If $f(x,y,z)=0, g(x,y,z)=0$ then we have:

$$df=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy+\frac{\partial f}{\partial z}dz=0$$
$$dg=\frac{\partial g}{\partial x}dx+\frac{\partial g}{\partial y}dy+\frac{\partial g}{\partial z}dz=0$$

Then we we can use basic simultaneous equations to eliminate $dz$ to find a relation between $dx$ and $dy$ and can rearrage that to find $\frac{dy}{dx}$. If $f(x,y,z)=c, g(x,y,z)=d$ then we can use the same proccess to find how $x$ and $y$ will change if $z$ is slightly increased or decreased.

## 1.4 *Taylors theorem*

In one dimension, Taylors theorem is:

$$f(x+\delta x)=f(x)+\delta x f_x+\frac{1}{2!}(\delta x)^2 f_{xx}+\dots$$

which generalises to n dimensions as:

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

where $H$ is the hessian matrix with entry given by:

$$H_{ij}=\frac{\partial^2 f}{\partial x_i \partial x_j}$$

in two variables this is:

$$f(X+h,Y+k)=f(X,Y)+hf_x+kf_y+\frac{1}{2}(h^2f_{xx}+2hkf_{xy}+k^2f_{yy})+\dots$$

the linear term is

$$hf_x+kf_y$$

and the quadratic term:

$$\frac{1}{2}\begin{pmatrix} h & k \end{pmatrix}\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}\begin{pmatrix} h \\ k \end{pmatrix}$$

the 2x2 Matrix here is the Hessian matrix for two variables, in 3 dimensions it is:

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

## 1.5 *Tangent planes and First Order Conditions:*

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

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

where $\nabla g$ is evaluated at $(x_0, y_0, z_0)$
Local maxima and minima of $f(x, y)$ occur where the tangent plane is horizontal, so $f_x = 0$ and $f_y = 0$. if $\mathbf{x_0}$ is a local max or min then you solve the first order conditions for $\mathbf{x_0}$. Not all critical points are maxima or minima, in order to categorise them we look to the quadratic terms.

## 1.6 *Quadratic forms;*

When FOC are satisfied then the behavior near the critical point is given by:

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

$\delta\mathbf{x} \cdot H\delta\mathbf{x}$ is called the quadratic form. We will usually consider just $\mathbf{x} \cdot A\mathbf{x} = Q(\mathbf{x})$ where A is the symmetric matrix. In the 2 and 3 dimensional cases we have:

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

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

respectively.
There are 5 classes of quadratic forms:
Positive Definite: $Q > 0$ for all non-zero $\mathbf{x}$
Negative Definite: $Q < 0$ for all non-zero $\mathbf{x}$
Indefinite: If $Q$ can take either positive or negative values.
Positive-Semidefinite: $Q \geq 0$ for all non-zero $\mathbf{x}$
Negative-Semidefinite: $Q \leq 0$ for all non-zero $\mathbf{x}$
Positive and negative definite represent local minimum and maximum respectively. In dimensions Indefinite represents a saddle point.

To classify a quadratic form we look at their leading principal minors, For a $3 \times 3$ matrix these are:

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

$$\begin{aligned} LPM_1 &= a \\ LPM_2 &= \begin{vmatrix} a & b \\ d & e \end{vmatrix} \\ LPM_3 &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \end{aligned}$$

If all leading principal minors are greater then 0 then the matrix is positive definite, if $LPM_1 < 0$ and the signs the rest signs of the $LPM$s alternate then the matrix is negative definite. If there is any other sign pattern then it is indefinite.
If $\det A = 0$ then Onno has written some bulshit about principal minors that I don't get so I'm choosing not to cover it. Look at the eigenvalues.
Eigenvalue cases:
Positive definite if all eigenvalues are positive
Negative definite if all eigenvalues are negative
PSD if some eigenvalues are 0 and the rest positive
NSD if some eigenvalues are 0 and the rest negative
Indefinite if any 2 have opposite signs.

# 2 Production functions and unconstrained optimisation:

## 2.1 *Production functions:*

A firm produces Q items per year which it sells at a price $p$, the revenue is $R = pQ$. $Q$ is a function of variable $(x_1, x_2, \ldots, x_n)$, these variables are called the input bundle and represent things such as employees. The cost $C$ also depends on the input bundle. The profit is given by:

$$\begin{aligned} \Pi &= R - C \\ &= pQ(x_1, x_2, \ldots, x_n) - C(x_1, x_2, \ldots, x_n) \end{aligned}$$

To maximise this we solve the set of equations:

$$p\frac{\partial Q}{\partial x_i} = \frac{\partial C}{\partial x_i}, i = 1, \ldots, n$$

## 2.2 *The discriminating monopolists:*

A monopolists produces two products at a rate $Q_1$ and $Q_2$. The price obtained is not constant but reduces if the monopolist floods the market, so in a linear model $p_1 = a_1 - b_1Q_1$ and $p_2 = a_2 - b_2Q_2$ are the prices for the products. The cost is taken as $C = c_1Q_1 + c_2Q_2$. $a_i, b_i, c_i$ are all constants.The profit is then:

$$\begin{aligned}\Pi &= R - C = p_1Q_1 + p_2Q_2to - c_1Q_1 - c_2Q_2 \\ &= (a_! - c_1)Q_1 + (a_2 - c_2)Q_2 - b_1Q_1^2 - b_2Q_2^2\end{aligned}$$

so to maximise profit we find:

$$\begin{aligned}\Pi_{Q_1} &= (a_! - c_1) - 2b_1Q_1 = 0 \\ \Pi_{Q_2} &= (a_2 - c_2) - 2b_2Q_2 = 0\end{aligned}$$

Looking to the hessians we find that that $\Pi$ has a maximum here provided $b_1$ and $b_2$ are positive. We also require $a_i \geq c_i$.

## 2.3 The Cobb-Douglas function

Writing this is slowly killing me. I hate this module. The Cobb-Douglas function is:

$$Q(x_1, x_2) = x_1^a x_2^b$$

where $a > 0, b > 0$. The input bundle here is $(x_1, x_2)$ and we assume the price $p$ is independant of $Q$. If the cost is linear in the input bundle, $C = w_1x_1 + w_2x_2$ then the profit is:

$$\Pi = R - C = px_1^a x_2^b - w_1x_1 - w_2x_2$$

and the conditions for a stationary point are:

$$\begin{aligned}\Pi_{x_1} &= apx_1^{a-1}x_2^b = w_1 \\ \Pi_{x_2} &= bpx_1^a x_2^{b-1} = w_2\end{aligned}$$

giving the positive critical values $x_1^* = \frac{apQ^*}{w_1}$ and $x_2^* = \frac{bpQ^*}{w_2}$ where $Q^*$ is the production function at the critical value. Examining the Hessian:

$$\begin{pmatrix} a(a-1)px_1^{a-2}x_2^b & abpx_1^{b-1}x_2^{b-1} \\ abpx_1^{b-1}x_2^{b-1} & b(b-1)px_1^a x_2^{b-2} \end{pmatrix}$$

Which is negative definite if:

$$a(a-1)px_1^{a-2}x_2^b < 0$$
$$ab(1-a-b)p^2x_1^{2a-2}x_2^{2b-2} > 0$$

which requires $0 < a < 1$ and $0 < b < 1$ and $a+b < 1$. If these conditions are satisfied then the Cobb-Douglas function gives a profit maximising strategy.

# 3 Constrained Optimisation:

We did alot of this last year so I wont go too in depth with Largrange multipliers.

## 3.1 Single equality constraint

If we have $f(\mathbf{x})$ and $h(\mathbf{x}) = c$ then we find this through the Lagrangian:

$$\nabla f = \lambda \nabla h$$
$$\nabla(f - \lambda h) = 0$$

So solve for $\mathbf{x}$ using the Lagrangian and $h(\mathbf{x}) = c$.

## 3.2 Multiple equality constaints:

if we have have $m$ equality contraints:

$$h_i(\mathbf{x}) - c_i = 0$$

then we need $m$ Lagrange multipliers, and the Lagrangian is:

$$L = f - \lambda_1 h_1 - \lambda_2 h_2 \ldots - \lambda_m h_m$$

Then we solve the system of equations:

$$\frac{\partial L}{\partial x_i} = 0$$
$$h_i(\mathbf{x}) = c_i$$

## 3.3 Second order conditions

In order to classify the stationary points then we use second order conditions on the Lagrangian:

### 3.3.1 $2 \times 2$ with one linear constraint:

If we have

$$Q = ax^2 + 2bxy + cy^2$$

and then we have constraint:

$$ux + vy = 0$$

Then we have minimum if $(av^2 - 2buv + cu^2) > 0$ and maximum if $(av^2 - 2buv + cu^2) < 0$

### 3.3.2 Bordered Hessian

If we have $n$ variables in the Lagrangian with $m$ constraints then the Borded Hessian is:

$$H_B = \begin{pmatrix} 0 & B \\ B^T & H \end{pmatrix}$$

where $B$ is the matrix of the coefficients of the constraints and 0 is the zero matrix.
To classify a Bordered Hessian we look at the last $(n - m)$ LPM's starting with $LPM_{2m+1}$ abd going up to and including $LPM_{n+m} = \det H_B$
We classify this then as
Positive definite if the sign of $\det H_B = (-1)^m$ and successive LPM's have the same sign.
Negative definite if the sign of $\det H_B = (-1)^n$ and successive LPM's alternate in sign.
Otherwise it's indefinite.

## 3.4 Inequality Constraints:

Given a function $f(x, y)$ to maximise subject to $g(x, y) \leq b$ then we need $\nabla f$ is parallel to $\nabla g$ and not anti-parallel, i.e. $\nabla f = \lambda \nabla g$ with $\lambda > 0$. We call this then a binding constraint.

### 3.4.1 Complementary slackness condition:

We define the lagrangian $L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$ and this is solved by:

$$\frac{\partial L}{\partial x} = 0$$
$$\frac{\partial L}{\partial y} = 0$$
$$\lambda(g(x, y) - b) = 0$$

The last of these is called the complementary slackness condition.
In general if we have a function of n variables $f(\mathbf{x}) = f(x_1, \cdots, x_n)$ subject to m conditions:

$$g_1(\mathbf{x}) \leq b_1, \cdots, g_m(\mathbf{x}) \leq b_m$$

NOTE, THE CONSTANT IS LARGER THAN THE FUNCTION, THIS IS HOW THE RESTRAINTS WORK.
We then define the lagrangian:

$$L = f(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \cdots - \lambda_m g_m(\mathbf{x})$$

and solve the $n$ first order equality conditions and $m$ complementary slackness conditions:

$$\frac{\partial L}{\partial x_i} = 0$$
$$\lambda_i(g_i(\mathbf{x}) - b_i) = 0$$

We only care about solutions which satisfy:

$$\lambda_i \geq 0$$
$$g_i(\mathbf{x}) \leq b_i$$

We classify these using the same techniques as the equality constraints using the bordered Hessian. Note that we only allow the coefficients of the binding constraints in the $B$ part of the bordered Hessian.

### 3.4.2 Kuhn-Tucker Method:

If any of the constraints are simply positivity constraints, i.e.

$$x_i \geq 0$$

We define the Kuhn-Tucker Lagrangian, $\bar{L}$, to be the normal Lagrangian ignoring the positivity constraints. So the Lagrangian is then:

$$L = \bar{L} + \mu_1 x_1 + \cdots + \mu_n x_n$$

with $\mu_i \geq 0$ and $x_i \geq 0$. So the complementary slackness equations for the positivity constraints are:

$$x_i \frac{\partial \bar{L}}{\partial x_i} = 0$$

and the remaining CSC's are as before, but now can be expressed:

$$\lambda_i \frac{\partial \bar{L}}{\partial \lambda_i} = 0$$

### 3.4.3 Mixed Equalities and Inequalities:

This is a trivial to find by mixing methods from the equality and inequality above.

# 4 Final Remarks

I'm not gonna lie, I hate this module and writing this up took so long because I genuinely cannot stand this. I apologise if these notes are not as robust but I just feel frustrated at Onno for his lack of resources.