FHS Microeconomic Analysis Notes

Harry Folkard, Keble College

2022

Abstract

These are my microeconomic analysis notes made for my finals in 2022. They cover all topics. Feel free to use these notes and pass them on to others. Please note, however, that these have just been made by a student and not checked over. They likely contain errors, so it will be worth checking things for yourself. Thanks to Ines Morena de Barreda and Miguel Ballester - these notes are just my interpretation of their lectures and tutorials.

Contents

Linear Algebra

Linear Systems

Definition

A linear system is a system of n linear equations is of the form

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n &= b_2 \\ \vdots \quad\quad \vdots \quad\quad \vdots \quad &\quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

It can be written as:

$$A\boldsymbol{x} = \boldsymbol{b}$$

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

Solutions

A linear system $A\boldsymbol{x} = \boldsymbol{b}$ must either have

(1) exactly one solution.

(2) no solutions.

(3) infinitely many solutions.

If the number of equations $<$ number of unknowns,

- $A\boldsymbol{x} = \mathbf{0}$ has infinitely many solutions,
- $A\boldsymbol{x} = \boldsymbol{b}$ for any $\boldsymbol{b}$ has zero or infinitely many solutions,
- If $\text{rank}(A) =$ number of equations (rows) then $A\boldsymbol{x} = \boldsymbol{b}$ has infinitely many solutions for every $\boldsymbol{b}$.

If the number of equations $>$ number of unknowns,

- $A\boldsymbol{x} = \mathbf{0}$ has one or infinitely many solutions,
- For any $\boldsymbol{b}$, $A\boldsymbol{x} = \boldsymbol{b}$ has 0, 1, or infinitely many solutions,
- If $\text{rank}(A) =$ number of unknowns (columns), then $A\boldsymbol{x} = \boldsymbol{b}$ has 0 or 1 solution for every b.

If the number of equations $=$ number of unknowns,

- $A\boldsymbol{x} = \mathbf{0}$ has one or infinitely many solutions,
- For any given b, $A\boldsymbol{x} = \boldsymbol{b}$ has 0, 1, or infinitely many solutions,
- If $\text{rank}(A) =$ number of unknowns $=$ number of equations then $A\boldsymbol{x} = \boldsymbol{b}$ has exactly one solution for every b.
  - We call this matrix A **non-singular**.

Invertibility

Definition

Matrix A is invertible iff square and non-zero determinant,

$$A^{-1} \text{ exists iff } |A| \neq 0$$

Notes

Invertibility is equivalent to having one solution while non-invertibility is equivalent to having no or infinitely many solutions.

Homogeneity

Definition

When all the b_j's on the right-hand side are zero,

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= 0 \\ a_{21}x_1 + \ldots + a_{2n}x_n &= 0 \\ \vdots \qquad \vdots \qquad &\vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= 0 \end{aligned}$$

Notes

A homogenous linear system *always has at least one solution:* $x_1 = x_2 = \ldots = x_n = 0$, known as the trivial solution.

$A\boldsymbol{x} = \mathbf{0}$, with A square has non-trivial solutions iff $|A| = 0$.

If it has more unknowns than equations then it must have infinitely many solutions. Why?

- Because it always has one solution: $x_1 = x_2 = \ldots = x_n = 0$
- Hence if it doesn't have enough equations and we have one solution, to have more than one we must have infinitely many.

Gaussian & Gauss-Jordan Elimination

Row echelon form

Each row has more leading zeros than the row before it.

$$\begin{pmatrix} 1 & -0.4 & -0.3 & |130 \\ 0 & 1 & -0.25 & |125 \\ 0 & 0 & 1 & |300 \end{pmatrix} \Rightarrow \begin{cases} x_1 - 0.4x_2 - 0.3x_3 = 130 \\ x_2 - 0.25x_3 = 125 \\ x_3 = 300 \end{cases}$$

Reduced row echelon form

Each 'pivot' is a one and everything else is a zero.

$$\begin{pmatrix} 1 & 0 & 0 & |300 \\ 0 & 1 & 0 & |200 \\ 0 & 0 & 1 & |300 \end{pmatrix} \Rightarrow \begin{cases} x_1 = 300 \\ x_2 = 200 \\ x_3 = 300 \end{cases}$$

Example: Solve this system of equations via Gauss-Jordan Elimination

$$\left.\begin{aligned} w + 4x + 17y + 4z &= 38 \\ 2w + 12x + 46y + 10z &= 98 \\ 3w + 18x + 69y + 17z &= 153 \end{aligned}\right\} \Leftrightarrow \begin{pmatrix} 1 & 4 & 17 & 4 \\ 2 & 12 & 46 & 10 \\ 3 & 18 & 69 & 17 \end{pmatrix} \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 38 \\ 98 \\ 153 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 4 & 17 & 4 & | & 38 \\ 2 & 12 & 46 & 10 & | & 98 \\ 3 & 18 & 69 & 17 & | & 153 \end{pmatrix} \xrightarrow[R_3 = R_3 - 3R_1]{R_2 = R_2 - 2R_1} \begin{pmatrix} 1 & 4 & 17 & 4 & | & 38 \\ 0 & 4 & 12 & 2 & | & 22 \\ 0 & 6 & 18 & 5 & | & 39 \end{pmatrix}$$

$$\xrightarrow{R_3 = 2R_3 - 3R_2} \begin{pmatrix} 1 & 4 & 17 & 4 & | & 38 \\ 0 & 4 & 12 & 2 & | & 22 \\ 0 & 0 & 0 & 4 & | & 12 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 4 & 17 & 4 & | & 38 \\ 0 & 1 & 3 & 0.5 & | & 5.5 \\ 0 & 0 & 0 & 1 & | & 3 \end{pmatrix}$$

$$\xrightarrow[R_2 = R_2 - 0.5R_1]{R_1 = R_1 - 4R_3} \left(\begin{array}{cccc|c} 1 & 4 & 17 & 0 & 26 \\ 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right) \xrightarrow{R_1 = R_1 - 4R_2} \left(\begin{array}{cccc|c} 1 & 0 & 5 & 0 & 10 \\ 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right)$$

$$\begin{aligned} & w + 5y = 10 \\ & x + 3y = 4 \\ & z = 3 \end{aligned}$$

- We tend to call z : determined; x and w: basic variables; y : free variable.
- Note that it does have to be as neat as this, row one could be the same and then row 2 could be $\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$ since that still has more leading zeros than the row above. If a matrix has a row of all zeros, then all rows below it must be zeros for row echelon form (remember you can swap over rows).
- Also note that the zero matrix is also in reduced row echelon form (with no pivots).

Rank

Definition

The rank of A is the *maximal number of linearly independent rows or columns.*

The rank of A is the number of *nonzero rows in its row echelon form.*

Full Rank

An $n \times m$ matrix has full rank iff $\text{rank}(A) = \min\{n, m\}$.

Note it must always be the case that

- $\text{rank}(A) \leq n$
- $\text{rank}(A) \leq m$
- $\text{rank}(A) = \text{rank}(A^T)$

Implications for Linear systems

A system of linear equations with coefficient matrix A and augmented matrix A^* has solutions iff,

$$\text{rank}(A) = \text{rank}(A^*)$$

Linear independence iff full rank.

Example: Finding Rank

- Reduce to echelon form and see how many rows, or just look and see what the maximum number of linearly independent rows or column is.

Linear Combination

Definition

A vector $\boldsymbol{x}$ is a linear combination of a collection of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{\text{m}}$ in $\mathbb{R}^n$ if there exist real numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that $\boldsymbol{x} = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_m \boldsymbol{x}_m$.

Example 1: Show that $\boldsymbol{x}$ is a linear combination $\boldsymbol{u}$ and $\boldsymbol{v}$

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ \boldsymbol{v} = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix}, \ \boldsymbol{x} = \begin{pmatrix} 5 \\ 9 \\ 13 \end{pmatrix}$$

- $\boldsymbol{x}$ is a linear combination of $\boldsymbol{u}$ and $\boldsymbol{v}$, since $\boldsymbol{x} = 2\boldsymbol{u} + \boldsymbol{v}$.

Linear Independence

Definition

A collection of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ in $\mathbb{R}^n$ is linearly independent if the *only* scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that, $\alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \ldots + \alpha_m\boldsymbol{x}_m = \mathbf{0}$ are $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$

A collection of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{\mathrm{m}}$ in $\mathbb{R}^n$ is linearly independent if none of them is a linear combination of the others.

Notes

This second claim can be shown since if we suppose $\boldsymbol{x}_1$ were a linear combination of the others such that $\boldsymbol{x}_1 = \alpha_2\boldsymbol{x}_2 + ... + \alpha_m\boldsymbol{x}_m$, then if we set $\alpha_1 = -1$ in the equation then, $(-1)\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + ... + \alpha_m\boldsymbol{x}_m = 0$. This holds for nonzero values of α, hence the set $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_m$ is linearly dependent.

It is also useful to know that a set of n vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ in $\mathbb{R}^n$ (notice the matrix of these will be square) is linearly independent iff $\det(\boldsymbol{x}_1\boldsymbol{x}_2 \ldots \boldsymbol{x}_n) \neq 0$. This is the same as the matrix of the vectors being non-singular as well as invertible.

Further for any set of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ in $\mathbb{R}^n$ (which must produce a $n \times m$ matrix), if $m > n$ then the set of vectors is linearly dependent. (These vectors produce a matrix with more columns than rows, hence it will have a free variable and thus will have infinitely many solutions, all bar one of which are nonzero.)

Example 1: Show that these are not linearly independent

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ \boldsymbol{v} = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix}, \ \boldsymbol{x} = \begin{pmatrix} 5 \\ 9 \\ 13 \end{pmatrix}$$

- These three vectors are *not* linearly independent because $\boldsymbol{x}$ is a linear combination of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Example 2: Show that these are linearly independent

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ \boldsymbol{v} = \begin{pmatrix} 3 \\ 2 \\ 9 \end{pmatrix}, \ \boldsymbol{w} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

- Need to show that $r\boldsymbol{u} + s\boldsymbol{v} + t\boldsymbol{w} = 0$ only if $r = s = t = 0$

$$r\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + s\begin{pmatrix} 3 \\ 2 \\ 9 \end{pmatrix} + t\begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} = 0\,, \ \left(\begin{array}{ccc|c} 1 & 3 & 4 & 0 \\ 2 & 2 & 2 & 0 \\ 3 & 9 & -1 & 0 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 2 & 2 & 2 & 0 \\ 3 & 9 & -1 & 0 \end{array}\right) \xrightarrow[R_3-3R_1\to R_3]{R_2-2R_1\to R_2} \left(\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 0 & -4 & -8 & 0 \\ 0 & 0 & -16 & 0 \end{array}\right) \xrightarrow[-\frac{1}{16}R_3\to R_3]{-\frac{1}{4}R_2\to R_2} \left(\begin{array}{ccc|c} 1 & 3 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \xrightarrow[R_1-3R_2-5R_3\to R_3]{R_2-2R_3\to R_2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

- This implies that $r = s = t = 0$, and hence the vectors are linearly independent.

Span

Definition

The span of a collection of n-dimensional vectors, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$, is the set of all their linear combinations,

$$\operatorname{span}[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m] = \{\boldsymbol{x} : \boldsymbol{x} = a_1\boldsymbol{x}_1 + \ldots + a_m\boldsymbol{x}_m\}$$

Notes

We need at least n vectors in order to span $\mathbb{R}^n$.

A collection of n-dimensional vectors, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ spans $\mathbb{R}^n$ iff $\operatorname{rank} X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = n$

Intuitively,

- The set spanned by a collection of vectors is all the possible linear combinations of those vectors, that is all the places we could get to using those vectors.

Basis

Definition

A collection of linearly independent vectors $x_1, x_2, \ldots, x_{\mathrm{m}}$ is called basis of $\mathbb{R}^n$ if any other vector $x \in \mathbb{R}^n$ is a linear combination of these vectors $x = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_{\mathrm{m}}$

Notes

Informally we can say,

(1) *A* basis of $\mathbb{R}^n$ is a linearly independent spanning set for $\mathbb{R}^n$

(2) A basis of $\mathbb{R}^n$ is a minimal spanning set for $\mathbb{R}^n$

   - Minimal spanning set = if you remove any element from the set then it no longer spans $\mathbb{R}^n$

(3) A basis of $\mathbb{R}^n$ is a maximal linearly independent subset of $\mathbb{R}^n$

   - Maximal linearly independent subset = add any element of $\mathbb{R}^n$ to this set and it will become linearly dependent

A basis of $\mathbb{R}^n$,

- Must have n vectors,
- The set of vectors must be linearly independent,
- The set of vectors must span $\mathbb{R}^n$.

Example 1: Show that these vectors are a basis for $\mathbb{R}^3$

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ \boldsymbol{v} = \begin{pmatrix} 3 \\ 2 \\ 9 \end{pmatrix}, \ \boldsymbol{w} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

- Check for linear independence: need to show that $r\boldsymbol{u} + s\boldsymbol{v} + t\boldsymbol{w} = 0$ only if $r = s = t = 0$.

$$r\begin{pmatrix}1\\2\\3\end{pmatrix}+s\begin{pmatrix}3\\2\\9\end{pmatrix}+t\begin{pmatrix}5\\2\\-1\end{pmatrix}=0\Rightarrow\left\{\begin{array}{c}r+3s+5t=0\\r+s+t=0\\3r+9s-t=0\end{array}\right.\Rightarrow\begin{array}{c}r+3s+5(3r+9s)=16r+48s=0\Rightarrow 4r+12s=0\\r+s+(3r+9s)=4r+10s=0\end{array}$$

$$\left\{\begin{array}{c}4r+12s=0\\4r+10s=0\end{array}\right.\Rightarrow 2s=0\Rightarrow s=0\Rightarrow r=0\Rightarrow t=0$$

- So $s=r=t=0$ and hence linearly independent.
- Any linearly independent set of n vectors in $\mathbb{R}^n$ spans $\mathbb{R}^n$
- Hence we have a basis for $\mathbb{R}^3$

Example 2: Find the basis for the space spanned by these vectors

$$\boldsymbol{w}_1=\begin{pmatrix}1\\1\\0\end{pmatrix},\ \boldsymbol{w}_2=\begin{pmatrix}0\\1\\1\end{pmatrix},\ \boldsymbol{w}_3=\begin{pmatrix}2\\3\\1\end{pmatrix},\ \boldsymbol{w}_4=\begin{pmatrix}1\\1\\1\end{pmatrix}$$

- Check for linear independence: $r_1\boldsymbol{w}_1+r_2\boldsymbol{w}_2+r_3\boldsymbol{w}_3+r_4\boldsymbol{w}_4=0$ only if $r_1=r_2=r_3=r_4=0$.

  $$\left(\begin{array}{cccc|c}1&0&2&1&0\\1&1&3&1&0\\0&1&1&1&0\end{array}\right)\to\left(\begin{array}{cccc|c}1&0&2&1&0\\0&1&1&0&0\\0&1&1&1&0\end{array}\right)\to\left(\begin{array}{cccc|c}1&0&2&1&0\\0&1&1&0&0\\0&0&0&1&0\end{array}\right)\to\left(\begin{array}{cccc|c}1&0&2&0&0\\0&1&1&0&0\\0&0&0&1&0\end{array}\right)$$

  $$\left\{\begin{array}{c}r_1+2r_3=0\\r_2+r_3=0\\r_4=0\end{array}\right.\Leftrightarrow\left\{\begin{array}{c}r_1=-2r_3\\r_2=-r_3\\r_4=0\end{array}\right.$$

  General solution: $(r_1,r_2,r_3,r_4)=(-2t,-t,t,0)\ \ t\in\mathbb{R}$

  Particular solution: $(r_1,r_2,r_3,r_4)=(2,1,-1,0)$

- This implies that $2w_1+w_2-w_3=0$ hence linearly dependent, therefore drop any of them.
- Try $\boldsymbol{w}_1,\boldsymbol{w}_2,\boldsymbol{w}_4$, check for independence: - Could have checked determinant was not zero instead.

  $$\left(\begin{array}{ccc|c}1&0&1&0\\1&1&1&0\\0&1&1&0\end{array}\right)\xrightarrow{R_2-R_1\to R_2}\left(\begin{array}{ccc|c}1&0&1&0\\0&1&0&0\\0&1&1&0\end{array}\right)\xrightarrow[R_1-R_3+R_2\to R_1]{R_3-R_2\to R_3}\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}$$

  - This implies linear independence, since $r_1=r_2=r_4=0$ is the only solution.
- Conclusion is that w_1,w_2,w_4 is a basis for V, and V $=\mathbb{R}^3$

Standard Basis

$$\text{Standard basis of }\mathbb{R}^2\text{ is }\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right\}$$

$$\text{Standard basis of }\mathbb{R}^3\text{ is }\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$$

$$\text{Standard basis of }\mathbb{R}^n\text{ is }\left\{\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix},\dots,\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}\right\}$$

Inverse

Definition

Let A be an $K \times K$ square matrix (inversion is only for square matrices), the inverse of matrix A is matrix A^{-1} such that $A^{-1}A = AA^{-1} = I_K$.

Notes

Does not necessarily always exist, but when it does it is unique,

Non-singular iff invertible.

Determinant

Definition

$$|A| = \sum_{k=1}{}^{k} a_{ik}(-1)^{i+k} |A_{ik}|$$

Where $|A_{ik}|$ is the minor of A.

- i.e. the determinant of the matrix you are left with when you delete row i and column k

Notes

Again this is just for square matrices.

Also note that the determinant is the same no matter which row i you do it for.

Non-singular ($A\boldsymbol{x} = \boldsymbol{b}$ has exactly one solution for every $\boldsymbol{b}$) iff determinant is non-zero.

Invertible iff determinant is not zero.

Example 1: Diagonal matrix

- Determinant is just the product of the diagonal entries

Example 2: 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},\ |A| = ad - bc$$

Eigenvalues

Definition

An eigenvalue of a n-square matrix A is a scalar λ such that $A \cdot \boldsymbol{x} = \lambda \boldsymbol{x}$.

Notes

Eigenvalues are roots of the characteristic polynomial,

$$|A - \lambda I| = 0$$

There are n roots and some may be complex or repeated.

Diagonal entries of a diagonal matrix D are the eigenvalues of D.

A square matrix A is singular iff 0 is an eigenvalue of A.

Example: 2×2 square matrix

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}$$

$$\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = (3-\lambda)(2-\lambda) - (0)(1)$$

$$(3-\lambda)(2-\lambda) = 0$$

$$\lambda = 3, \lambda = 2$$

Determinant & trace

Let A be a $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then,

$$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$$

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n$$

For 2x2 matrix

$\det A < 0$ eigenvalues have opposite sign,

$\det A > 0$ eigenvalues have the same sign,

- $\mathrm{tr}(A) > 0$ they are all positive,
- $\mathrm{tr}(A) < 0$ they are all negative.

Eigenvectors

Definition

An eigenvector of an n-square matrix A is an n-dimensional vector $\boldsymbol{x}$ such that $A \cdot \boldsymbol{x} = \lambda \boldsymbol{x}$.

Notes

We care only about the non-zero eigenvectors since the zero ones are uninteresting.

Or such that $(A - \lambda I)x = 0$.

Example: 2×2 square matrix

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}$$

$$\lambda = 3, \lambda = 2$$

$$\begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} 3x_1 + x_2 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 3x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} 3x_1 + x_2 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Diagonalizability

Definition

If a matrix has all the eigenvalues real and different, the eigenvectors are independent and span $\mathbb{R}^n$, then by forming the eigenvectors in columns as V, and denoting D as the diagonalised matrix, we have,

$$
\begin{aligned}
&V^{-1}AV^{-1} = D \\
&A = VDV^{-1} \\
&A^n = VD^nV^{-1}
\end{aligned}
$$

Quadratic Forms

Two variable case

$$
\begin{aligned}
Q\left(x_1, x_2\right) &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 \\
&= a_{11}x_1^2 + \left(a_{12} + a_{21}\right)x_1x_2 + a_{22}x_2^2 \\
&\text{let } a_{12} \text{ and } a_{21} \text{ be } 1/2\left(a_{12} + a_{21}\right) \\
&\text{then the new } a_{12} \text{ and } a_{21} \text{ are equal without changing } Q\left(x_1, x_2\right) \\
&= a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \\
Q\left(x_1, x_2\right) &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}
\end{aligned}
$$

Given the replacement of the a's in 12 and 21 then the matrix A is symmetric.

$$
Q\left(x_1, x_2\right) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x'Ax
$$

$Q\left(x_1, x_2\right)$ and A are:

- PD if $Q\left(x_1, x_2\right) > 0$ PSD if $Q\left(x_1, x_2\right) \geq 0$
- ND if $Q\left(x_1, x_2\right) < 0$, NSD if $Q\left(x_1, x_2\right) \leq 0$.

For all $\left(x_1, x_2\right) \neq 0$,

- ID if there exists $\left(x_1^*, x_2^*\right), \left(y_1^*, y_2^*\right)$ such that $Q\left(x_1^*, x_2^*\right) < 0$ and $Q\left(y_1^*, y_2^*\right) > 0$.

General case

$$
Q\left(x_1, x_2\right) = \boldsymbol{x}'A\boldsymbol{x}
$$

$Q\left(x_1, x_2\right)$ and A (symmetric) are:

- PD if $Q(x) > 0$
- PSD if $Q(x) \geq 0$
- ND if $Q(x) < 0$
- NSD if $Q(x) \leq 0$

For all $x \neq 0$,

- ID if there exists x^* and y^* such that $Q\left(x^*\right) < 0$ and $Q\left(y^*\right) > 0$

Definiteness

Eigenvalue Test

- A is PD iff all its eigenvalues are strictly positive.
- A is PSD iff all its eigenvalues are weakly positive.
- A is ND iff all its eigenvalues are strictly negative.
- A is NSD iff all its eigenvalues are weakly negative.
- A is indefinite otherwise.

Determinant Test

- A is positive definite iff $|A_k| > 0$ for all $k = 1, \ldots, n$
- A is positive semi-definite iff $|A_k| \geq 0$ for all $k = 1, \ldots, n$ and all row+col permutations.
- A is negative definite iff $(-1)^k |A_k| > 0$ for all $k = 1, \ldots, n$
- A is negative semi-definite iff $(-1)^k |A_k| \geq 0$ for all $k = 1, \ldots, n$ and all row +col | permutations.
- A is indefinite otherwise.

Where A_k is a square submatrix of A retaining only the first k rows and columns.

$$A_1 = (a_{11}) \ , \ A_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \ ,$$
$$\ldots \ , \quad A_k = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}$$

Note in the ND and NSD cases the determinants of the submatrices need to oscillate.

Multivariate Calculus

Continuity

Definition 1

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\boldsymbol{x}$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that, if $\|\boldsymbol{y} - \boldsymbol{x}\| < \delta$, then $\|f(\boldsymbol{y}) - f(\boldsymbol{x})\| < \varepsilon$.

Definition 2

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\boldsymbol{x}$ if for any sequence $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ converging to $\boldsymbol{x}$, the sequence $\{f(\boldsymbol{x}_k)\}_{k=1}^{\infty}$ converges to $f(\boldsymbol{x})$.

Note the claim is ***any*** converging sequence, hence one counterexample proves discontinuity.

Example 1: Epsilon Delta proof

Show that $f(x, y) = xy$ is continuous at $(0, 0)$

- For any $\varepsilon > 0$ we need to find a $\delta > 0$ such that if $\|(x, y) - (0, 0)\| = \|x, y\| = \sqrt{x^2 + y^2} < \delta$ then $\|f(x, y) - f(0, 0)\| = \|xy\| = |xy| = |x||y| < \varepsilon$
- We know that,

$$|x| = \sqrt{x^2} \leq \sqrt{x^2 + y^2} < \delta$$
$$|y| = \sqrt{y^2} \leq \sqrt{x^2 + y^2} < \delta$$

- Hence $|x||y| < \delta^2$, let $\delta = \sqrt{\varepsilon}$

$$|x||y| < (\sqrt{\varepsilon})^2 = \varepsilon$$

Example 2: Epsilon Delta proof

Show that $f(x) = \sqrt{x}$ is continuous at 0

- For any $\varepsilon > 0$ we need to find a $\delta > 0$ such that if $\|(x) - (x_0)\| < \delta$ then $\|f(x) - f(x_0)\| = \left\|\sqrt{x} - \sqrt{x_0}\right\| < \varepsilon$
- We know that we have an $\varepsilon > 0$ and we need a delta
- We know that $|x - 0| < \delta$ and we want it to imply that $\|f(x) - f(0)\| = |\sqrt{x} - \sqrt{0}| = \sqrt{x} < \varepsilon$ or $x < \varepsilon^2$
- So if we choose $\delta = \varepsilon^2$ we should be able to prove it.
- Let $\varepsilon > 0$ and $\delta = \varepsilon^2$. When $|x - 0| = |x| < \delta$ then we have $|\sqrt{x} - \sqrt{0}| = \sqrt{x} < \sqrt{\delta} = \sqrt{\varepsilon^2} = \varepsilon$

Example 3: Proving sum, subtraction, and product are continuous

- Let f and g be functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ that are continuous at $\boldsymbol{x}$. Then $f + g$, $f - g$, $f \cdot g$ (inner product) are all continuous at $\boldsymbol{x}$.
- Let $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ be a sequence converging to $\boldsymbol{x}$. By continuity $f(\boldsymbol{x}_n)$ converges to $f(\boldsymbol{x})$, and $g(\boldsymbol{x}_n)$ converges to $g(\boldsymbol{x})$. $f(\boldsymbol{x}_n) + g(\boldsymbol{x}_n) = (f + g)(\boldsymbol{x}_n)$ converges to $f(\boldsymbol{x}) + g(\boldsymbol{x}) = (f + g)(\boldsymbol{x})$ therefore $f + g$ is continuous at $\boldsymbol{x}$

Differentiation

Partial Derivative

Consider a real valued function $f : \mathbb{R}^n \to \mathbb{R}$.

The kth partial derivative of f at $\boldsymbol{x} \in \mathbb{R}^n$ is

$$\frac{\partial f}{\partial x_k}(\boldsymbol{x}) = \underset{h\to 0}{lim} \frac{f(x_1, ..., x_k + h, ..., x_n) - f(x_1, ..., x_n)}{h}$$

Other notion:

$$\frac{\partial f}{\partial x_k}(\boldsymbol{x}) = f_k(\boldsymbol{x}) = f_{x_k} = \partial_{x_k} f = \partial_k f = D_k f$$

Young's Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$. If $\partial_i \partial_j f$ and $\partial_j \partial_i f$ exist and are continuous then

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \text{ for } i, j = 1, ..., n$$

Differentiability

The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\boldsymbol{x}$ if there exists an $m \times n$ matrix A such that

$$\underset{||h||\to 0}{lim} \frac{||f(\boldsymbol{x} - \boldsymbol{h}) - f(\boldsymbol{x}) - A\boldsymbol{h}||}{||\boldsymbol{h}||} = 0$$

This might be less confusing in the univariate case: $f : \mathbb{R} \to \mathbb{R}$. If $f'(x)$ is to exist then it must be the case that

$$f'(x) = \underset{h\to 0}{lim} \frac{f(x-h) - f(x)}{h}$$
$$0 = \underset{h\to 0}{lim} \frac{f(x-h) - \left[f(x) + hf'(x)\right]}{h}$$

Sufficient Conditions

A function f is differentiable at $\boldsymbol{x}$ if all the partial derivatives of f exist in th neighbourhood of $\boldsymbol{x}$ and are continuous at $\boldsymbol{x}$.

A function f is differentiable at $\boldsymbol{x}$ if f is C^1: all its partial derivatives exist and are continuous at all points of its domain.

Jacobian

If $f : \mathbb{R}^n \to \mathbb{R}^m$ where $f = (f_1, ..., f_m)$, we define the Jacobian matrix as the $m \times n$ matrix $Df(x)$, where

$$Df(x) = \begin{pmatrix} \partial_1 f_1 & ... & \partial_n f_1 \\ \vdots & \ddots & \vdots \\ \partial_1 f_m & ... & \partial_n f_m \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & ... & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & ... & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

In the case when $f : \mathbb{R}^n \to \mathbb{R}$ then the Jacobian matrix is the $1 \times n$ matrix

$$Df(x) = \begin{pmatrix} \partial_1 f & ... & \partial_n f \end{pmatrix}$$

We call the transpose of this matrix in the real-valued function case the gradient matrix

$$\nabla f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_n f \end{pmatrix}$$

Example:

$$f(x, y) = (xy\ ,\ x^2 + y^2)$$
$$Df(x, y) = \begin{pmatrix} y & x \\ 2x & 2y \end{pmatrix}$$

Example:

$$f(x, y) = (x^2 y + y^2 x)$$
$$Df(x, y) = \begin{pmatrix} 2xy + y^2 & x^2 + 2yx \end{pmatrix}$$

Example:

$$f(x, y) = (xy^2)$$
$$\nabla f = \begin{pmatrix} y^2 \\ 2xy \end{pmatrix}$$

Hessian

The second-order partial derivatives can be written as a matrix,

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Providing that the function f is C^2 (all partial derivatives exist and are continuous at all points in the domain) the matrix is symmetric.

Example:

$$f(x, y) = \left(x^2 y + y^2 x\right)$$
$$Df(x, y) = \left(2xy + y^2 x^2 + 2yx\right)$$

$$H = \begin{pmatrix} 2y & 2x + 2y \\ 2x + 2y & 2x \end{pmatrix}$$

Chain Rule

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $x : \mathbb{R}^m \to \mathbb{R}^n$ be C^1 functions. Define $z : \mathbb{R}^m \to \mathbb{R}$ such that $z(\boldsymbol{t}) = f(x(\boldsymbol{t}))$, then z is a C^1 function, and,

$$\frac{\partial z}{\partial t_k} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_k} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_k} + \ldots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t_k} \text{ for } 1 \leq k \leq m$$

Example:

$$\begin{aligned}
f(x,y) &= x^2 + y, x(s,t) = s+t, y(s+t) = s^2 + t^2 \\
\frac{\partial z}{\partial s} &= 2x \cdot 1 + 1 \cdot 2s = 2x + 2s = 2(s+t) + 2s \\
&= 4s + 2t \\
\frac{\partial z}{\partial t} &= 2x \cdot 1 + 1 \cdot 2t = 2x + 2t = 2(s+t) + 2s \\
&= 4t + 2s
\end{aligned}$$

Total Differential

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function. If there is an infinitesimal change $d\boldsymbol{x}$ in $\boldsymbol{x}$, the corresponding change in f is given by the total differential,

$$\mathrm{d}f = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \ldots + \frac{\partial f}{\partial x_n}dx_n \text{ or } \mathrm{d}f = Df \cdot d\boldsymbol{x}$$

Example:

$$\begin{aligned}
f(x,y) &= x^2 y + y^2 x, \\
\mathrm{d}f &= \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y \\
\mathrm{d}f &= \left(2xy + y^2\right)\mathrm{d}x + \left(x^2 + 2yx\right)\mathrm{d}y
\end{aligned}$$

Taylor Approximations

Taylor's Theorem

First Order

Suppose $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$, is a C^1 function. Then for any $\boldsymbol{a}, \boldsymbol{x} \in U$

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + Df(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a}) + R_1(\boldsymbol{x}, \boldsymbol{a})$$
$$\text{where } \lim_{\boldsymbol{x}\to\boldsymbol{a}} \frac{R_1(\boldsymbol{x}, \boldsymbol{a})}{\|\boldsymbol{x} - \boldsymbol{a}\|} \to 0$$

Second Order:

Suppose $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$, is a C^2 function. Then for any $\boldsymbol{a}, \boldsymbol{x} \in U$

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + Df(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{a})^T D^2 f(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a}) + R_2(\boldsymbol{x}, \boldsymbol{a})$$
$$\text{where } \lim_{\boldsymbol{x}\to\boldsymbol{a}} \frac{R_2(\boldsymbol{x}, \boldsymbol{a})}{\|\boldsymbol{x} - \boldsymbol{a}\|^2} \to 0$$

Example:

Find a Taylor approximation of $z = f(x, y)$ at (a, b)

(1) First order:

$$f(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) + R_1$$

(2) Second order:

$$\begin{aligned} f(x, y) =& f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) \\ &+ \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(a, b)(x - a)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a, b)(x - a)(y - b) + \frac{\partial^2 f}{\partial y^2}(a, b)(y - b)^2\right] + R_2 \end{aligned}$$

Example:

Find a Taylor approximation of $z = x^2y + y^2x$ at $(1, 2)$ where

$$Dz = \begin{pmatrix} 2xy + y^2 & x^2 + 2yx \end{pmatrix} \quad H = \begin{pmatrix} 2y & 2x + 2y \\ 2x + 2y & 2x \end{pmatrix}$$

First order:

$$\begin{aligned} f(x, y) &= f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) + R_1 \\ z &= \left\{1^2 2 + 2^2 1\right\} + \left\{2(1)(2) + 2^2\right\}(x - 1) + \left\{1^2 + 2(2)(1)\right\}(y - 2) + R_1 \\ z &= 6 + 8(x - 1) + 4(y - 2) + R_1 \end{aligned}$$

Second order:

$$\begin{aligned} f(x, y) &= f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) \\ &+ \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(a, b)(x - a)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a, b)(x - a)(y - b) + \frac{\partial^2 f}{\partial y^2}(a, b)(y - b)^2\right] + R_2 \\ z &= \left\{1^2 2 + 2^2 1\right\} + \left\{2(1)(2) + 2^2\right\}(x - 1) + \left\{1^2 + 2(2)(1)\right\}(y - 2) \\ &+ \frac{1}{2}\left[2(2)(x - 1)^2 + 2(2(1) + 2(2))(x - 1)(y - 2) + 2(1)(y - 2)^2\right] + R_2 \\ z &= 6 + 8(x - 1) + 4(y - 2) + \frac{1}{2}\left[4(x - 1)^2 + 12(x - 1)(y - 2) + 2(y - 2)^2\right] + R_2 \end{aligned}$$

Implicit Function Theorem

If it is the case that the equation $F(x, y) = 0$ has a solution y^* ***for a particular*** x^*, and we want to know how y^* changes with x^*, then rather than finding each x^* and it's corresponding y^*, we can think of y **implicitly** as a function of x.

Univariate Case

Definition

Suppose $x \in \mathbb{R}^m$, and $(\boldsymbol{x}^*, y^*)$ is a solution to $F(\boldsymbol{x}, y) = 0$. Suppose that F is C^1 in an open ball around $(\boldsymbol{x}^*, y^*)$, with $\frac{\partial F}{\partial y}(\boldsymbol{x}^*, y^*) \neq 0$. Then there is a C^1 function $y = y(\boldsymbol{x})$ defined in an open ball around $\boldsymbol{x}^*$ such that

$$\begin{aligned}
&y(\boldsymbol{x}^*) = y^* \\
&F(\boldsymbol{x}, y(\boldsymbol{x})) = 0 \\
&\frac{\partial y}{\partial x_i}(\boldsymbol{x}, y) = -\frac{\frac{\partial F}{\partial x_i}(\boldsymbol{x}, y)}{\frac{\partial F}{\partial y}(\boldsymbol{x}, y)}
\end{aligned}$$

Notes

This essentially says if $F(\boldsymbol{x}, y) = 0$ has a solution y^* for a particular $\boldsymbol{x}^*$, then it is legitimate to think of y as a function of x around that point, and therefore you can determine how y changes from y^* when x changes a little from x^* by implicitly differentiation.

Example:

Consider the equation $x^2 + y^2 = 10$ with a solution at $x = 3$ and $y = -1$. Find $\frac{\mathrm{d}y}{\mathrm{d}x}$ at this point.

- Define $F(x, y) = x^2 + y^2 - 10$, hence $(3, -1)$ is a solution to $F(x, y) = 0$

(1) Check $F(x, y) = x^2 + y^2 - 10$ is C^1 :
  - $\frac{\partial F}{\partial x}(x, y) = 2x$, which is continuous,
  - $\frac{\partial x}{\partial y}(x, y) = 2y$, which is continuous,
  - Hence F is C^1.

(2) Check that $\frac{\partial F}{\partial y}(x^*, y*) \neq 0$
  - $\frac{\partial F}{\partial y}(3, -1) = 2(-1) = -2 \neq 0$

- Conclusion: there exists a $y = y(x)$ such that

(1) $y(3) = -1$

(2) $F(x, y(x)) = 0$

(3) $\frac{\partial y}{\partial x}(x, y) = -\frac{\frac{\partial F}{\partial x}(x,y)}{\frac{\partial F}{\partial y}(x,y)}$

- Giving $\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{2x}{2y} = -\frac{x}{y} = 3$ (at the solution).

Note the IFT does not say that $y(x)$ is the only solution to the equation for a given value of x, but rather that it is the only solution near $(3, -1)$

Multivariate Case

Definition

Consider $\boldsymbol{F} : \mathbb{R}^{m+n} \to \mathbb{R}^n$. Suppose that $\boldsymbol{F}(\boldsymbol{x}^*, \boldsymbol{y}^*) = 0$ and that each F_i is C^1 around $(\boldsymbol{x}^*, \boldsymbol{y}^*)$, and $D_y\boldsymbol{F}$ is invertible at this point $(\boldsymbol{x}^*, \boldsymbol{y}^*)$. Then, there exists a collection of C^1 functions $\boldsymbol{y}(\boldsymbol{x})$ $[y_1(x_1, ..., x_n), ..., y_m(x_1, ..., x_n)]$ such that,

$$\begin{aligned}
&\boldsymbol{y}(\boldsymbol{x}^*) = \boldsymbol{y}^* \\
&\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x})) = \mathbf{0} \\
&D_x\boldsymbol{y}(\boldsymbol{x}) = -\left[D_y\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x}))\right]^{-1} D_x\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x}))
\end{aligned}$$

Find partials by

$$\begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} & \cdots & \frac{\partial F_1}{\partial y_n} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} & \cdots & \frac{\partial F_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial y_1} & \frac{\partial F_n}{\partial y_2} & \cdots & \frac{\partial F_n}{\partial y_n} \end{pmatrix} \begin{pmatrix} \frac{\partial y_1}{\partial x_j} \\ \frac{\partial y_2}{\partial x_j} \\ \vdots \\ \frac{\partial y_n}{\partial x_j} \end{pmatrix} = - \begin{pmatrix} \frac{\partial F_1}{\partial x_j} \\ \frac{\partial F_2}{\partial x_j} \\ \vdots \\ \frac{\partial F_n}{\partial x_j} \end{pmatrix}$$

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_j} \\ \vdots \\ \frac{\partial y_n}{\partial x_j} \end{pmatrix} = - \left[\begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial y_1} & \cdots & \frac{\partial F_n}{\partial y_n} \end{pmatrix} \right]^{-1} \begin{pmatrix} \frac{\partial F_1}{\partial x_j} \\ \vdots \\ \frac{\partial F_n}{\partial x_j} \end{pmatrix}$$

Notes

Consider a system of n equations with $n + m$ variables

$$\begin{aligned}
&F_1(x_1, x_2, \ldots, x_m; y_1, y_2, \ldots, y_n) = 0 \\
&F_2(x_1, x_2, \ldots, x_m; y_1, y_2, \ldots, y_n) = 0 \\
&\cdots \\
&F_n(x_1, x_2, \ldots, x_m; y_1, y_2, \ldots, y_n) = 0
\end{aligned}$$

These equations implicitly define y_i as a function of $(x_1, \ldots, x_m)$. We can hence generalise the Implicit function theorem to see how y_k changes with x_j

Implicitly differentiating the ith equation gives:

$$\frac{\partial F_i}{\partial x_j} + \frac{\partial F_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial F_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \ldots + \frac{\partial F_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} = 0$$

And

$$\begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} & \cdots & \frac{\partial F_1}{\partial y_n} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} & \cdots & \frac{\partial F_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial y_1} & \frac{\partial F_n}{\partial y_2} & \cdots & \frac{\partial F_n}{\partial y_n} \end{pmatrix} \begin{pmatrix} \frac{\partial y_1}{\partial x_j} \\ \frac{\partial y_j}{\partial x_j} \end{pmatrix} = \begin{pmatrix} \frac{\partial F_1}{\partial x_j} \\ \vdots \frac{\partial F_2}{\partial x_j} \\ \vdots \\ \frac{\partial F_n}{\partial x_j} \end{pmatrix}$$

Which has a solution if the Jacobian is non-singular.

Unconstrained Optimisation

Throughout this section we are considering $f(\boldsymbol{x})$: a real valued function (n arguments but 1 value outputted) which is C^2, hence we have

$$Df = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{pmatrix}, \; D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Weierstrass Theorem

A continuous real valued function $f : S \to \mathbb{R}$, where is S is a compact subset of $\mathbb{R}^n$ attains a maximum and a minimum value.

First Order Conditions (FOCs)

If f has a local maximum or minimum at an interior point $\boldsymbol{x}^*$ of S, then,

$$Df\left(\boldsymbol{x}^*\right) = 0$$

Second Order Conditions (SOCs)

Suppose that $\boldsymbol{x}^*$ is an interior point of S and that $Df\left(\boldsymbol{x}^*\right) = 0$.

Second Order Necessary Conditions

- If f has a local maximum, at an interior point $\boldsymbol{x}^*$ of S, then $D^2 f\left(\boldsymbol{x}^*\right)$ is negative semidefinite.
- If f has a local minimum, at an interior point $\boldsymbol{x}^*$ of S, then $D^2 f\left(\boldsymbol{x}^*\right)$ is positive semidefinite.

Second Order Sufficient Conditions

- If $D^2 f\left(\boldsymbol{x}^*\right)$ is negative definite, then $\boldsymbol{x}^*$ is a strict local maximum.
- If $D^2 f\left(\boldsymbol{x}^*\right)$ is positive definite, then $\boldsymbol{x}^*$ is a strict local minimum.
- If $D^2 f\left(x^*\right)$ is indefinite, then $\boldsymbol{x}^*$ is a saddle point.

Checking the definiteness of a symmetric matrix A

(A) Eigenvalue test
- A is positive definite iff all its eigenvalues are strictly positive.
- A is positive semi-definite iff all its eigenvalues are weakly positive.
- A is negative definite iff all its eigenvalues are strictly negative.
- A is negative semi-definite iff all its eigenvalues are weakly negative.
- A is indefinite otherwise.
- https://www.symbolab.com/solver/matrix-eigenvalues-calculator

(B) Determinant test

- A is positive definite iff $|A_k| > 0$ for all $k = 1, \ldots, n$.
- A is positive semi-definite iff $|A_k| \geq 0$ for all $k = 1, \ldots, n$ and all row+col permutations.
- A is negative definite iff $(-1)^k\,|A_k| > 0$ for all $k = 1, \ldots, n$.
- A is negative semi-definite iff $(-1)^k\,|A_k| \geq 0$ for all $k = 1, \ldots, n$ and all row+col permutations.
- A is indefinite otherwise.
- https://www.symbolab.com/solver/matrix-determinant-calculator

Examples

- The leading minors of $A = \begin{pmatrix} 5 & 4 \\ 4 & 11 \end{pmatrix}$ are 5 and $(55 - 16) = 39$, so A is positive definite.
- The leading minors of $A = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}$ are -1 and 2, so A is negative definite.
- The leading minors of $A = \begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix}$ are -1 and -4, so A is indefinite.

Global Optima for concave functions

In the case of concave functions, stationary points are automatically global maxima.

Let f be a C^2 function defined on a convex set S : - $D^2 f$ is positive definite for all x in S $\Rightarrow$ f is strictly convex on S.

- $D^2 f$ is positive semi-definite for all $\mathbf{x}$ in S $\Leftrightarrow$ f is convex on S.
- $D^2 f$ is negative definite for all $\mathbf{x}$ in S $\Rightarrow$ f is strictly concave on S.
- $D^2 f$ is negative semi-definite for all $\mathbf{x}$ in S $\Leftrightarrow$ f is concave on S.

Unconstrained Optimisation: Example

Find all the local maxima and minima of $f(x,y) = x^3 + y^3 - 3x - 3y + 3xy$

Does the global maximum or minimum exist?

FOCs

For a real valued function: $Df(x) = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right)$

$$Df(x,y) = \left(\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}\right) = \left(3x^2 - 3 + 3y \quad 3y^2 - 3 + 3x\right)$$

Solution is at $Df(x^*, y^*) = 0$

$$\begin{aligned}
&3x^2 - 3 + 3y = 0 \Rightarrow x^2 - 1 + y = 0 \Rightarrow y = 1 - x^2 \\
&3y^2 - 3 + 3x = 0 \Rightarrow y^2 - 1 + x = 0 \\
&\left(1 - x^2\right)^2 - 1 + x = 0 \Rightarrow (1-x)(1+x)(1-x)(1+x) - (1-x) \\
&\Rightarrow (1-x)\left\{(1+x)^2(1-x) - 1\right\} \\
&\Rightarrow (1-x)\left\{\left(x^2 + 2x + 1\right)(1-x) - 1\right\} \\
&\Rightarrow (1-x)\left\{\left(\left(x^2 + 2x + 1\right) - x\left(x^2 + 2x + 1\right)\right) - 1\right\} \\
&\Rightarrow (1-x)\left\{x^2 + 2x + 1 - x^3 - 2x^2 - x - 1\right\} \\
&\Rightarrow (1-x)x\left\{x + 2 - x^2 - 2x - 1\right\} \\
&\Rightarrow (1-x)x\left\{1 - x - x^2\right\}
\end{aligned}$$

Hence we have four solutions that meet the FOCs

$$\begin{array}{ll}
\text{(A) } x = 1, y = 0 & \text{(C) } x = y = \frac{-1+\sqrt{5}}{2} \\
\text{(B) } x = 0, y = 1 & \text{(D) } x = y = \frac{-1-\sqrt{5}}{2}
\end{array}$$

SOCs

$$D^2 f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 6x & 3 \\ 3 & 6y \end{pmatrix}$$

For each solution:

(A) $(1,0) \Rightarrow H = \begin{pmatrix} 6 & 3 \\ 3 & 0 \end{pmatrix} \Rightarrow |H_1| = 6 > 0$ and $|H| = (6 \cdot 0 - 3 \cdot 3) = -9 < 0 \Rightarrow$ indefinite $\Rightarrow$ saddle point.

(B) $(0,1) \Rightarrow H = \begin{pmatrix} 0 & 3 \\ 3 & 6 \end{pmatrix} \Rightarrow |H_1| = 0$ and $|H| = (0 \cdot 6 - 3 \cdot 3) = -9 < 0 \Rightarrow$ indefinite $\Rightarrow$ saddle point.

(C) $\left(\frac{-1+\sqrt{5}}{2}, \frac{-1+\sqrt{5}}{2}\right) \Rightarrow H = \begin{pmatrix} 3(-1+\sqrt{5}) & 3 \\ 3 & 3(-1+\sqrt{5}) \end{pmatrix} \Rightarrow |H_1| = 3(-1+\sqrt{5}) > 0$ and $|H| = \left(9(-1+\sqrt{5})^2 - 3 \cdot 3\right) = 9(6 - 2\sqrt{5} - 1) = 9(5 - 2\sqrt{5}) > 0 \Rightarrow$ positive definite $\Rightarrow$ strict local minimum.

(D) $\left(\frac{-1-\sqrt{5}}{2}, \frac{-1-\sqrt{5}}{2}\right) \Rightarrow H = \begin{pmatrix} 3(-1-\sqrt{5}) & \\ 3 & 3(-1-\sqrt{5}) \end{pmatrix} \Rightarrow |H_1| = -3(1+\sqrt{5}) < 0$ and $|H| = \left(9(1+\sqrt{5})^2 - 3 \cdot 3\right) = 9(6 + 2\sqrt{5} - 1) = 9(5 + 2\sqrt{5}) > 0 \Rightarrow$ negative definite $\Rightarrow$ strict local maximum.

Constrained Optimisation: Equality Constraints

Maximise $f(\boldsymbol{x}) = f(x_1, \ldots, x_n)$ subject to $h(\boldsymbol{x}) = \boldsymbol{a}$, $h_1(x_1, \ldots, x_n) = a_1, \ldots, h_m(x_1, \ldots, x_n) = a_m$ $[\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m]$

Lagrange's Theorem FOCs

Let,

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{j=1}^{m} \lambda_j \left(h_j(\boldsymbol{x}) - a_j\right)$$

$$D_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = Df(\boldsymbol{x}) - \sum_{j=1}^{m} \lambda_j Dh_j(\boldsymbol{x})$$

If $\boldsymbol{x}^*$ is a local maximum of f subject to the constraints $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{a}$ and the matrix $D\boldsymbol{h}(\boldsymbol{x}^*)$ is of full rank m, then there exists a $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)$ such that,

$$D_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$$

Constraint Qualification

"$D\boldsymbol{h}(\boldsymbol{x}^*)$ is of full rank m" is known as the **constraint qualification** (CQ).

Recall that,

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_m(x) \end{pmatrix} \text{ and } D\boldsymbol{h} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix}$$

Lagrange's Theorem SOCs (sufficient conditions)

Suppose that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ satisfies the CQ and FOCs, then:

(a) If $v^T D_x^2 L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) v < 0 \; \forall v \neq 0$ in $\mathbb{R}^n$ for which $D\boldsymbol{h}(\boldsymbol{x}^*) v = \mathbf{0}$ then $\boldsymbol{x}^*$ is a strict local maximum.

(b) If $v^T D_x^2 L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) v > 0 \; \forall v \neq 0$ in $\mathbb{R}^n$ for which $D\boldsymbol{h}(\boldsymbol{x}^*) v = \mathbf{0}$ then $\boldsymbol{x}^*$ is a strict local maximum.

Borderer Hessian Test (Checking for the SOCs)

Construct the bordered Hessian:

$$\begin{pmatrix} \mathbf{0} & D\boldsymbol{h}(\boldsymbol{x}^*) \\ D\boldsymbol{h}(\boldsymbol{x}^*)^T & D_x^2 L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \end{pmatrix}$$

Check the sign of the last $n - m$ leading principal minors:

(a) If they alternate in sign ending with $(-1)^n$ then H is negative definite $\left[v^T A v < 0 \; \forall v \neq 0\right]$ and $\boldsymbol{x}^*$ is a strict local maximum.

(b) If they have the same sign as $(-1)^m$ then H is positive definite $\left[v^T A v > 0 \; \forall v \neq 0\right]$ and $\boldsymbol{x}^*$ is a strict local minimum.

Notes

Where n is the number of arguments (x's and y's), and m is the number of constraints.

If $n - m = 0$ then you have nothing to check!

Example: Bordered Hessian

$$f(x_1, x_2, x_3, x_4) = x_1^2 - x_2^2 + x_3^2 + x_4^2 + 4x_2x_3 - 2x_1x_4) \text{ on } x_2 + x_3 + x_4 = 0 \text{ and } x_1 - 9x_2 + x_4 = 0$$

Given that $n = 4$ and $m = 2$ we need to check the last $n - m = 2$ principals: $H_5,\ H_6$ where,

$$H_6 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & -9 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 1 & -9 & 0 & -1 & 2 & 0 \\ 1 & 0 & 0 & 2 & 1 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 \end{pmatrix}$$

$$H_5 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -9 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & -9 & 0 & -1 & 2 \\ 1 & 0 & 0 & 2 & 1 \end{pmatrix}$$

Given that $n = 4$ and that $(-1)^4 > 0$ we would need $|H_6| > 0$ and $|H_5| < 0$ to verify negative definiteness.

Given that $m = 2$ and that $(-1)^2 > 0$ we would need $|H_6| > 0$ and $|H_5| > 0$ to verify positive definiteness.

In fact the determinant of $H_6 = 24$ and $H_5 = 77$ hence f is positive definite and $\boldsymbol{x}^*$ is a local minimum.

Equality Constraint: Example

$$\min rK + wL \ (r > 0, w > 0) \text{ s.t. } K^{1/2}L^{1/2} = y > 0 \Rightarrow KL = y^2 \text{ since } K > 0\ ,\ L > 0$$

Weirestrass Theorem

The constraint set is not compact hence we cannot call on the Weirestrass theorem.

Constraint Qualification

$Dh = (LK) \neq 0$ because $K > 0$ and $L > 0$.

This implies full rank because as long as this matrix is nonzero it has full rank.

Lagrangian & FOCs

$$\mathbf{L}(K, L, \lambda) = rK + wL - \lambda\left(KL - y^2\right)$$

FOCs

$$\left.\begin{array}{c} r - \lambda L = 0 \\ w - \lambda K = 0 \\ KL = y^2 \end{array}\right\} \lambda = \frac{r}{L} = \frac{w}{K} > 0, \text{ so } y^2 = \frac{K^2 r}{w} = \frac{L^2 w}{r}$$

Hence,

$$K^* = y(w/r)^{1/2}\ ,\ L^* = y(r/w)^{1/2}$$

Evaulate and Compare

If you know a global maximum exists (by the Weirestrass Theorem), then evaluate f at the points which satisfy the FOCs and at which one f is largest.

We do not know this in our case, hence we need to check the Bordered Hessian instead.

Bordered Hessian

If you do not know a global maximum exists, use Bordered Hessian to check SOCs.

Bordered Hessian:

$$H = \begin{pmatrix} 0 & L & K \\ L & 0 & -\lambda \\ K & -\lambda & 0 \end{pmatrix}$$

This is because $D_{K,L}\mathbf{L} = \begin{pmatrix} r - \lambda L \\ w - \lambda K \end{pmatrix}, D^2_{K,L}\mathbf{L} = \begin{pmatrix} \frac{\partial}{\partial K}(r - \lambda L) & \frac{\partial}{\partial L}(r - \lambda L) \\ \frac{\partial}{\partial K}(w - \lambda K) & \frac{\partial}{\partial L}(w - \lambda K) \end{pmatrix} = \begin{pmatrix} 0 & -\lambda \\ -\lambda & 0 \end{pmatrix}$

How many of the last leading principle minors should we check?

- $n = 2$ (we have two variables), $m = 1$ (one constraint)
- Hence $n - m = 1$,
- So we only need to check the last leading principle minor, in other words the sign of $\det(H)$

$$|H| = 0(0 \cdot 0 - (-\lambda)(-\lambda)) - L(L \cdot 0 - (-\lambda)K) + K(L \cdot (-\lambda) - 0 \cdot K) = -2\lambda KL < 0$$

The sign coincides with the sign $(-1)^m$

Hence positive definite hence local minimum.

Global?

If we let c^* be the cost of using K^* and L^*,

$$c^* = rK^* + wL^* = 2y(rw)^{1/2}$$

And define $\bar{K}, \bar{L}$, by $r\bar{K} = c^*$, $w\bar{L} = c^*$, using $K \geq \bar{K}, L \geq \bar{L}$, then to produce y would cost more than $c^*(y)$ and could not therefore be optimal.

Hence $K \leq \bar{K}, L \leq \bar{L}$, but these constraints do not bind so local minimum is a global minimum.

Constrained Optimisation: Inequality Constraints

$$max\ f(x) = f(x_1, \ldots, x_n) \text{ subject to}$$
$$\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{b}\ :\ g_1(x_1, \ldots, x_n) \leq b_1, \ldots, g_k(x_1, \ldots, x_n) \leq b_k \quad \left[\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^k\right]$$

Maximisation vs Minimisation

For *maximisation* problems it MUST be the case that constraints are written as *less than or equal to zero*, that is $g(x) - b \leq 0$.

For *minimisation* problems it MUST be the case that constraints are written as *greater than or equal to zero* that is $b - g(x) \geq 0$.

Kuhn Tucker's FOCs

Suppose now for one constraint we have the Lagrangian,

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{j=1}^{m} \lambda_j \left(g_i(\boldsymbol{x}) - b_j\right)$$
$$D_x L(\boldsymbol{x}, \boldsymbol{\lambda}) \text{ and } D_x^2 L(\boldsymbol{x}, \boldsymbol{\lambda})$$

Suppose $\boldsymbol{x}^*$ is a local maximum of f subject to the constraint $g(\boldsymbol{x}) \leq b$ and the constraint qualification is satisfied at $\boldsymbol{x}^*$. Then there exists $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)$ such that,

$$D_x L\left(\boldsymbol{x}^*, \boldsymbol{\lambda}\right) = 0$$
$$\lambda_j^* \geq 0 \text{ and } \lambda_j^*\left(g_j\left(\boldsymbol{x}^*\right) - b_j\right) = 0 \text{ for } j = 1, \ldots, m$$

Obviously for inequality constraints you don't have to impose that the constraints bind (which are focs for equality constraint optimisation)

Rather here we have **complementary slackness conditions**.

Complementary Slackness Conditions

$$\lambda_j \geq 0 \ \text{ and } \lambda_j^*\left(g_j(\boldsymbol{x}^*) - b_j\right) = 0 \text{ for } j = 1, ..., m$$

(1) λ_j for all j is non-negative as an agent only needs to be penalised if she wishes the constraint to exceed b (for it to be the case that $g_j(\boldsymbol{x}) > b_j$).

(2) Note that if $g_j(\boldsymbol{x}^*) < b_j$, then, $\lambda_j = 0$.

That is λ_j is only positive if the jth constraint is satisfied with equality.

These conditions allow for,

(1) Interior optima

$$\lambda_j^* = 0\ \forall j \ \text{ so } D_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = Df(\boldsymbol{x}^*) = 0$$

(2) Constrained optima

Constraint Qualification

Let g_E be all the constraints which bind $\boldsymbol{x}^*$, and e be the number of binding constraints. The constraint qualification is satisfied if the matrix $Dg_E\left(\boldsymbol{x}^*\right)$ is of full rank e.

$$Dg_E\left(\boldsymbol{x}^*\right) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}\left(\boldsymbol{x}^*\right) & \dots & \frac{\partial g_1}{\partial xn}\left(\boldsymbol{x}^*\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_e}{\partial x_1}\left(\boldsymbol{x}^*\right) & \dots & \frac{\partial g_e}{\partial x_n}\left(\boldsymbol{x}^*\right) \end{pmatrix}$$

Kuhn Tucker's SOCs

Suppose $\left(\boldsymbol{x}^*, \boldsymbol{\lambda}^*\right)$ satisfy the constraint qualification and the FOCs,

(1) If some constraints bind

- If some of the constraints g_E bind at $\boldsymbol{x}^*$ and $v^T D_x^2 L\left(\boldsymbol{x}^*, \boldsymbol{\lambda}^*\right) v < 0 \forall v \neq 0$ in $\mathbb{R}^n$ **for which** $Dg_E\left(\boldsymbol{x}^*\right) v = 0$ then $\boldsymbol{x}^*$ is a strict local maximum.
- Check via Bordered Hessian
$$\begin{pmatrix} 0 & Dg_E(\boldsymbol{x}^*) \\ Dg_E(\boldsymbol{x}^*)^T & D_x^2 L(\boldsymbol{x}, \boldsymbol{\lambda}^*) \end{pmatrix}$$
- Check the sign of the last $n-e$ leading principal minors:
  (a) If they alternate in sign ending with $(1)^n$ then Q is negative definite $\left[v^T A v < 0 \forall v \neq 0\right]$ and $\boldsymbol{x}^*$ is a local maximum.
  (b) If they have the same sign as $(1)^e$ then Q is positive definite $\left[\ v^T A v > 0 \forall v \neq 0\right]$ and $\boldsymbol{x}^*$ is a local minimum.

(2) If no constraints bind

- If no constraints bind at $\boldsymbol{x}^*$ and $v^T D_x^2 L\left(x^*, \lambda^*\right) v = v^T D_x^2 f\left(x^*\right) v < 0 \forall v \neq 0$ in $\mathbb{R}^n$ (is negative definite) then $\boldsymbol{x}^*$ is a strict local maximum. (CHECK NORMAL HESSIAN IF NONE BIND)

Constrained Optimisation: Equality & Inequality Constraints

$$\begin{aligned}
&max\ f(\boldsymbol{x}) = f\left(x_1, \ldots, x_n\right) \text{ subject to} \\
&\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{a}\ ,\ h_1\left(x_1, \ldots, x_n\right) = a_1, \ldots, h_m\left(x_1, \ldots, x_n\right) = a_m \quad \left[\boldsymbol{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m\right] \\
&\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{b}\ ,\ g_1\left(x_1, \ldots, x_n\right) \leq b_1, \ldots, g_k\left(x_1, \ldots, x_n\right) \leq b_k \quad \left[\boldsymbol{g}: \mathbb{R}^n \rightarrow \mathbb{R}^k\right]
\end{aligned}$$

Kuhn Tucker's FOCs

Our Lagrangian and first and second order conditions are given by,

$$\begin{aligned}
&L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{j=1}^{m} \lambda_j \left(h_j(\boldsymbol{x}) - a_j\right) - \sum_{j=1}^{k} \mu_j \left(g_j(\boldsymbol{x}) - b_j\right) \\
&D_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ and } D_x^2 L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})
\end{aligned}$$

If $\boldsymbol{x}^*$ is a local maximum of f on S, and CQ holds, then $\boldsymbol{x}^*$ is a solution to:

(1) The FOCs,

$$\frac{\partial L}{\partial x}\left(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) = 0 \text{ for } i = 1, \ldots, m$$

(2) The equality constraints,

$$h_j\left(\boldsymbol{x}^*\right) = a_j \text{ for } j = 1, \ldots, m$$

(3) The complementary slackness conditions for the inequality constraints,

$$\mu_j^* \geq 0\ ,\ g_j\left(x^*\right) \leq b_j \text{ and } \mu_j^*\left(g_j\left(x^*\right) - b_j\right) = 0 \text{ for } j = 1, \ldots, k$$

Constraint Qualification

The constraint qualification is satisfied if the matrix of binding constraints is of full rank $m + e$.

$$\begin{pmatrix} D\boldsymbol{h}\left(\boldsymbol{x}^*\right) \\ D\boldsymbol{g}_E\left(\boldsymbol{x}^*\right) \end{pmatrix} \text{ has full rank } m + e, \text{ where,}$$

$$\begin{pmatrix} D\boldsymbol{h}(\boldsymbol{x}^*) \\ D\boldsymbol{g}_E(\boldsymbol{x}^*) \end{pmatrix} = \begin{pmatrix}
\frac{\partial g_1}{\partial x_1}\left(\boldsymbol{x}^*\right) & \cdots & \frac{\partial g_1}{\partial x_n}\left(\boldsymbol{x}^*\right) \\
\vdots & \ddots & \vdots \\
\frac{\partial g_e}{\partial x_1}\left(\boldsymbol{x}^*\right) & \cdots & \frac{\partial g_e}{\partial x_n}\left(\boldsymbol{x}^*\right) \\
\frac{\partial h_1}{\partial x_1}\left(\boldsymbol{x}^*\right) & \cdots & \frac{\partial h_1}{\partial x_n}\left(\boldsymbol{x}^*\right) \\
\vdots & \ddots & \vdots \\
\frac{\partial h_m}{\partial x_1}\left(\boldsymbol{x}^*\right) & \cdots & \frac{\partial h_m}{\partial x_n}\left(\boldsymbol{x}^*\right)
\end{pmatrix}$$

Kuhn Tucker's SOCs

Suppose $\left(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right)$ satisfies CQ and FOCs.

Considering the **binding** constraints: If $\boldsymbol{v}^T D_x^2 L\left(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) \boldsymbol{v} < 0 \quad \forall \boldsymbol{v} \neq \boldsymbol{0}$ in $\mathbb{R}^n$ for which $D\boldsymbol{h}\left(\boldsymbol{x}^*\right) \boldsymbol{v} = 0$ and $D\ \boldsymbol{g}_E\left(\boldsymbol{x}^*\right) \boldsymbol{v} = 0$ then $\boldsymbol{x}^*$ is a strict local maximum.

[Negative definite for a maximum, positive definite for a minimum]

Check via Bordered Hessian

$$\begin{pmatrix} 0 & 0 & D\boldsymbol{h}\left(\boldsymbol{x}^*\right) \\ 0 & 0 & D\boldsymbol{g}_E\left(\boldsymbol{x}^*\right) \\ D\boldsymbol{h}\left(\boldsymbol{x}^*\right)^T & D\boldsymbol{g}_E\left(\boldsymbol{x}^*\right)^T & D_x^2 L\left(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) \end{pmatrix}$$

Check the sign of the last $n-(m+e)$ leading principal minors:

(a) If they alternate in sign ending with $(-1)^n$ then Q is negative definite $\left[\boldsymbol{v}^T A\boldsymbol{v} < 0 \ \ \forall \boldsymbol{v} \neq \mathbf{0}\right]$ and $\boldsymbol{x}^*$ is a local maximum.

(b) If they have the same sign as $(-1)^{e+m}$ then Q is positive definite $[\ \boldsymbol{v}^T A\boldsymbol{v} > 0 \ \ \forall \boldsymbol{v} \neq \mathbf{0}\ ]$ and $\boldsymbol{x}^*$ is a local minimum.

Equality & Inequality Constraints: Maximisation Example

$$max\ x^3 + y^3 - 3x - 3y + 3xy \text{ subject to } x^2 \leq 4 \text{ and } y^2 \leq 4 \;\; [2 \leq x \leq 2\ ,\ 2 \leq y \leq 2]$$

Weierstrass Theorem

Compact constraint hence a global max and min exist.

Constraint Qualification

Let $\boldsymbol{g}_E$ be all the constraints which bind at $\boldsymbol{x}^*$, and e be the number of binding constraints. The constraint qualification is satisfied if the matrix $D\boldsymbol{g}_E(\boldsymbol{x}^*)$ is of full rank e. Let $D\boldsymbol{g}$ be the same matrix of all constraints (binding and not),

$$D\boldsymbol{g} = \begin{pmatrix} \partial g_1/\partial x & \partial g_1/\partial y \\ \partial g_2/\partial x & \partial g_2/\partial y \end{pmatrix} = \begin{pmatrix} 2x & 0 \\ 0 & 2y \end{pmatrix}$$

Now we need to consider what binds:

- If $x^2 = 4$ and $y^2 < 4$ then $D\boldsymbol{g}_E = \begin{pmatrix} 2x & 0 \end{pmatrix}$ with $x = +2$ or -2 therefore rank is 1.
  - Which is what it needs to be, since $n - e = 2 - 1 = 1$,
  - Also the fact $x = +2$ or -2 is important because if the Jacobian was 0 it would have rank 0 (hence not full rank).

- If $y^2 = 4$ and $x^2 < 4$ then $D\boldsymbol{g}_E = \begin{pmatrix} 0 & 2y \end{pmatrix}$ with $y = +2$ or -2 therefore rank is 1.
  - Which is what it needs to be, since $n - e = 2 - 1 = 1$,
  - Also the fact $y = +2$ or -2 is important because if the Jacobian was 0 it would have rank 0 (hence not full rank).

- If $x^2 = 4$ and $y^2 = 4$ then $D\boldsymbol{g}_E = D\boldsymbol{g}$ with x and $y = +2$ or -2 therefore rank is 2.

Hence CQ holds: In this circumstance it cannot fail.

Of course if no constraints bind then we have no constraint qualification to look at, we just have an unconstrained maximisation problem!

Kuhn Tucker FOCs

$$L(x, y, \lambda_1, \lambda_2) = f(x, y) - \lambda_1 \left(x^2 - 4\right) - \lambda_2 \left(y^2 - 4\right)$$

[For maximisation problems it MUST be the case that constraints are less than or equal to zero, that is $x^2 - 4 \leq 0$]

FOCs:

$$\begin{aligned} &FOC_x : 3x^2 - 3 + 3y - 2\lambda_1 x = 0 \\ &FOC_y : 3y^2 - 3 + 3x - 2\lambda_2 y = 0 \\ &CS_x : \lambda_1 \geq 0\ ,\ x^2 \leq 4\ ,\ \lambda_1 \left(x^2 - 4\right) = 0 \\ &CS_y : \lambda_2 \geq 0\ ,\ y^2 \leq 4\ ,\ \lambda_1 \left(y^2 - 4\right) = 0 \end{aligned}$$

Possible cases:

(1) Neither constraint binds

- Both Lambdas are zero.
- Function is has critical points at: $3x^2 - 3 + 3y = 0$, $3y^2 - 3 + 3x = 0 \Rightarrow y = (1-x)^2$, $y^2 - 1 + x = 0$
- Using SOCs as well as the FOCs above the function is maximised at $\left(\frac{-1-\sqrt{5}}{2}, \frac{-1-\sqrt{5}}{2}\right)$ for a value of 9.1.

(2) Both constraints bind

- This implies: $\lambda_1 \geq 0$, $x^2 = 4$, $\lambda_2 \geq 0$, $y^2 = 4$.
- FOC_x can be rearranged as $\lambda_1 = \frac{9+3y}{2x}$,
  - With $y = +2$ or -2 the numerator is +ve hence denominator must be +ve hence $x = +2$.
- FOC_y can be rearranged as $\lambda_2 = \frac{9+3x}{2y}$,
  - With $x = +2$ or -2 the numerator is +ve hence denominator must be +ve hence $y = +2$.
- So our critical point is $(2, 2)$, with $n = 2$ and $e = 2$, hence $n - e = 0$ and so nothing to check, we have a constrainted maximum at $f(2, 2)$.

(3) One constraint binds

- Say x binds and y doesn't, so $x^2 = 4$ and $y^2 < 4$
- Again FOC_x can be written as $\lambda_1 = \frac{9+3y}{2x}$, with $y = +2$ or -2 the numerator is positive hence denominator must be +ve hence $x = +2$.
- But there is no real solution for FOC_y which becomes $y^2 - 1 + 2 = 0$.
- By symmetry this will be the same for the other way around.

Evaluate and Compare

NO NEED TO CHECK THEM SOCs. Recall that the Weierstrass theorem held, hence we don't have to check any SOCs since we know that a global maximum must exist, and it must be one of the points found by the FOCs.

Therefore just evaluate and see at which point the function is larger.

It turns out it is $f(2, 2)$.

Equality & Inequality Constraints: Minimisation Example

$$min\ x^3 + y^3 - 3x - 3y + 3xy \text{ subject to } x^2 \leq 4 \text{ and } y^2 \leq 4 \ \ [2 \leq x \leq 2\ ,\ 2 \leq y \leq 2]$$

Weierstrass Theorem

Compact constraint hence a global max and min exist.

Constraint Qualification

Let $\boldsymbol{g}_E$ be all the constraints which bind at $\boldsymbol{x}^*$, and e be the number of binding constraints. The constraint qualification is satisfied if the matrix $D\boldsymbol{g}_E(\boldsymbol{x}^*)$ is of full rank e. Let $D\boldsymbol{g}$ be the same matrix of all constraints (binding and not),

$$D\boldsymbol{g} = \begin{pmatrix} \partial g_1/\partial x & \partial g_1/\partial y \\ \partial g_2/\partial x & \partial g_2/\partial y \end{pmatrix} = \begin{pmatrix} 2x & 0 \\ 0 & 2y \end{pmatrix}$$

Now we need to consider what binds:

- If $x^2 = 4$ and $y^2 < 4$ then $D\boldsymbol{g}_E = \begin{pmatrix} 2x & 0 \end{pmatrix}$ with $x = +2$ or -2 therefore rank is 1.
  - Which is what it needs to be, since $n - e = 2 - 1 = 1$,
  - Also the fact $x = +2$ or -2 is important because if the Jacobian was 0 it would have rank 0 (hence not full rank).
- If $y^2 = 4$ and $x^2 < 4$ then $D\boldsymbol{g}_E = \begin{pmatrix} 0 & 2y \end{pmatrix}$ with $y = +2$ or -2 therefore rank is 1.
  - Which is what it needs to be, since $n - e = 2 - 1 = 1$,
  - Also the fact $y = +2$ or -2 is important because if the Jacobian was 0 it would have rank 0 (hence not full rank).
- If $x^2 = 4$ and $y^2 = 4$ then $D\boldsymbol{g}_E = D\boldsymbol{g}$ with x and $y = +2$ or -2 therefore rank is 2.

Hence CQ holds: In this circumstance it cannot fail.

Of course if no constraints bind then we have no constraint qualification to look at, we just have an unconstrained maximisation problem!

Kuhn Tucker FOCs

$$L(x, y, \lambda_1, \lambda_2) = f(x, y) - \lambda_1 \left(4 - x^2\right) - \lambda_2 \left(4 - y^2\right)$$

[For minimisation problems it MUST be the case that constraints are greater than or equal to zero, that is $0 \leq 4 - x^2$]

FOCs:

$$\begin{aligned} &FOC_x : 3x^2 - 3 + 3y + 2\lambda_1 x = 0 \\ &FOC_y : 3y^2 - 3 + 3x + 2\lambda_2 y = 0 \\ &CS_x : \lambda_1 \geq 0\ ,\ x^2 \leq 4\ ,\ \lambda_1 \left(x^2 - 4\right) = 0 \\ &CS_y : \lambda_2 \geq 0\ ,\ y^2 \leq 4\ ,\ \lambda_1 \left(y^2 - 4\right) = 0 \end{aligned}$$

Possible cases:

(1) Neither constraint binds

- Both Lambdas are zero.
- Using SOCs as well as the FOCs above the function is minimised at $\left(\frac{-1+\sqrt{5}}{2}, \frac{-1+\sqrt{5}}{2}\right)$.

(2) Both constraints bind

- This implies: $\lambda_1 \geq 0$, $x^2 = 4$, $\lambda_2 \geq 0$, $y^2 = 4$.
- FOC_x can be rearranged as $\lambda_1 = -\frac{9+3y}{2x}$,
  - With $y = +2$ or -2 the numerator is +ve hence denominator must be +ve hence $x = -2$.
- FOC_y can be rearranged as $\lambda_2 = -\frac{9+3x}{2y}$,
  - With $x = +2$ or -2 the numerator is +ve hence denominator must be +ve hence $y = -2$.
- So our critical point is $(2-, -2)$, with $n = 2$ and $e = 2$, hence $n - e = 0$ and so nothing to check, we have a constrainted maximum at $f(-2, -2)$.

(3) One constraint binds

- Say x binds and y doesn't, so $x^2 = 4$ and $y^2 < 4$
- Again FOC_x can be written as $\lambda_1 = -\frac{9+3y}{2x}$, with $y = +2$ or -2 the numerator is positive hence denominator must be +ve hence $x = +2$.
- Using $FOC_y : y^2 - 1 - 2 = 0$ which implies $y = \pm\sqrt{3}$.
- **Bordered Hessian**

$$Dg_E = \begin{pmatrix} 2x & 0 \end{pmatrix} \;,\; D^2L = \begin{pmatrix} 6x + 2\lambda_1 & 3 \\ 3 & 6y + 2\lambda_2 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -4 & 0 \\ -4 & -12 + 2\lambda_1 & 3 \\ 0 & 3 & 6y \end{pmatrix}$$

  - $n = 2$ and $e = 1$, hence we need to check the last determinant. Specifically we need $(-1)^e$ for a minimum.
  - Insert the values for y and λ_2.
- Solution is $y = \sqrt{3}$ hence we get $(-2, \sqrt{3})$ and by symmetry $(\sqrt{3}, 2)$.

Evaluate and Compare

Therefore just evaluate and see at which point the function is smaller.

It turns out it is both $(-2, \sqrt{3})$ and $(\sqrt{3}, 2)$.

Concavity and Convexity

So far we have used the necessary FOCs and the sufficient SOCs to find local optima. Now we look to see under what circumstances we can guarantee that a local optima is also a global one.

Concavity/Convexity

Consider a convex set $S \subseteq \mathbb{R}^n$. A realvalued function $f : S \to \mathbb{R}$ is [strictly] concave (convex) if for any $x, y \in S$ and any $t \in (0,1)$,

$$f(tx + (1-t)y) \geq (\leq) t f(x) + (1-t) f(y)$$

[Concave if all points of the function lie above the straight line drawn between two points]

Example Concavity Proof

$$\begin{aligned}
&f(x,y) = 1 - x^2 \\
&\text{let } (x_1, y_1) \text{ and } (x_2, y_2) \\
\\
&f\left(\lambda x_1 + (1-\lambda)x_2\ ,\ \lambda y_1 + (1-\lambda)y_2\right) \geq \lambda f\left(x_1, y_1\right) + (1-\lambda) f\left(x_2, y_2\right) \\
&\lambda \in [0,1] \\
\\
&1 - \left[\lambda x_1 + (1-\lambda)x_2\right]^2 \geq \lambda\left(1 - x_1^2\right) + (1-\lambda)\left(1 - x_2^2\right) \\
&-\lambda^2 x_1^2 - 2(1-\lambda)\lambda x_2 x_1 - (1-\lambda)^2 x_2^2 \geq -\lambda x_1^2 - x_2^2 + \lambda x_2^2 \\
&(1-\lambda)\left[\lambda x_1^2 - x_2^2[1 - 1 + \lambda] - 2\lambda x_2 x_1\right] \geq 0 \\
&(1-\lambda)\lambda\left[x_1^2 - 2x_2x_1 + x_2^2\right] \geq 0 \\
&(1-\lambda)\lambda\left(x_1 - x_2\right)^2 \geq 0
\end{aligned}$$

Definiteness Definition of Concavity/Convexity

If $f : S \to \mathbb{R}$ is C^2

$D^2 f$ is negative semidefinite for all x in $S \Leftrightarrow f$ is concave on S.

$D^2 f$ is negative definite for all x in $S \Rightarrow f$ is concave on S.

$D^2 f$ is positive semidefinite for all x in $S \Leftrightarrow f$ is convex on S.

$D^2 f$ is positive definite for all x in $S \Rightarrow f$ is convex on S.

Global Optima

Let f be (strictly) concave/convex

If a local maximum/minimum exists, it is a (unique) global maximum/minimum.

If a stationary point exists, it is a (unique) global maximum/minimum.

[Note we need to look at SOCs for boundary optima though - the FOCs are not sufficient because of the constrained (boundary) optima cases]

Lagrangian Sufficiency

If there exists $\boldsymbol{x}^*$ and $\boldsymbol{\lambda}^* \geq \mathbf{0}$ such that,

$$L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \geq L(\boldsymbol{x}, \boldsymbol{\lambda}^*) \quad \forall \boldsymbol{x}$$
$$\boldsymbol{\lambda}^* (\boldsymbol{b} - g(\boldsymbol{x}^*)) = 0$$

And $\boldsymbol{x}^*$ is feasible, then $\boldsymbol{x}^*$ is optimal for the maximisation problem.

Sufficient FOC

Let f be concave and each g_j be convex.

The Kuhn-Tucker FOCs are sufficient for a global maximum.

Notes

Idea behind this: *if f is concave and g_j convex*, and further $\boldsymbol{\lambda}$ *is great than or equal to zero*, then the **Lagrangian is a concave function** of $\boldsymbol{x}$. Given that the KT conditions are satisfied then $\boldsymbol{x}^*$ and $\boldsymbol{\lambda}^*$ are stationary points of the Lagrangian. Since the Lagrangian is concave these points must maximise it globally.

Quasi-Concavity/Quasi-Convexity

A function f is quasiconcave if,

$$f(tx + (1-t)y) \geq \min\{f(x), f(y)\} \quad \forall t \in (0,1)$$

A function f is quasiconvex if,

$$f(tx + (1-t)y) \leq \max[f(x), f(y)\} \quad \forall t \in (0,1)$$

[Properties are strict if inequalities are strict]

Notes

Essentially imagine a straight line drawn from the y axis across and through $f(x)$ and $f(y)$. if all the values of the function between these two points lie above the lowest point out of $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{y})$ then it is quasiconcave.]

Some properties:

- Function f is quasiconcave iff f is quasiconvex
- If function f is concave then f is quasiconcave
- If f is convex then f is quasiconvex
- Any (strictly) monotonic function of one variable is both (strictly) quasiconcave and quasiconvex.

Global Optima

Let f be strictly quasiconcave.

If a local maximum exists, it is a unique global maximum.

Sufficient FOC

Let f be quasiconcave and let each g_j be quasiconvex.

The KuhnTucker FOCs are sufficient for a global maximum (at $\boldsymbol{x}^*$) provided at least one of the following holds.

(1) f is concave

OR

(2) $Df(\boldsymbol{x}^*) \neq 0$

Testing for Concavity/Convexity

2x2 Matrix

$\det A < 0$ eigenvalues have opposite signs.

$\det A > 0$ eigenvalues have the same sign,

  $\text{tr}(A) > 0$ they are all positive,

  $\text{tr}(A) < 0$ they are all negative.

Eigenvalue test

A is positive definite iff all its eigenvalues are strictly positive.

A is positive semidefinite iff all its eigenvalues are weakly positive.

A is negative definite iff all its eigenvalues are strictly negative.

A is negative semidefinite iff all its eigenvalues are weakly negative.

A is indefinite otherwise.

Determinant test

A is positive definite iff $|A_k| > 0$ for all $k = 1, \ldots, n$

A is positive semidefinite iff $|A_k| \geq 0$ for all $k = 1, \ldots, n$ and all row + col permutations.

A is negative definite iff $(-1)^k |A_k| > 0$ for all $k = 1, \ldots, n$

A is negative semidefinite iff $(-1)^k |A_k| \geq 0$ for all $k = 1, \ldots, n$ and all row + col permutations.

A is indefinite otherwise.

Notes

Where A_k is a square submatrix of A retaining only the first $\boldsymbol{k}$ rows and columns.

$$A_1 = (a_{11}) \ , \ A_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \ , \ \ldots \ , \ A_k = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}$$

Note also in the ND and NSD cases the determinants of the submatrices need to oscillate.

Result

$D^2 f$ is negative semidefinite for all $\boldsymbol{x}$ in $S \Leftrightarrow f$ is concave on S.

- $NSD =$ concave

$D^2 f$ is negative definite for all x in $S \Rightarrow f$ is concave on S.

- $ND =$ concave

$D^2 f$ is positive semidefinite for all $\boldsymbol{x}$ in $S \Leftrightarrow f$ is convex on S.

- $PSD =$ convex

$D^2 f$ is positive definite for all $\boldsymbol{x}$ in $S \Rightarrow f$ is convex on S.

- $PSD =$ convex

Envelope Theorems

$M(\boldsymbol{a})$ is the maximum value function and we assume $\boldsymbol{x}^*(\boldsymbol{a})$ is the optimal value of $\boldsymbol{x}^*$ in terms of $\boldsymbol{a}$.

$$\begin{aligned}
&\max_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{a}) \text{ subject to } \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{a}) \leq \boldsymbol{b} \\
&M(\boldsymbol{a}) = \{\max f(\boldsymbol{x}, \boldsymbol{a}) \mid \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{a}) \leq \boldsymbol{b}\} \\
&M(\boldsymbol{a}) = f\left(\boldsymbol{x}^*(\boldsymbol{a}), \boldsymbol{a}\right)
\end{aligned}$$

Envelope Theorem

$$\frac{\partial M(\boldsymbol{a})}{\partial a_i} = \frac{\partial f\left(\boldsymbol{x}^*(\boldsymbol{a}), \boldsymbol{a}\right)}{\partial a_i}$$

An Example:

$$\begin{aligned}
&f(x, a, b) = 4 + 2bx - ax^2 \\
&f_x\left(x^*, a, b\right) = 2b - 2ax^* = 0 \Rightarrow x^*(a, b) = \frac{b}{a} \\
&ET: \frac{\partial M(a, b)}{\partial a} = \frac{\partial f\left(x^*(a, b), a, b\right)}{\partial a} = -x^*(a, b)^2
\end{aligned}$$

See the usefulness of this: we never even needed to calculate the max value function.

Envelope Theorem for Constrained Problems

$$\begin{gathered}
\text{For, } L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{a}) = f(\boldsymbol{x}, \boldsymbol{a}) - \boldsymbol{\lambda}(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{a}) - \boldsymbol{b}) \\
\frac{\partial M(\boldsymbol{a})}{\partial a_i} = \frac{\partial L\left(\boldsymbol{x}^*(\boldsymbol{a}), \boldsymbol{\lambda} * (\boldsymbol{a}), \boldsymbol{a}\right)}{\partial a_i}
\end{gathered}$$

Interpretation of Lagrange Multipliers

λ_j represents the effect on the objective function of relaxing the constraint $g_j(\boldsymbol{x}) \leq b_j$

This is in fact a special case of the ET:

$$\begin{aligned}
&\frac{\partial M}{\partial b_j} = \underbrace{\frac{\partial f}{\partial b_j}\left(\boldsymbol{x}^*(\boldsymbol{b})\right)}_{=\frac{\partial f}{\partial \boldsymbol{x}^*}\frac{\partial \boldsymbol{x}^*}{\partial b_j}=0} - \sum_i \lambda_i^*(\boldsymbol{b}) \frac{\partial}{\partial b_j}\left(g_i\left(\boldsymbol{x}^*(\boldsymbol{b})\right) - b_i\right) \\
&\frac{\partial M}{\partial b_j} = 0 - \lambda_j^*(\boldsymbol{b})(0 - 1) = \lambda_j^*
\end{aligned}$$

Decision Theory: Preferences & Utility

Preferences

Generally considered to be:

- Ordinal: Says which option is better without quantifying the preference.
- Stable: Do not change from choice problem to choice problem.

Binary Relation: Given X (a set of economic objects), a binary relation is a description of pairs of elements in X that are related.

- Denoted $(x, y) \in B$, or xBy
- When binary relations represent preferences we can use $\succ$, $\succsim$, P

Preference relations are said to be *rational* if they have this set of properties:

- **Reflexivity:** For every object $x \in X, xBx$ must be true.
  - You don't actually need reflexivity since completeness implies reflexivity.
- **Completeness:** For every $x, y \in X$, either xBy, yBx or both must be true.
  - Completeness is made up of connectedness and reflexivity. To be connected is to mean every pair of distinct (different) elements can be related in one way or another. To be reflexive is to mean that every distinct element is related to itself.
- **Transitivity:** For every $x, y, z \in X$, whenever xBy and yBz, it must also be true that xBz

Strict (part of) Rational Preference: Given the rational preference $\succsim$, the strict part of $\succsim$ is the binary relation $\succ$ defined by, $x \succ y$ iff $x \succsim y$ but not $y \succsim x$.

- These preferences are:
  - Always transitive.
  - May or may not be connected. [If you were indifferent between two elements then the relation would not be connected, if you were indifferent between no elements then the relation would be connected.]
  - Never reflexive.
  - Always antisymmetric.

Note: Antisymmetric means that for any *distinct* x and y, xBy and not yBx, but notice that when $x = y$ then xBy and yBx and yet we still have antisymmetry. Asymmetry says for *any* x and y, xBy and not yBx.

Indifferent (part of) Rational Preference: Given the rational preference $\succsim$, the indifferent part of $\succsim$ is the binary relation $\sim$ defined by, $x \sim y$ iff $x \succcurlyeq y$ and $y \succsim x$.

- Often called an equivalence relation, these preferences are:
  - Always transitive.
  - Always reflexive.
  - Always symmetric.

Equivalence class: On X is given by $[a] = \{x \in X \mid x \sim a\}$, that is the set of all x in X such that x and a are equivalent to one another.

- We might also use the notion of the quotient set/space which is the set of equivalence classes, denoted by $X/\sim$.
- For example, consider X as the set of cars and the equivalence relation as 'has the same colour as', in this case $X/\sim$ denotes the set of all the colours of cars.

Utility

Utility Representation

Given the rational preference $\succsim$ over X, we say that the real-valued function $u : X \to \mathbb{R}$ is a utility representation (utility function) whenever, for every $x, y \in X$:

$$x \succsim y \text{ iff } u(x) \geq u(y)$$

Note that while only rational preferences can have a utility representation; rational preferences do not imply a utility representation if they exist over an infinite set of objects.

Transformations

Rational preferences are ordinal and, if $u(.)$ represents this preference, then any monotone transformation of $u(.)$ also represents the same preference.

That is, providing $f : \mathbb{R} \to \mathbb{R}$ is strictly increasing,

$$x \succsim y \text{ iff } u(x) \geq u(y) \text{ iff } f(u(x)) \geq f(u(y))$$

Maximal Elements

A finite set always has a maximal element.

That is given that $\succsim$ is a rational preference over a finite set X every subset $A \subseteq X$ has a maximal element.

For an infinite set to have a maximal element,

(1) Rational preferences $\succsim$ must be continuous.

(2) Subsets must be compact (closed and bounded) - $[0, 1]$ rather than $(0, 1)$

- If not bounded then we must have an argument to determine that the solution must belong to the compact set $[-K, K]$ and then we can go from there.

Conditions for representation

(1) **Only rational binary relations** can be represented by a utility function.

- Let B be a binary relation on X, and $u : X \to \mathbb{R}$ such that, for every $x, y \in X, xBy$ iff $u(x) \geq u(y)$. Then, B must be a rational preference.
- Proof
  - B must be complete because, for any pair x, y, it is either $u(x) \geq u(y), u(y) \geq u(x)$, or both. Similarly B must be transitive because if xBy and yBz it must be the case that $u(x) \geq u(y) \geq u(z)$, and hence $u(x) \geq u(z)$ which implies xBz as desired.

(2) If X **is finite, every rational preference** $\succsim$ can be represented.

(3) If the rational preference $\succsim$ has a **finite number of equivalence classes**, the preference can be represented.

- Equivalence class: $[a] = \{x \in X \mid x \sim a\}$ the set of all x in X such that x and a are equivalent to one another.
- This conditions mean that not all rational preferences can be represented:
  - E.g. Lexicographic preferences:

$$\{(x_1, x_2) : x_1, x_2 \in [0, 1]\}$$
$$\text{where } (x_1, x_2) > (y_1, y_2) \text{ iff } x_1 > y_1 \text{ or } (x_1 = y_1 \text{ and } x_2 > y_2)$$

Revealed Preference

A choice problem is a subset of alternatives $A \subseteq X$ over which the analyst can observe the behaviour of the individual. It is often called the ***menu***.

If we let $\Delta = \{A, B, \ldots\}$ be the menus over which the analyst has information. Then **choice data** is defined as a map of the form $c : \Delta \to X$ with the obvious assumption that $c(A) \in A$.

- **Universal choice data** is such that $\Delta = \chi$ where χ is the collection of all subsets of X. [Essentially the domain is all possible menus].
- **Binary choice data** is such that $\Delta = \beta$, where β is the collection of all choice problems formed by two alternatives.
- **Bundle consumption choice data** is such that Δ is formed by some budget sets of the form $p_1x_1 + \ldots + p_nx_n \leq m$

We then can learn about preferences from this choice data with revelation.

- **Direct Revelation:** Given choice data c, we say that x is *directly revealed preferred* to $y \neq x$ and write $x \succ_c y$ whenever it is the case that $x = c(A)$ and $y \in A$.
- **Indirect Revelation:** Given choice data c, we say that x is indirectly revealed preferred to $y \neq x$ whenever there is a sequence of elements $z_1 = x,\ z_2, \ldots,\ z_n = y$ such that $z_i = c\,(A_i)$ and $z_{i+1} \in A_i \backslash \{z_i\}$. That is whenever $z_1 = x \succ_c z_2 \succ_c \ldots \succ_c y = z_n$.
  - Where $S \backslash \{a\} := \{x \in S \mid x \neq a\}$, i.e. S backslash element a is the subset of all elements of S except a.

Binary and Universal choice data are both enough to learn all the preferences of the individual.

If X contains n elements, then $n - 1$ choice problems may be enough to learn all the preferences.

The analyst can make assumptions, for example if two things are identical but one is of worse quality then we can assume $x_h \succ x_l$, (high quality vs low quality).

Rational Choice Data

Choice data $c : \Delta \to X$ is said to be ***rationalizable*** if there exists a rational preference $\succsim$ such that $c(A)$ is always the maximal element in A according to $\succsim$.

OR

Choice data $c : \Delta \to X$ is said to be ***rationalizable*** if there exists a utility representation $u()$ such that $c(A)$ is always the maximal element in A according to $u()$.

Choice data is rationalizable iff [note the *iff*, hence both directions of the biconditional need proving] it satisfies,

(1) **Property $\boldsymbol{\alpha}$**

- Property alpha is satisfied by c if for any pair of choice problems $B \subseteq A$ it is the case that: If $c(A) \in B$, then $c(B) = c(A)$.
- [If you remove some options from the menu A but the maximal element remains then the maximal element of this new set should stay the same as what it was for A.]
- Proof of this:

  (A) Rationalizability implying Property alpha

  - Start from c being rationalizable. Hence there exists a preference $\succsim$ explaining all choices. Hence $c(A)$ is the maximal element of A. It is obvious then that $c(A) > y$ holds for all other elements in A. From this given B is a subset of A, containing $c(A)$, then hence $c(A) > y$ for all elements y in B, hence $c(A)$ must be the optimal element of B.

  (B) Property alpha implying Rationalizability

  - Start from c satisfying the property α. We can start with $X_1 = X$ and $x_1 = c\left(X_1\right)$. This is the best alternative and property α guarantees that it is chosen when available. Now consider $X_2 = X_1 \backslash c\left(X_1\right)$ and $x_2 = c\left(X_2\right)$. This the best remaining alternative, and property α guarantees it is chosen. Therefore $x_1 \succ x_2$ Then repeat this process such that $x_1 \succ x_2 \succ \ldots \succ x_n$

(Notice that this proof requires that X is finite).

(2) **Weak Axiom of Revealed Preference**

- Menu A and B cannot reveal $x > y$ and $y > x$.

Bundles

Uses the space $X = \mathbb{R}^n_+$ and it tends to be the case that $n = 2$.

In this space the variation of menus tends to be because of variation in income and/or prices.

- **Strict Monotonicity:** We say that $\succsim$ is strictly monotone if $x_1 \geq x_2$ and $y_1 \geq y_2$[2] with at least one of them strict, implies that $(x_1, y_1) > (x_2, y_2)$.

Varying *just* price or income to test rationality is generally unhelpful,

- In the case of just income the choice of the big set is never available in the small set, hence property α or the weak axiom never say anything.
- In the case of just price the choice of the big set is not available in the small set aside for the very extreme case of the choice being on the axis (in this case property α and therefore rationality can be rejected).

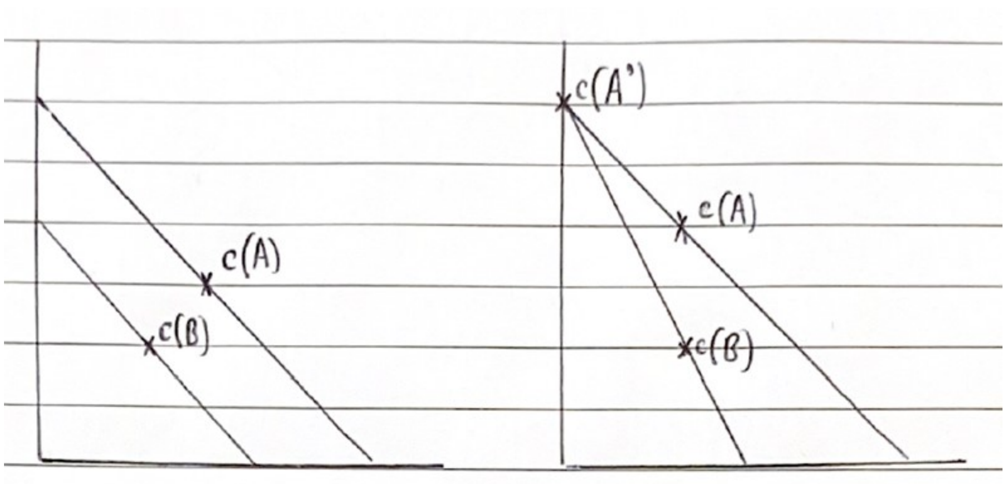

Figure 1: Left: varying only incomes; Right: varying only prices (we can learn something but only if the choice for menu A is c(A'))

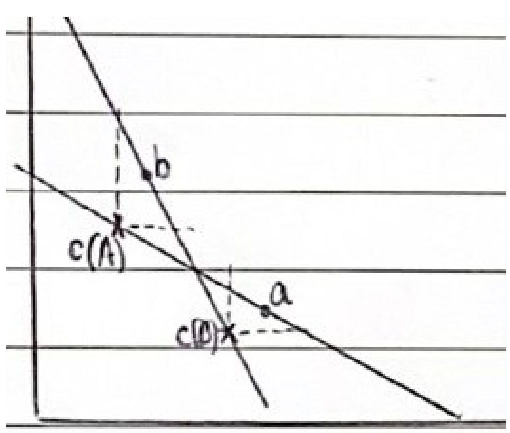

Figure 2: Varying price and income

Combined variations are informative,

- These sets of preferences plus monotonicity plus weak axiom can be shown to be not rational. If $c(A)$ is preferred to a, and $c(B)$ is preferred to b, but b is better than $c(A)$ and a is better than $c(B)$.
- Hence $c(A)$ is preferred to $c(B)$ and $c(B)$ is preferred to $c(A)$.
- Which implies inconsistency.

(Strict) Convexity: For every x, y such that $x > y$ and every $\alpha \in (0, 1)$, $x \succ \alpha x + (1 - \alpha)y > y$. - [If x is preferred to y then a mixture of some x and some y is still preferred to y averages are better than extremes.]

Cobb-Douglas: The fraction of income spent in each good is constant. Test by varying m and checking fraction is constant.

Quasi-Linearity: Utility in the representation $U(x, m) = u(x) + m$.

Time

Use the simplest set $X = \mathbb{R}^2$.

An element (x, t) describes the amount of money x to be paid at time t.

- **Discounted Utility:** With $u : \mathbb{R} \to \mathbb{R}$ and $\delta \in (0, 1]$, discounted utility says that $(x, t) \succsim (y, s)$ iff $U(x, t) = \delta^t u(x) \geq \delta^s u(y) = U(y, s)$.

Discounted utility is a ***rational preference***.

It also has the property of:

- **Stationarity**: We say that the preference $\succsim$ over dated prizes satisfies stationarity if for every x, y, t, s, r, it is the case that: $(x, t) \succsim (y, s)$ iff $(x, t + r) \succsim (x, s + r)$.

Other important properties

- **Single-Peakedness**: Let $X = [0, 1]$. We say that $\succsim$ is single peaked if there exists x^* such that:

$$y_2 < y_1 < x^* \Rightarrow y_1 \succ y_2 \text{ and } x^* < y_1 < y_2 \Rightarrow y_1 \succ y_2$$

$y(1)$ is always preferred to $y(2)$ when it is closer to the bliss point. i.e. the ordering is always $y(2), y(1), x^*, y(1), y(2)$.

Decisions Under Risk

Probability

Sample space

The set of all possible results of an experiment, denoted Z.

Finite Random Variables

A FRV over Z is a variable that can adopt, randomly, a finite number of different values of the sample space (often called the support of a random variable), with some preassigned probability (often called the probability mass of each possible result).

It can hence be defined as a map (function),

$p : Z \to [0,1]$

Where $p(z) = 0$ except for a finite number of results, and with,

$\sum_{z \in Z} p(z) = 1$

Expected Value

$$EV(p) = \sum_{z \in Z} p(z)z$$

Given a FRV p and a map $h : Z \to Z$, we can denote by p_h the FRV that assigns probability $p(z)$ to the result $h(z)$.

If h is a linear map: $h(z) = a + bz$ then $EV(p_h) = a + bEV(p)$

Variance

$$\operatorname{Var}(p) = \sum_{z \in Z} p(z)(z - EV(p))^2$$

Given a FRV p and a map $h : Z \to Z$, we can denote by p_h the FRV that assigns probability $p(z)$ to the result $h(z)$.

If h is a linear map: $h(z) = a + bz$ then $\operatorname{Var}(p_h) = b^2 \operatorname{Var}(p)$

Jensen's Inequality

If h is strictly concave, then $EV(p_h) < h(EV(p))$.

If h is strictly convex, then $h(EV(p)) < EV(p_h)$.

Intuitive proof: If there are only two possible results z_1 and z_2 in an experiment, and $h(.)$ is concave, then,

$EV(p_h) = p(z_1)\, h(z_1) + [1 - p(z_1)]\, h(z_2)$.

This is strictly less than,

$h\Big(p(z_1)\, z_1 + [1 - p(z_1)]\, z_2\Big) = h(EV(p))$.

Jenson's inequality can be rewritten in a few useful ways:

For $\varphi(.)$ concave { convex } and p_i's as positive weights of x_i's such that $\sum_{i=1}^{n} p_i = 1$,

$$\sum_{i=1}^{n} p_i \varphi(x_i) \leq \varphi\left(\sum_{i=1}^{n} p_i x_i\right) \quad , \quad \left\{\varphi\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i \varphi(x_i)\right\}$$

For $\varphi(.)$ concave { convex } and a_i's as positive weights of x_i's (not summing to one),

$$\frac{\sum_{i=1}^{n} a_i \varphi(x_i)}{\sum_{i=1}^{n} a_i} \leq \varphi\left(\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i}\right) \quad , \quad \left\{\varphi\left(\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i}\right) \leq \frac{\sum_{i=1}^{n} a_i \varphi(x_i)}{\sum_{i=1}^{n} a_i}\right\}$$

For $\varphi(.)$ concave { convex } and a_i 's as positive weights of x_i 's where $a_i = a \ \forall i$,

$$\frac{\sum_{i=1}^{n} \varphi(x_i)}{n} \leq \varphi\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \quad , \quad \left\{\varphi\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \leq \frac{\sum_{i=1}^{n} \varphi(x_i)}{n}\right\}$$

Events

An event E is a subset of the sample space. Given a finite random variable p, the probability of event E is simply $p(E) = \sum_{z \in t} p(z)$.

Independent events

E and F are independent events if $p(E \cap F) = p(E)p(F)$.

Conditional probability

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}$$

Bayes' Rule

$$p(E \mid F) = \frac{p(F \mid E)p(E)}{p(F)}$$

Cumulative Distribution Function

The map $F : Z \to [0, 1]$ where $F(z)$ is the total mass of experiencing a result below or equal to z.

Continuous Random Variables

A CRF is a continuous and increasing map $F : Z \to [0, 1]$ with values 0 and 1 in the lower and upper limits of the sample space.

The CDF is given by $F(z)$ and the pdf by $f(z)$, where:

$$\text{For } a \leq Z \leq b \ \ F(t) = P(z \leq t) = \int_{a}^{t} f(z)dz$$

$$EV(f) = \int_{\infty}^{-\infty} zf(z)dz$$

$$\text{Var}(f) = \int_{\infty}^{-\infty} (z - EV(f))^2 f(z)dz$$

Rational Preferences over Lotteries

Let Z be a finite space of prizes.

Lottery: A lottery is a (finite) random variable over the prize space.

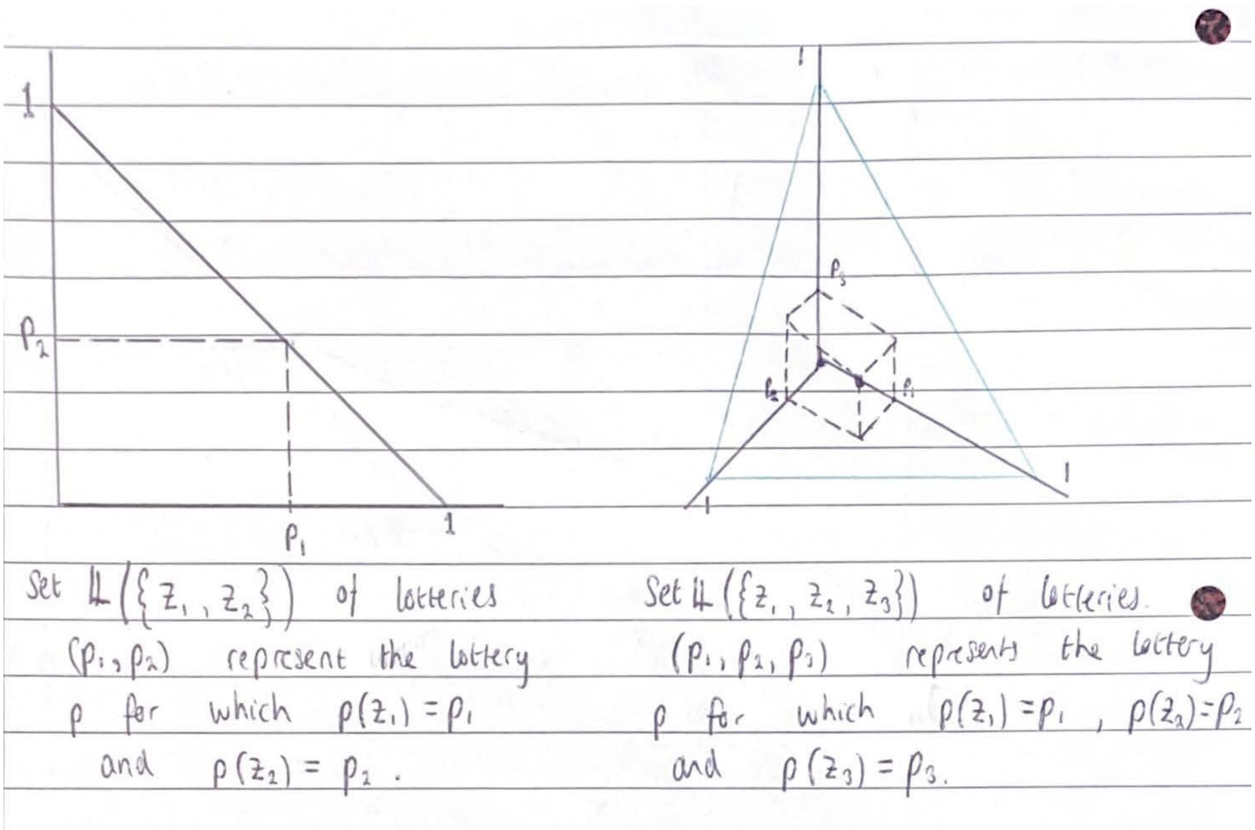

Figure 3: Lotteries

We now study preferences $\succsim$ over $L(Z)$.

Rational Preferences: We assume that individuals have rational preferences over $L(Z)$, i.e. $\succsim$ is transitive and complete.

- Rational preferences do not always guarantee that a utility function exists, this is the case again for rational preferences over lotteries even though the prize space is finite!
- Example: Lexicographic preferences
  - First compare the probability of z_1, if equal then compare $z_{2'}$, if equal then z_3 and so on.
  - These preferences are clearly transitive and complete (rational).
  - We can't construct a utility function though.
- Example: 'Minimising mental contemplation'
  - An individual prefers a lottery if there are less results to be contemplated (less mental burden)
  - These preferences are rational and the utility function is $U(p) =$ the number of prizes with zero probability.

Expected Utility Theorem

The binary relation $\succsim$ is rational and satisfies the properties of independence and continuity iff it can be represented by an expected utility.

$$p \succsim q \text{ iff } \sum_{z \in Z} p(z)v(z) \geq \sum_{z \in Z} q(z)v(z)$$

Properties

(1) Independence

Let p, q, r be three lotteries and $\lambda \in [0, 1]$. Let p' and q' be the simple lotteries associated, respectively, to the compound lotteries $\lambda p \oplus (1\lambda)r$ and $\lambda q \oplus (1\lambda)r$.

Then,

$p \succsim q$ iff $p' \succsim q'$

(2) Continuity

Let $[a]$ denote the lottery in which prize a is given with certainty. Suppose that $[a] > p > [c]$. Then there exists a λ_p such that the simple lottery associated to $\lambda_p \oplus (1\lambda_p)\,[c]$ is indifferent to p

Marschack-Marina Triangle

Suppose three possible outcomes x_1, x_2, x_3 such that $x_3 > x_2 > x_1$. These occur with probabilities p_1, p_2, p_3 where $\sum_l p_i = 1$. Because of the fact that $p_3 = 1p_1p_2$ these can be plotted on the unit triangle.

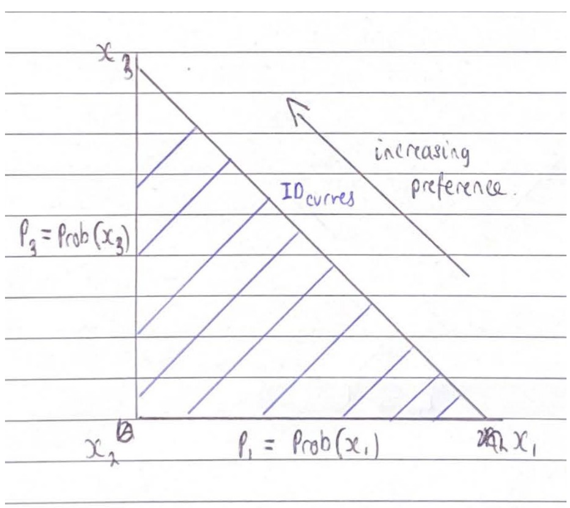

Figure 4: Marschack-Marina Triangle

- On the triangle a point = a lottery.
- Northwest = increasing utility (increasing probability of best option).
- $(0, 0) = x_2$ with certainty, $(1, 0) = x_2$, with certainty, $(0, 1) = x_3$ with certainty.

EUT implies that indifference curves must be parallel straight lines. Why?

(1) Linear in probabilities:

- Because when $U(p) = U(p')$ we know that $U(ap + (1a)p') = aU(p) + (1a)U(p')$ which is constant if they are the same utility and linear in probabilities.

(2) Independence:

- If $p \sim p'$ then $p \sim \alpha p + (1\alpha)p' \sim p'$ (straight line)

Continuity implies that for any lottery there exists a point on the frontier that is indifferent to it (a point where a combo of a better and worse lottery is indifferent to said lottery). A finite monetary lottery is a finite random variable over the monetary prize space.

Monetary Lotteries

A **finite monetary lottery** is a finite random variable over the monetary prize space.

We can reproduce the EUT with a function over monetary space $v : Z \to \mathbb{R}$ with,

$$p \succsim q \text{ iff } \sum_{z \in Z^p} p(z)v(z) \geq \sum_{z \in Z} q(z)v(z)$$
$$p \succsim q \text{ iff } \int_{z \in Z} p(z)v(z)dz \geq \int_{z \in Z} q(z)v(z)dz$$

Certainty equivalent

The certainty equivalent of a lottery is the amount of money that is indifferent to the lottery.

$$v(CE(p)) = EU(p)$$
$$CE(p) = v^1[EU(p)]$$

If v is monotone and continuous then every lottery has a certainty equivalent.

- Proof:
  - Consider z_l and z_h the lowest and highest payoffs in the support of a lottery. If v is monotone then the utility of z_l and z_h must be lower and higher respectively than the expected utility of the lottery. We can then use continuity to find the CE between them.

Risk Attitudes

Risk neutral if every lottery is indifferent to its expected value.

Risk averse if every lottery is worse than its expected value.

Risk loving if every lottery is better than its expected value.

A Bernoulli utility (utility function over monetary space) is risk averse iff v is concave.

- Proof:
  - Concavity implies that $\text{EV}(u(p)) < u(EV(p))$ i.e. the expected utility of every lottery is less than the expected utility of receiving the expected value (with certainty). So you would prefer the expected value than to play the lottery.

Arrow-Pratt Results

Measuring concavity of utility functions and hence risk aversion.

Need for normalisation:

- A property of Expected utility is that u and v are equivalent where $u(.)$ and $v(.) = \alpha + \beta u$ with $\beta > 0$.
- When this is the case we face a problem of using the second derivative to measure concavity since,

$$v'(x) = \beta u'(x) \text{ and } v''(x) = \beta u''(x) \text{ hence } v'' \neq u''$$

- By normalising with the first derivative we can remove the β.

Measures:

$$\text{Absolute Arrow-Pratt: } A(x) = -\frac{u''(x)}{u'(x)}$$

$$\text{Relative Arrow- Pratt: } R(x) = A(x)x = -\frac{u''(x)}{u'(x)}x$$

Alternatively:

If it is the case that $u_1 = f(u_2)$ with $f(.)$ concave $\Big[$ that means that $f''() < 0$ (second derivative less than zero $\Rightarrow$ getting more negative $\Rightarrow$ concave) $\Big]$ then u_1 is more concave than u_2.

Arrow-Pratt Theorem

The following statements are equivalent:

(1) For every lottery, the certainty equivalent associated to u_1 is smaller than that of u_2.

- (You would accept less money to remove the risk)

(2) u_1 is more concave than u_2.

- (More risk averse)

(3) The Absolute Arrow-Pratt measure of u_1 is, as a function, greater than that of u_2.

- (Higher $A(x)$ means more concave i.e. more risk averse)

(4) The risk premium associated to u_1 is larger than that of u_2.

Given wealth x, subject to additive, normally distributed, fluctuations with variance σ^2 and mean zero,

$$u(CE(p)) = u(x - RP(p)) = EU(p)$$

$$u(x - RP(p)) \approx u(x) - RP(p)u'(x) \text{ (by a first order Taylor approximation)}$$

$$\begin{aligned} EU(p) &\approx u(x) + u'(x)0 + \frac{1}{2}u''(x)\sigma^2 \text{ (by a second order Taylor approximation)} \\ &\approx u(x) + \frac{1}{2}u''(x)\sigma^2 \end{aligned}$$

Therefore,

$$\begin{aligned} u(x) - RP(p)u'(x) &\approx u(x) + \frac{1}{2}u''(x)\sigma^2 \\ RP(p) &\approx -\frac{1}{2}\frac{u''(x)}{u'(x)}\sigma^2 = \frac{1}{2}\sigma^2 A(x) \end{aligned}$$

(See microeconomics notes for a more complete version of this proof).

CARA: Constant Absolute Risk Aversion

Larger a means more risk averse.

$$\begin{aligned} &A(x) = a \\ &\text{Eg} : u_a(x) = -e^{ax} \text{ when } a \neq 0 \end{aligned}$$

DARA: Decreasing Absolute Risk Aversion

Attitude towards the lottery $\left[\frac{1}{2}, \frac{1}{2}; x + K, xH\right]$ is more favourable when x is larger

$$\text{Eg: } u(x) = \sqrt{x}\ ,\ A(x) = \frac{1}{2x}$$

CRRA: Constant Relative Risk Aversion

$R(x) = r$ where larger r implies more risk aversion.

$$\text{Eg: } u_r(x) = x^{1-r} \text{ with } r < 1$$

If the r is negative then we have a risk lover.

If $r = 0$ we have risk neutral.

When r tends to one we get logarithmic preferences.

Information Economics

Dynamic Optimisation & Differential Equations