Introduction to optimisation

Onno Bokhove

School of Mathematics, University of Leeds, UK
Room 8.04; E-mail: o.bokhove@leeds.ac.uk

Organisation

- See the handout online.
- Semester: 1; No. of credits: 10; Level: 2
- Prerequisites: MATH1010 or (MATH1050 & MATH1060) or (MATH1050 & MATH1331) or equivalent.
- All material will be available electronically via the MATH2640 Minerva pages.
- If you wish to have paper copies, then please email me, given our low-carbon use policy.

Introduction

- **What is optimisation**?
- Answer: Optimisation is about "the quest for the best", i.e. designing strategies for obtaining the best result. (Optimum is Latin for best.) In many cases of human endeavour we search for conditions under which we attain the optimum. So often, we are trying to find maxima (or minima).
- In economics, this can concern the highest profit, best rewards and maximum growth.
- In river or coastal flood-mitigation, this concerns mitigation or minimisation of flood damage, expressed in terms of staying below a specified water level at given locations, properties saved or economic damage saved, along a river or a coast.
- For wave-energy devices, this concerns maximising energy output given incoming waves.
- In real life, optimisation will include uncertainty, but we will not deal with the effects of uncertainty in this course.

Introduction

- In mathematics, optimisation is the subject of dealing with finding maxima or minima of functions.
- The definition of a function or functions: an expression or expressions in terms of variables $x, y, \ldots$ that can be altered, often also under additional conditons.
- Then we are tasked to find maxima or minima of the expression $f(x, y)$ (or expressions $f(x, y)$ and $g(x, y)$) subject to such additional conditions.
- Here the functions f, g can can describe the quantities we are interested in, such as profit, revenue, tax, energy output, flood damage, etc.
- We will (mainly) consider applications in economics.

Breakdown of the course

Chapter 1: Several variable calculus (~ 6 lectures)

- Develop techniques of multivariate functions, partial derivatives, total derivatives, gradients, directional derivatives, implicit differentiation, chain rule (in many variables), Taylor series, Hessian and stationary points.

Chapter 2: Unconstrained optimization (~ 6 lectures)

- Quadratic forms, eigenvalues, definiteness and principal minor test, stationary points, local extrema, applications to economics (Cobb-Douglas functions).

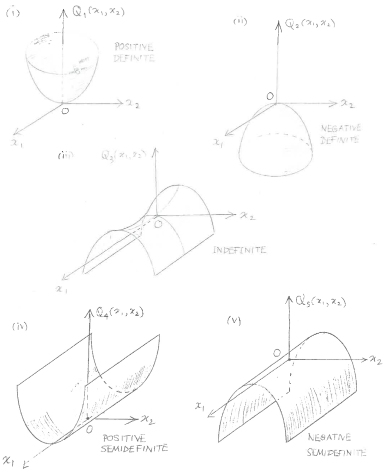
Chapters 3 & 4: Constrained optimization (~ 10 lectures)

- A) Equality constraints (conditions involving an equality $=$–symbol)
- B) Inequality constraints (conditions involving an inequality $<$–symbol)
- Lagrange multiplier formalism to find:
  - critical points and maximisers;
  - (bordered) Hessians to classify critical points;
  - Kuhn-Tucker theory; and,
  - applications to economics, etc.

Updates and feedback

- In 2018, the course structure was updated: e.g., 5 homework assignments.
- In 2019, some notes/assignment solutions will be typed out, with accompanying Python graphs.
- Every lecture, I will ask two volunteers to give me feedback immediately after the lecture (email/in person).
- There will be online questionnaires regarding comments and feedback, halfway and at the end of the course.
- Please feel free to ask questions.

2.2 Quadratic forms: sketches (by Prof M. Kelmanson)



2.5 Applications to economics & Cobb-Douglas function

- Unconstrained optimisation in economics: we are interested in optimising functions describing profits based on production, costings, revenues, etc. These can involve various kinds of quantities (denoted by $x_1, x_2, \ldots, x_n$, say) describing production outputs or variables influencing production. We use symbols:
- $R =$ revenue $R(x_1, \ldots, x_n)$
- $p_i =$ price per product
- $Q_i =$ production functions $\#$ outputs
- Π –profit function
- C –cost function (cost to produce items).

Applications to economics & Cobb-Douglas function

Situation $\#1$:

- A firm produces Q items per year, which sell at a price p per item
- The revenu is:

$$R(x_1, \ldots, x_n) = pQ(x_1, \ldots, x_n)$$

- R, Q may depend on variables x_1, x_2 influencing the production (such as $\#$ of employees, amount spent on equipment, etc.).
- Vector $(x_1, \ldots, x_n)$ is called <u>input bundle</u> vector.
- <u>Cost function</u> $C(x_1, \ldots, x_n)$ measures how much the firm spends on production, giving a negative contribution.

Applications to economics & Cobb-Douglas function

- Hence, the profit is given by:

$$\Pi(x_1, \ldots, x_n) = R - C = pQ(x_1, \ldots, x_n) - C(x_1, \ldots, x_n)$$

- To maximise profit we want to determine variables $x_1, \ldots, x_n$ such that Π has maximum value.
- Stationary points follow from first-order conditions (FOCs):

$$\frac{\partial \Pi}{\partial x_i} = 0 \Longrightarrow p\frac{\partial Q}{\partial x_i} = \frac{\partial C}{\partial x_i}, \quad \text{for} \quad i = 1, \ldots, n.$$

- A stationary points Hessian needs to be ND to maximise profit.

Applications to economics & Cobb-Douglas function

- In the case we have multiple products, we can have several production functions $Q_j(x_1, \ldots, x_n)$ each with its price p_j:

$$R(x_1, \ldots, x_n) = \sum_{j=1}^{m} p_j Q_j(x_1, \ldots, x_n) = \mathbf{p} \cdot \mathbf{Q}$$

  with cost function $C(\mathbf{x}) = C(x_1, \ldots, x_n)$.
- We then get as FOCs:

$$\Pi(\mathbf{x}) = \sum_{j=1}^{m} p_j Q_j - C = \mathbf{p} \cdot \mathbf{Q}(\mathbf{x}) - C(\mathbf{x})$$

$$\implies \frac{\partial \Pi}{\partial x_i} = \sum_{j=1}^{m} p_j \frac{\partial Q_j}{\partial x_i} - \frac{\partial C}{\partial x_i} = 0$$

$$\implies \mathbf{p} \cdot \frac{\partial \mathbf{Q}}{\partial x_i} - \frac{\partial C}{\partial x_i} = 0.$$

Applications to economics & Cobb-Douglas function

Situation $\#2$ (discriminating monopolist):

- In some cases the price per product p_i is not constant but can be influenced by the demand.
- E.g. if a monopolist floods the market the price may be negatively influence by the production, so we get for example $p_i = a_i - b_i Q_i$ with constants a_i, b_i.
- Or we get a situation where the demand, given by the demand function $F_i(p_i)$ is influenced by the price, which then influences the production.
- Thus $Q_i = F_i(p_i)$ –output of product x_i is influenced by the price p_i.
- Usually we can then invert, with inverse function $G_i(Q_i)$, determining the price:

$$Q_i = F_i(p_i) \Longleftrightarrow p_i = G_i(Q_i)$$

and we can try and determine optimal output. (i.e. find maximum of profit as function of the productions Q_i).

Cobb-Douglas production function

- The Cobb-Douglas production function is used in many economic models and involves two (or more) variables.
- For instance in models where production Q of a product depends on: $x_1 = K =$ Capital input and $x_2 = L =$ Labour input, so $Q = Q(K, L) = Q(x_1, x_2)$.
- The form of the Cobb-Douglas (CD, 1928) function in two variables is:

$$Q(x_1, x_2) = x_1^a x_2^b = K^a L^b, \quad \text{with constants} \quad a, b > 0.$$

- In more variables, we would have:

$$Q(x_1, x_2, \ldots, x_n) = x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}, \quad \text{with} \quad a_1, a_2, \ldots, a_n > 0.$$

Cobb-Douglas production function

- Consider the two-variables case. Assume the price of the product is p and the cost function

$$C(x_1, x_2) = w_1x_1 + w_2x_2 \qquad \text{(constants } w_1, w_2 > 0).$$

- Hence revenue $R(x_1, x_2) = pQ(x_1, x_2) = px_1^a x_2^b$ and the profit function Π becomes

$$\Pi(x_1, x_2) = R(x_1, x_2) - C(x_1, x_2) = px_1^a x_2^b - w_1x_1 - w_2x_2.$$

- Let us find the critical points of Π, to investigate the strategy to maximise profit. FOCs:

$$\begin{aligned}\frac{\partial \Pi}{\partial x_1} &= apx_1^{a-1}x_2^b - w_1 = 0 \Longrightarrow w_1 = apQ/x_1\\ \frac{\partial \Pi}{\partial x_2} &= bpx_1^a x_2^{b-1} - w_2 = 0 \Longrightarrow w_2 = bpQ/x_2.\end{aligned}$$

- Denote the critical point by $(\cdot)^*$: $x_1^* = apQ^*/w_1, x_2^* = bpQ^*/w_2$.

Cobb-Douglas production function

- Furthermore, we have that $Q^* = (x_1^*)^a (x_2^*)^b$ and hence we can find (x_1^*, x_2^*) by reinserting

$$Q^* = \left(\frac{apQ^*}{w_1}\right)^a \left(\frac{bpQ^*}{w_2}\right)^b \Longrightarrow 1 = \left(\frac{ap}{w_1}\right)^a \left(\frac{bp}{w_2}\right)^b (Q^*)^{a+b-1}.$$

- The critical production Q^* subsequently follows and then reinsert back into the formulas for $x_1^* = apQ^*/w_1, x_2^* = bpQ^*/w_2$ to obtain the critical values. But: $a + b < 1$!
- We are, however, interested in the conditions on a, b such that this critical point is a maximum.
- Hence, we need to examine the Hessian at the critical point (Why?).

Cobb-Douglas production function

- Second-order conditions require the Hessian:

$$H = \begin{pmatrix} \Pi_{x_1x_1} & \Pi_{x_1x_2} \\ \Pi_{x_1x_2} & \Pi_{x_2x_2} \end{pmatrix} = \begin{pmatrix} a(a-1)px_1^{a-2}x_2^b & abpx_1^{a-1}x_2^{b-1} \\ abpx_1^{a-1}x_2^{b-1} & b(b-1)px_1^a x_2^{b-2} \end{pmatrix}.$$

- Conditions for maximum (ND ...):

$$\begin{aligned} LPM_1 =& a(a-1)px_1^{a-2}x_2^b < 0 \\ LPM_2 =& ab(a-1)(b-1)p^2x_2^{2a-2}x_2^{2b-2} - a^2b^2p^2x_1^{2a-2}x_2^{2b-2} > 0 \\ \Longrightarrow LPM_2 =& abp^2x_1^{2a-2}x_2^{2b-2}\left((a-1)(b-1) - ab\right) \\ =& abp^2x_1^{2a-2}x_2^{2b-2}(1-a-b) > 0. \end{aligned}$$

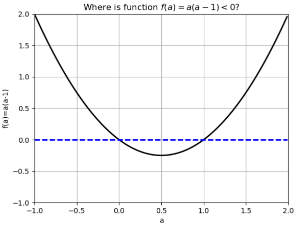
- Since all variables are positive, we have a maximum when (draw a graph of $a(a-1)$ versus a)

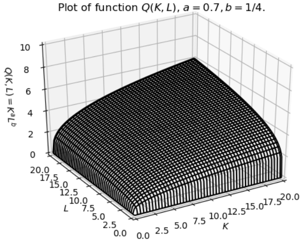
$$a(a-1) < 0, \qquad a+b < 1.$$

- So, we have $0 < a < 1$ and $0 < b < 1$ such that $a+b < 1$.

Cobb-Douglas production function

- If these conditions are not satisfied, then profit increases as x_1, x_2 get large.
- Felipe and Adams (2005) reanalysed the data from Cobb and Douglas (1928) and found a surprising result (links via wiki on CD).
- See exercises! Caution: bespoke production functions required based on appropriate/approximate fitting to data.





Cobb-Douglas production function: references

- Cobb and Douglas 1928: A Theory of Production. American Economic Review **18**.
- Felipe and Adams 2005: The estimation of the Cobb-Douglas function: a retrospective view. Eastern Economic Journal **31**, 427-445.
- ...

Ch. 3: Constrained optimisation

- In many situation in economics we deal with optimisation where variables $x_1, \dots, x_n$ are not varying freely but are subjected to restrictions, i.e. <u>constraints</u>.
- These could be constraints on their values (such as $x_i \geq 0$) or a relationship between variables (such as $x_1 + x_2 = 1$).
- There are two possible strategies for the latter:
  - eliminating some of the variables and get an optimisation problem in fewer variables or
  - trying to deal with the constraints as such.
- We will set up the mathematical framework for constraints.
- Thus, we will treat the problem of maximising/minimising functions $f(\mathbf{x}) = f(x_1, \dots, x_n)$ of variables $x_1, \dots, x_n$ subject to constraints of the following forms.

Constrained optimisation

...subject to constraints of the following forms:

- equality constraints, i.e. an additional set of conditions involving constraint function $h_i(\mathbf{x})$, $i = 1, \ldots, m$, where we set

$$h_1(\mathbf{x}) = c_1, \quad h_2(\mathbf{x}) = c_2, \quad \ldots, \quad h_m(\mathbf{x}) = c_m,$$

  for some numbers $c_1, \ldots, c_m$, involving $=$ symbols.
- inequality constraints, i.e., set of additional functions $g_i(\mathbf{x})$ which are bounded by some constants b_i:

$$g_1(\mathbf{x}) \leq b_1, \quad g_2(\mathbf{x}) \leq b_2, \quad \ldots, \quad g_k(\mathbf{x}) \leq b_k,$$

  for some numbers $b_1, \ldots, b_k$ (if $g_i(\mathbf{x}) \geq 0$ then we rewrite it to $-g_i(\mathbf{x}) \leq 0$), involving $\leq$ symbols.
- mixed constraints, i.e. combinations of the above situations.

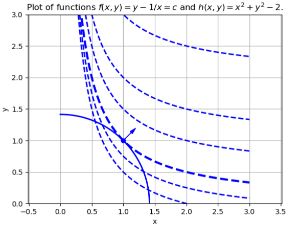
Example in economics with constraints

. . .

3.1 Single equality constraint ($n = 2, 3$)

Let us consider the problem of maximising/minimising a function $f(\mathbf{x})$ subject to a single constraint $h(\mathbf{x}) = c$. 2D case ($n = 2$): function $f(x, y)$ subject to $h(x, y) = c$.

- $h(x, y) = c \Longrightarrow$ curve in xy–plane.



- Consider now level curves $f(x, y) = cst = f_i$ for fixed values f_i. Let m be the value when the critical point is achieved $m = f_*$.

- As f scans through various values the critical point is obtained when curve $h = c$ touches at the critical level curve of f: $f = m$.
- At that point, since ∇f perpendicular to curve $f = m$ and ∇h perpendicular to the curve $h = c$, they must be parallel:

$$\nabla f = \lambda \nabla h$$

  for some scalar λ (note that λ can depend on the point).
- Argument holds equally for local minimum/maximum.
- Alternatively, at the max (min) value m of f, we have

$$df = 0 \iff f_x = f_y = 0$$

  and, hence, on level curve $f = m$, we have

$$\frac{dy}{dx} = \frac{f_x}{f_y}.$$

- Since the critical point also lies on the constraint curve $h = c$, we have

$$\implies \frac{dy}{dx} = \frac{f_x}{f_y} = \frac{h_x}{h_y}$$
$$\implies \frac{f_x}{f_y} = \frac{h_x}{h_y} = \lambda \quad \implies \quad \nabla f - \lambda \nabla h = 0.$$

- Now reformulate this as follows

$$f_x - \lambda h_x = 0 \quad f_y - \lambda h_y = 0$$

  because of constraint $h - c = 0$.

- Hence, introduce a new function, called the <u>Lagrangian</u>

$$L(x, y, \lambda) = f(x, y) - \lambda\,(h(x, y) - c)$$

- So critical points, subject to <u>constraint $h - c = 0$</u>, are obtained by considering the FOCs of this Lagrangian function $L(x, y, \lambda)$.

- Furthermore

$$\frac{\partial L(x,y,\lambda)}{\partial \lambda} = -\left(h-c\right) = 0 \Longleftrightarrow h = c.$$

- So either we consider the critical point (x^*, y^*) subject to the constraint $h(x,y) - c = 0$, obtained by considering the FOCs of L, or we can treat $L(x,y,\lambda)$ as a function of three independent variables: x, y, λ.
- Here λ is called the Lagrange multiplier, the FOCs of which reproduce the relevant critical point conditions for f as well as the constraint.
- The FOCs of L reproduce the relevant critical-point conditions plus the constraint:

$$L_x = 0, \quad L_y = 0, \quad L_\lambda = -(h-c) = 0.$$

- ...alternative (cf. Lanczos).

Single equality constraint (n variables)

- More generally, in case of n variables: $f(x_1, \ldots, x_n)$ subject to the constraint $h(x_1, \ldots, x_n) = c$ the critical points of f, subject to this contriant, are obtained from a FOCs of a Lagrangian

$$L(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - \lambda \left(h(x_1, \ldots, x_n) - c\right)$$

  is a function of $n+1$ variables.

- This leads to FOCs:

$$\left\{ \begin{array}{lr} \frac{\partial L(x_1, \ldots, x_n, \lambda)}{\partial x_i} = 0 & i = 1, \ldots, n \\ \frac{\partial L(x_1, \ldots, x_n, \lambda)}{\partial \lambda} = -\left(h - c\right) = 0 & \end{array} \right. , \tag{1}$$

  giving $n+1$ equations for $n+1$ unknowns $x_1, \ldots, x_n, \lambda$.

Non-degenerate constraint quantification

Remark:

- It is important to note that because of the requirement

$$\frac{f_{x_1}}{h_{x_1}} = \frac{f_{x_2}}{h_{x_2}} = \cdots = \frac{f_{x_n}}{h_{x_n}} = \lambda,$$

- we must demand that $\nabla h \neq \mathbf{0}$ at the critical point $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$, otherwise the argument used is no longer valid.
- This condition $\nabla h \neq \mathbf{0}$ is called the NDCQ: "non-degenerate constraint quantification".
- Examples ....

3.2 Several equality constraints (n–variables, m constraints)

- If we want to find the stationary points of $f(x_1, \dots, x_n)$ subject to m constraints simultaneously

$$h_i(x_1, \dots, x_n) = c_i \quad \text{for} \quad i = 1, \dots, m \quad \text{with } m < n$$

  with various values $c_1, \dots, c_m$ then we construct the Lagrangian

$$L(x_1, \dots, x_n; \lambda_1, \dots, \lambda_m) = f(x_1, \dots, x_n) - \lambda_1(h_1 - c_1) + \dots + \lambda_m(h_m - c_m)$$

  and m Lagrange multiplier $\lambda_1, \dots, \lambda_m$.
- We get $n + m$ FOCs

$$\frac{\partial L}{\partial x_1} = 0, \quad \dots \quad \frac{\partial L}{\partial x_m} = 0; \frac{\partial L}{\partial \lambda_1} = 0, \quad \dots \quad \frac{\partial L}{\partial \lambda_m} = 0,$$

  for $n + m$ variables $x_1, \dots, x_n; \lambda_1, \dots, \lambda_m$.

Several equality constraints

- Furthermore, in this case the NDCQ involves the gradients $\nabla h_1, \nabla h_2, \ldots, \nabla h_m$ of all constraints;
- the condition being that the $n \times m$–matrix (Jacobian derivative matrix)

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_n}{\partial x_1} & \cdots & \frac{\partial h_n}{\partial x_n} \end{pmatrix}$$

  has rank m, i.e. m gradients $\nabla h_1, \ldots, \nabla h_m$ are linearly independent.
- That is, none of the gradients can be expressed as a linear combination of the others at the critical point.
- Examples ....

Ch.3.3 Second-order conditions for problems with equality constraints

Characterisation of critical points arising from functions subject to constraints:

- Suppose we have found the critical points $\mathbf{x}^*$ of the function $f(\mathbf{x})$ subject to system of constraints

$$h_1(\mathbf{x}) = c_1, \dots, h_m(\mathbf{x}) = c_m \Longrightarrow$$

  vector of constraints $\mathbf{h} = (h_1, \dots, h_m)$ with vector of Lagrange multipliers $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$.
- Lagrangian:

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) =& f(\mathbf{x}) - \boldsymbol{\lambda} \cdot (\mathbf{h}(\mathbf{x}) - \mathbf{c}) \\ =& f(x_1, \dots, x_n) - \sum_{j=1}^{m} \lambda_j \left(h_j(x_1, \dots, x_n) - c_j \right). \end{aligned}$$

- Taylor expansion of Lagrangian L around the critical point $\mathbf{x}^*$ with values $\boldsymbol{\lambda}^*$ of Lagrange multipliers

$$L(\mathbf{x}+\delta\mathbf{x}) = L(\mathbf{x}) + d\mathbf{x}\cdot\nabla(f-\boldsymbol{\lambda}\cdot\mathbf{h})\Big|_{\substack{\mathbf{x}=\mathbf{x}^*\\ \boldsymbol{\lambda}=\boldsymbol{\lambda}^*}} + \frac{1}{2}d\mathbf{x}^T\cdot H d\mathbf{x} + \ldots,$$

where the Hessian is

$$\begin{pmatrix} L_{x_1x_1} & \ldots & L_{x_1x_n} \\ \vdots & & \vdots \\ L_{x_nx_1} & \ldots & L_{x_nx_n} \end{pmatrix} = \begin{pmatrix} f_{x_1x_1}-\lambda h_{x_1x_1} & . & f_{x_1x_n}-\lambda h_{x_1x_n} \\ \vdots & & \vdots \\ L_{x_nx_1}-\lambda h_{x_nx_1} & . & f_{x_nx_n}-\lambda h_{x_nx_n} \end{pmatrix}$$

- and at the stationary point $\mathbf{x}^*$ we have $\nabla(f-\boldsymbol{\lambda}\cdot\mathbf{h})\Big|_{\substack{\mathbf{x}=\mathbf{x}^*\\ \boldsymbol{\lambda}=\boldsymbol{\lambda}^*}} = 0$.
- Thus for a minimum we need $d\mathbf{x}^T\cdot H d\mathbf{x} > 0$ (PD)
- and for a maximum we need $d\mathbf{x}^T\cdot H\mathbf{x} < 0$ (ND).

- The key problem now is that $d\mathbf{x}$ in the Taylor expansion is not an arbitrary variation of $x_1, \ldots, x_n$, but $\delta\mathbf{x}$ is subject to the constraints $h_1(\mathbf{x}) = c_1, \ldots, h_m(\mathbf{x}) = c_m$!
- I.e. $d\mathbf{x}$ lies along the constrained surface, along the intersection of the tangent surfaces of $h_1(\mathbf{x}) = c_1, \ldots, h_m(\mathbf{x}) = c_m$ at $\mathbf{x}^*$!
- So $d\mathbf{x}$ is constrained, for all i, by:

$$\begin{aligned} h_i(\mathbf{x}^* + d\mathbf{x}) =& h_i(\mathbf{x}^*) + d\mathbf{x} \cdot \nabla h_i|_{\mathbf{x}=\mathbf{x}^*} + \cdots = c_i \text{ cst} \\ \Longleftrightarrow & d\mathbf{x} \cdot \nabla h_i|_{\mathbf{x}=\mathbf{x}^*} = 0 \Longrightarrow d\mathbf{x} \perp \nabla h_i|_{\mathbf{x}=\mathbf{x}^*}, \end{aligned}$$

- so $d\mathbf{x}$ must lie in tangent plane to all $h_i = c_i$.

- Conclusion:
  The character of critical point $\mathbf{x}^*$ is obtained by deciding whether the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T \cdot H\mathbf{x}$ is PD (or ND) subject to the constraints

$$\mathbf{x} \cdot \mathbf{u}^i, \quad \text{where} \quad \mathbf{u}^{(i)} = \nabla h_i(\mathbf{x}), \quad i = 1, \ldots, m.$$

- This requires a theory of constrained quadratic forms.
- Example: $\ldots Q(x, y) = x^2 - y^2$ (ID) with $\mathbf{u} = (0, 1)$ then $\mathbf{u} \cdot \mathbf{x} = \mathbf{0} \Longrightarrow y = 0$; constrained one PD (in x).
- Note: if the unconstrained quadratic form is PD then the constrained one would also be PD. But, conversely, the constrained Q being PD does not imply that the unconstrained Q is PD!

Constrained quadratic forms and bordered Hessians

- We now want to set up criteria to decide whether quadratic froms $Q(\mathbf{x})$ subject to a set of linear constraints $\mathbf{u} \cdot \mathbf{x} = 0, i = 1, \ldots, m$ is PD or ND (or otherwise).
- General two-variable case, subject to one constraint. Consider

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

  subject to constraint $ux + vy = 0$, i.e. $\mathbf{u} = (u, v)$.
- If $v \neq 0 \quad \Longrightarrow \quad y = -ux/v \quad \Longrightarrow$

$$Q = ax^2 - 2b\frac{u}{v}x^2 + c\left(-\frac{ux}{v}\right)^2 = \left(a - 2b\frac{u}{v} + c\frac{u^2}{v^2}\right)x^2$$

$$\Longrightarrow Q = (av^2 - 2buv + cu^2)\left(\frac{x}{v}\right)^2.$$

- Q is **PD** if $av^2 - 2buv + cu^2 > 0$ and **ND** if $av^2 - 2buv + cu^2 < 0$.
- Note that these expressions look similar to the expression for Q apart from the minus sign!
- Introducing the 3×3 matrix, called "bordered Hessian", $(n+m) \times (n+m)$ matrix with $n = 2, m = 1$ (number of variables & constraints)

$$H_B = \begin{pmatrix} 0 & u & v \\ u & a & b \\ v & b & c \end{pmatrix}$$

- then $\det(H_B)$ determines the sign (and note the sign reversal)

$$\begin{aligned} \det(H_B) &= -u(uc - bv) + v(ub - av) \\ &= -(av^2 - 2buv + cu^2) \\ \det(H_B) < 0 &\iff Q \quad \text{PD} \\ \det(H_B) > 0 &\iff Q \quad \text{ND}. \end{aligned}$$

General form of bordered Hessians

- Consider a quadratic form $Q(\mathbf{x}) = \mathbf{x}^T H \mathbf{x}$ with n variables $\mathbf{x}$ and subject to m constraints $\mathbf{h}$, i.e.

$$h_i = \mathbf{u}_i \cdot \mathbf{x} = 0, \quad i = 1, \ldots, m$$

- then the corresponding bordered Hessian H_B determines the sign of the quadratic form:

$$H_B = \begin{pmatrix} 0 & & \mathbf{u}_1^T \\ 0 & & \ldots \\ 0 & & \mathbf{u}_m^T \\ \mathbf{u}_1 \ldots \mathbf{u}_m & \ldots & H \end{pmatrix} = \begin{pmatrix} 0 & B \\ B^T & H \end{pmatrix},$$

- where B is the $m \times n$ matrix (with $\mathbf{u}_i = (u_1^{(i)}, \ldots, u_m^{(i)})$)

$$B = \begin{pmatrix} u_1^{(1)} & \ldots & u_n^{(1)} \\ \vdots & & \vdots \\ u_1^{(m)} & \ldots & u_n^{(m)} \end{pmatrix}.$$

- In fact, the form becomes transparant when seen from the point of view of the Lagrangian

$$
\begin{aligned}
L(\mathbf{x}, \boldsymbol{\lambda}) =& f - \sum_{i=1}^{m} \lambda_i (h_i - c_i) \\
& \text{such that} \\
H_B =& \begin{pmatrix} (L_{\lambda_i \lambda_j})_{i,j=1,\ldots,m} & (L_{\lambda_i x_j})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \\ (L_{x_i \lambda_j})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} & (L_{x_i x_j})_{i,j=1,\ldots,n} \end{pmatrix} \\
=& \begin{pmatrix} 0 & -\partial h_i / \partial x_j \\ -\partial h_j / \partial x_i & L_{x_i x_j} \end{pmatrix}.
\end{aligned}
$$

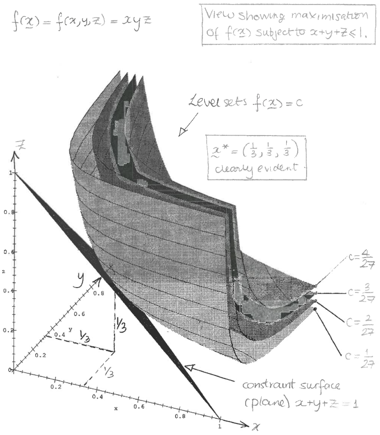
Principal minor test for bordered Hessians

- Classify quadratic forms $Q(\mathbf{x}) = \mathbf{x}^T \cdot H\mathbf{x}$ subject to constraints $\mathbf{u}_i \cdot \mathbf{x},\ i = 1, \ldots, m$.
- Construct the bordered Hessian

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

- We need $n - m$ LPM's starting with $LPM_{2m+1}, \ldots, LPM_{n+m}$ (note $n \geq m$), where $LPM_{n+m} = \det(H_B)$ then
- we have the following rules.

- We have the following rules. **(A)** Q is PD (under the constraints) if $sign(LPM_{n+m}) = sign(\det H_B) = (-1)^m$ and the successive $LPMs$ have the same sign.
  (B) Q is ND (under the constraints) if $sign(LPM_{n+m}) = sign(\det H_B) = (-1)^n$ and the successive $LPMs$ alternate in sign.
  (C) If both (A) and (B) are violated by nonzero $LPMs$ then Q is ID.
- We do not consider –here– the semi-definite conditions!



f(x) = f(x,y,z) = xyz
View showing maximisation of f(x) subject to x+y+z≤1.
Level sets f(x) = c
x* = (1/3, 1/3, 1/3) clearly evident
c = 4/27
c = 3/27
c = 2/27
c = 1/27
constraint surface (plane) x+y+z = 1
1/3
1/3
1/3
z
y
x