APPENDIX B: MAPPINGS OF VECTORS

B.1 VECTORS AND SETS

Key ideas: orthogonal vector, hyperplanes, convex sets, open and closed sets

We now extend our analysis to ordered *n*-tuples or "vectors." Each component of the vector $x = (x_1, ..., x_n)$ is a real number. Where it is important to make clear the dimension of the vector, we write $x \in \mathbb{R}^n$. If each component is positive, then *x* is positive and we write $x \in \mathbb{R}^n_+$.

We will describe the vector *y* as being larger than the vector *x* if every component of *y* is at least as large as *x* and write $y \geq x$. If at least one component of *y* is strictly larger we will say that *y* is strictly larger than *x* and write $y > x$. (We will occasionally use the notation $x \gg y$ to indicate that all components are strictly larger.)

A neighborhood of a vector is the set of points near it. Thus to define a neighborhood we need some measure of distance $d(y,z)$ between any two vectors $y = (y_1, ..., y_n)$ and $z = (z_1, ..., z_n)$. If $n = 1$, a natural measure is the absolute value of the difference so that $d(y,z) = |y - z|$. Similarly with $n = 2$, a natural measure is the Euclidean distance $\|y - z\|$ between the two vectors. Appealing to Pythagoras Theorem,

$$\|y - z\|^2 = (y_1 - z_1)^2 + (y_2 - z_2)^2 .$$

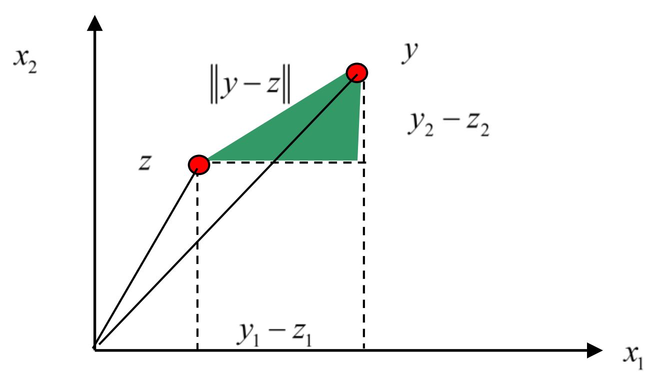


Figure B.1-1: Distance between 2 vectors

Extending this to *n*-dimensions, the square of Euclidean distance between ordered *n*-tuples is defined as follows:

$$\|y - z\|^2 = \sum_{j=1}^{n} (y_j - z_j)^2.$$

It is sometimes more convenient to express distance as a vector product.

Vector Product

The product (“inner product” or “sumproduct”) of two n-dimensional vectors y and z is.

$$y \cdot z = \sum_{i=1}^{n} y_i z_i.$$

From this definition, it follows that the distributive law for multiplying vectors is the same as for numbers.

Distributive law for vector multiplication

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

Orthogonal Vectors

Two vectors, x and p are depicted below in 3-dimensional space.

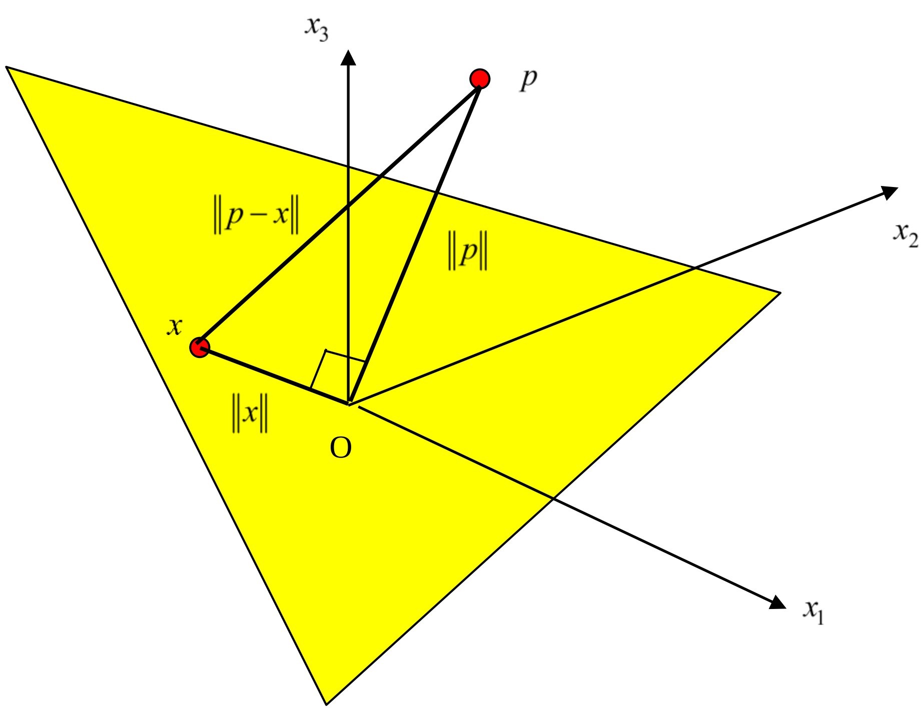


Figure B.1-2: Orthogonal Vectors

The length of each vector is its distance to the origin. If x and p are perpendicular it follows from Pythagoras Theorem that the square of the hypotenuse equals the sum of the squares of the other two sides, that is,

$$\|p-x\|^2 = \|p\|^2 + \|x\|^2 \quad . \qquad \text{(B.4-1)}$$

From the definition of a vector product, the square of Euclidean distance between two vectors can then be written as follows.[1]

$$\|p-x\|^2 = \sum_{i=1}^{n} (p_i - x_i)(p_i - x_i) = (p-x)\cdot(p-x) .$$

Substituting into (B.4-1),

$$(p-x)\cdot(p-x) = p\cdot p + x\cdot x \ .$$

Applying the rules of vector multiplication,

$$p\cdot p + 2p\cdot x + x\cdot x = p\cdot p + x\cdot x \ .$$

Hence

$$p\cdot x = 0 .$$

We generalize perpendicularity to higher dimensions as follows.

Definition: Orthogonal Vectors

The vectors x and p are orthogonal if their product $p\cdot x = \sum_{i=1}^{n} p_i x_i = 0$.

Consider Figure B.1-2 again. Let H be the set of vectors perpendicular to p. In mathematical terms, $H = \{x \mid p\cdot x = 0\}$. This is the plane created by extending the shaded

[1] The alternative distance measure $d(y,z) = (\sum_{j=1}^{n} (y_j - z_j)^{2i})^{\frac{1}{2i}}$, $i = 2,3...$ places more weight on larger differences. In the limit this is equivalent to placing all the weight on the maximum difference, that is, $d(y,z) = \underset{j=1,\dots,n}{Max} \left| y_j - z_j \right|$.

region. If *p* is two-dimensional, the set *H* is a line. If *p* is of dimension greater than three, the set *H* is called a hyperplane.

If we add the vector x^0 to each point in the hyperplane we shift the plane as depicted in Figure B.1-3. Note that the vector *x* is in the hyperplane through x^0, orthogonal to *p* if the vectors p and $x - x^0$ are orthogonal. Thus the hyperplane through x^0 is the set of vectors

$$H^0 = \{x \mid p \cdot (x - x^0) = 0\}.$$

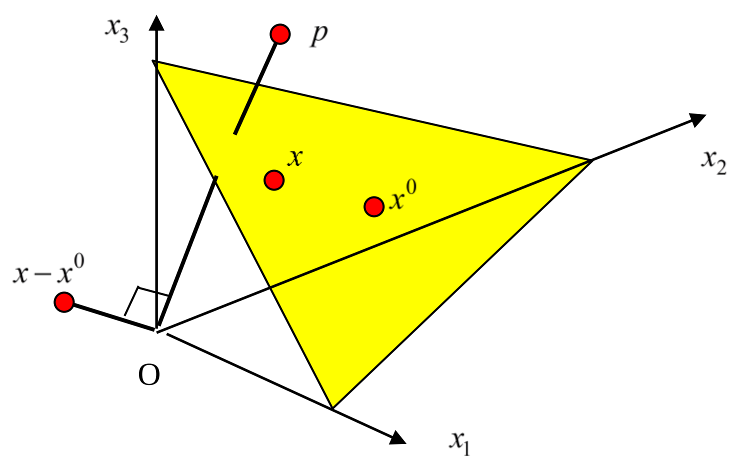


Figure B.1-3: Hyperplane through x^0

Hyperplanes have a central role in price theory. Take for example the expenditure of a consumer purchasing a consumption bundle *x* given a price vector *p*. The set of consumption bundles costing the same as the bundle x^0 is the set $H = \{x \mid p \cdot x = p \cdot x^0\}$, that is, the hyperplane through x^0, orthogonal to *p*.

Linear and Convex Combinations of Vectors

Let $x^0, ..., x^{m-1}$ be *m* vectors in $\mathbb{R}^n$. Then $y \in \mathbb{R}^n$ is a linear combination of these *m* vectors if for some $(\alpha_0, ..., \alpha_{m-1})$,

$$y = \sum_{i=0}^{m-1} \alpha_i x^i .$$

If $m < n$, the set of all possible linear combinations of *m* vectors in $\mathbb{R}^n$ is a space of lower dimension. To illustrate, consider the two vectors depicted below in $\mathbb{R}^3$. Taking all the linear combinations of these vectors a 2-dimensional plane is mapped out.

A convex combination of two vectors is a weighted average of the vectors where the weights are both positive and sum to one.

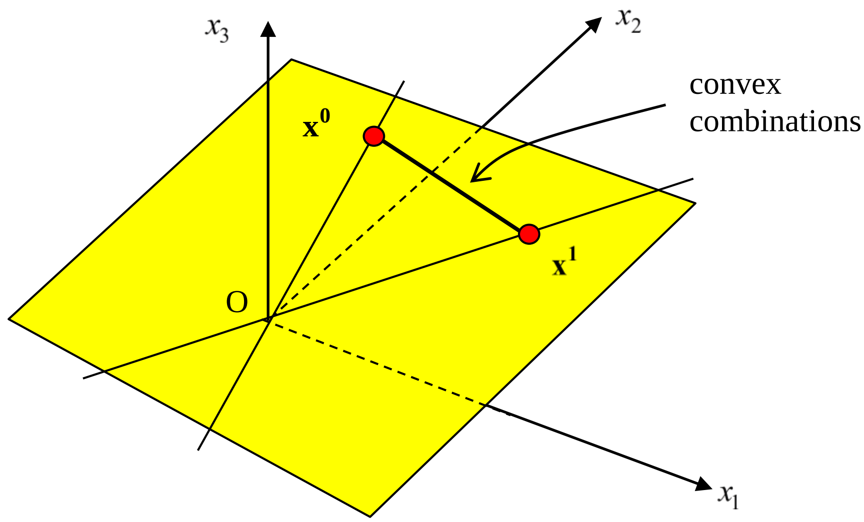


Figure B.1-4: Linear and convex combinations

Convex combination of two vectors

The vector *y* is a convex combination of x^0 and $x^1 \in \mathbb{R}^n$ if

$$y = (1-\lambda)x^0 + \lambda x^1, \;\; 0 \le \lambda \le 1.$$

It is often useful to rewrite a convex combination as follows:

$$y = x^0 + \lambda(x^1 - x^0), \;\; \lambda \in [0,1].$$

Thus the vector *y* is the vector x^0 plus some fraction of the vector $x^1 - x^0$. The weight λ thus indicates the distance along the line joining x^0 and x^1.

To further emphasize this connection, we will often write the convex combination of x^0 and x^1 as x^λ. As λ approaches zero, x^λ approaches x^0 and as λ approaches 1, x^λ approaches x^1.

Convex Sets of Vectors

A set S in $\mathbb{R}^n$ is any collection of vectors. In one dimension we have seen that two especially useful sets are the closed and open intervals. Both of these sets have the property that if points x^0 and x^1 are in the set then so is any convex combination,

$$x^\lambda = (1-\lambda)x^0 + \lambda x^1, \quad 0 \le \lambda \le 1.$$

This suggests the following generalization of an interval.

Convex Set

$X \subset \mathbb{R}^n$ is convex if, for any $x^0, x^1 \in X$, every convex combination $x^\lambda \in X$.

Graphically, if two points are in the set X then every point on the line joining these two points also lies in X. In two dimensions, the boundary of the set must have the "bowed out" or convex shape as depicted below.

A set is strictly convex if it is convex and, for every pair of vectors x^0 and $x^1 \in X$, no convex combination lies on the boundary of the set.

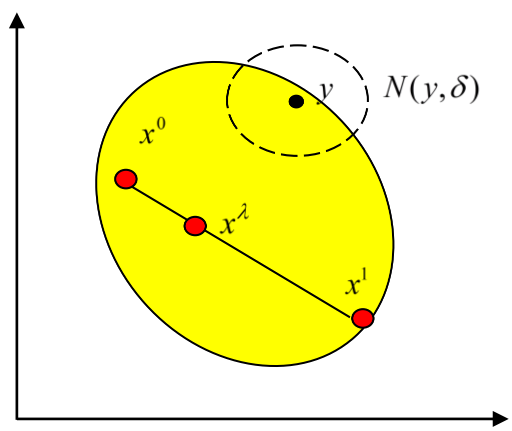


Figure B.1-5: Neighborhood and Boundary Point

To formalize this we need to make precise the idea of a boundary point. First we extend our definition of a neighborhood. A "$\delta-$neighborhood" of a vector *y* is the set of points whose distance from y is no greater than δ, that is,

$$N(y,\delta)=\{x\in\mathbb{R}^n,\ \|x-y\|\le\delta\}.$$

Boundary Point of a Set

The vector *y* is a boundary point of the set *X*, if every $\delta-$neighborhood of *y* contains points in *X* and points not in *X*.

Any point is a set that is not a boundary point is called an interior point. That is, *y* is an interior point of *X* if it has a $\delta-$neighborhood that lies in *X*.

Open and Closed sets

A subset of $\mathbb{R}^n$ is closed if it includes all its boundary points. A subset is open if it does not include any of its boundary points.

Compact set

A set is compact if it is both closed and bounded.

We have the following generalization of the Extreme Value Theorem.

Extreme Value Theorem ($\mathbb{R}^n$)

If f is continuous in some compact subset *S* of $\mathbb{R}^n$ then there are points x^0 and x^1 in *S* where *f* takes on its maximum and minimum values.

For example, the set $B=\{x\in\mathbb{R}^n_+ \mid p\cdot x\le I,\ I>0\}$ is closed. If $p\gg 0$, we can interpret this set as a budget set. Note that x_j is bounded from below by 0 and from above by I/p_j. Thus *B* is also bounded. Let *f* be a utility function defined on *B*. Then for the extreme value theorem, *f* takes on its maximum at some point x^0 in the budget set.

Exercise B.1-1: Convex Sets

Suppose the sets $X_1, X_2 \subset \mathbb{R}^n$ are both convex.

(a) Show that the intersection of the two sets is convex.

(b) Show by counter example that the union of the two sets need not be convex.

(c) Show that $Y = X_1 + X_2$ and $Z = X_1 - X_2$ are both convex.

B.2 FUNCTIONS OF VECTORS

Key Ideas: partial and total derivatives, functions of vectors, contour sets, concave and quasi-concave functions

A function f maps each vector $x = (x_1, ..., x_n)$ to a point on the real line. Typically we assume that the domain of the function, D_f is convex. For example, a production function $q = f(z)$ maps an input vector z into output q and a utility function $u = U(x)$ maps a consumption vector into consumer utility. In each case a natural domain of the function is the convex set of non-negative vectors $\mathbb{R}^n_+$.

Partial and Total Derivatives

For a function of *n* variables, we can fix all but one variable and consider the derivative with respect to this one variable. This is known as a partial derivative

$$\frac{\partial f}{\partial x_j} = \lim_{h\to 0} \frac{f(x_1,...x_{j-1}, x_j + h, x_{j+1},...x_n) - f(x)}{h} \ .$$

Higher-order partial derivatives are similarly defined.[2]

Definition: Gradient Vector

A function *f* with domain $D_f \subset \mathbb{R}^n$ is differentiable at *x* if it is partially differentiable with respect to x_i, $i = 1,...,n$. The vector of partial derivatives

$$\frac{\partial f}{\partial x}(x) \equiv (\frac{\partial f}{\partial x_1}(x),...,\frac{\partial f}{\partial x_n}(x))$$

is the gradient of *f* at *x*.

[2] While we will usually write the partial derivatives of $f(x, y)$ as $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, we will sometimes find it more helpful to write them as $f_x(x, y)$ and $f_y(x, y)$.

Next suppose all the components of x vary with some other variable α, that is, $x = x(\alpha)$.

This is depicted below for the two variable case.

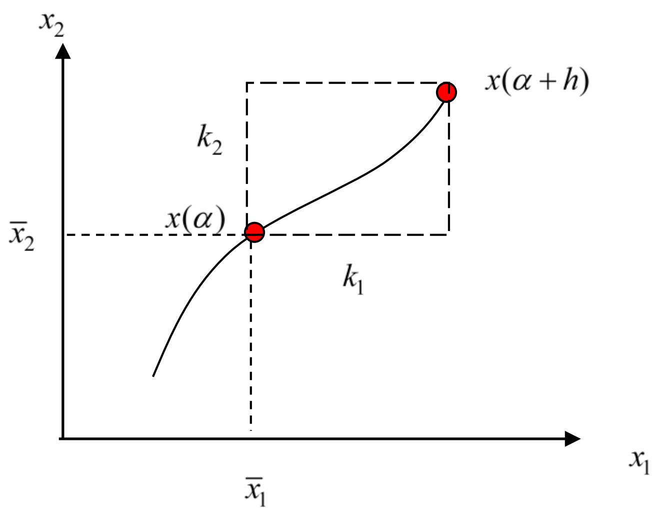


Figure B.2-1: The vector x changes with α

Following the definition of a derivative we seek to evaluate the limit of the following ratio

$$\frac{\Delta f}{\Delta \alpha} = \frac{f(x(\alpha+h)) - f(x(\alpha))}{h}.$$

Define $\bar{x} = x(\alpha)$, $k_1 = x_1(\alpha+h) - x_1(\alpha)$ and $k_2 = x_2(\alpha+h) - x_2(\alpha)$.

Then we can rewrite the above ratio as follows:

$$\frac{\Delta f}{\Delta \alpha} = \frac{f(\bar{x}_1 + k_1, \bar{x}_2 + k_2) - f(\bar{x}_1, \bar{x}_2)}{h}$$

$$= \frac{f(\bar{x}_1 + k_1, \bar{x}_2) - f(\bar{x}_1, \bar{x}_2) + f(\bar{x}_1 + k_1, \bar{x}_2 + k_2) - f(\bar{x}_1 + k_1, \bar{x}_2)}{h}$$

$$= [\frac{f(\bar{x}_1 + k_1, \bar{x}_2) - f(\bar{x}_1, \bar{x}_2)}{k_1}][\frac{x_1(\alpha+h) - x_1(\alpha)}{h}]$$

$$+ [\frac{f(\bar{x}_1 + k_1, \bar{x}_2 + k_2) - f(\bar{x}_1 + k_1, \bar{x}_2)}{k_2}][\frac{x_2(\alpha+h) - x_2(\alpha)}{h}].$$

Each of the four ratios becomes a partial derivative as $h \to 0$. Taking the limit,

$$\frac{df}{d\alpha} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial \alpha} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial \alpha}.$$

The argument generalizes immediately when there are more than two variables. Expressing the result in vector notation, the total derivative is

$$\frac{df}{d\alpha} = \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial \alpha} \; .$$

Rather than make explicit the underlying variable leading to the change in *x*, this is often left implicit as follows:

$$df = \frac{\partial f}{\partial x}\cdot dx \; .$$

This is called the total differential of *f*.

Special functions

The following two special mappings from $\mathbb{R}^n$ onto $\mathbb{R}$ have a very central role in the theory of optimization.

Linear Function[3]

$$f(x) = a_0 + \sum_{i=1}^{n} a_j x_j = a_0 + a \cdot x$$

For the two-variable case, the graph of the function $y = a_0 + a_1 x_1 + a_2 x_2$ is a plane in 3-dimensional space. An example is depicted below.

[3] Mathematicians typically reserve the term "linear function" for functions where each term is linear in one of the variables. However economists typically include the constant.

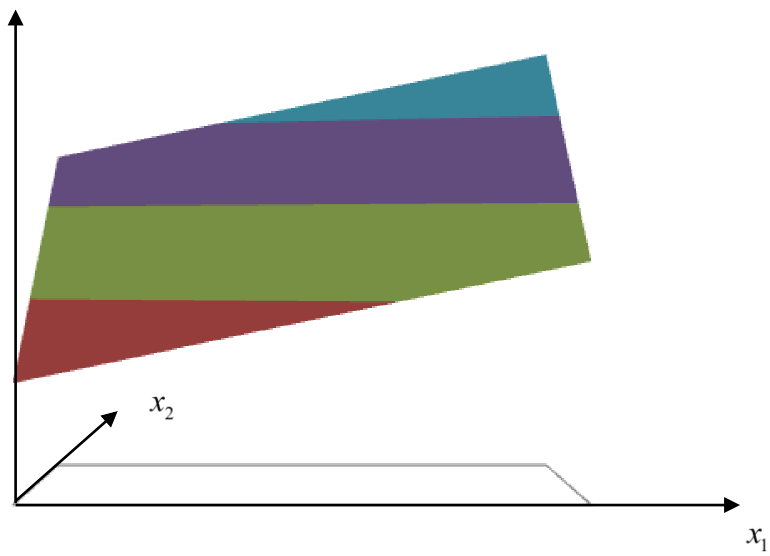


Figure B.2-2: Linear function

Quadratic Function

$$f(x) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Note that a quadratic function *f* is the sum of a linear function and the function

$$q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j .$$

This function is called a quadratic form. For every pair x_i and x_j, where $j \neq i$, the quadratic form includes the two terms $a_{ij} x_i x_j + a_{ji} x_j x_i = (a_{ij} + a_{ji}) x_i x_j$. Thus there is no loss of generality in assuming that $a_{ij} = a_{ji}$.

For the two-variable case, we can therefore write the quadratic form as follows:

$$q(x) = \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} x_i x_j = a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_{22} x_2^2 .$$

The figures below depict some of the possible graphs of a quadratic form. Both the graphs in the top row depict functions, which have a minimum at $x=0$. On the bottom left we have a maximum and on the bottom right a saddle point.

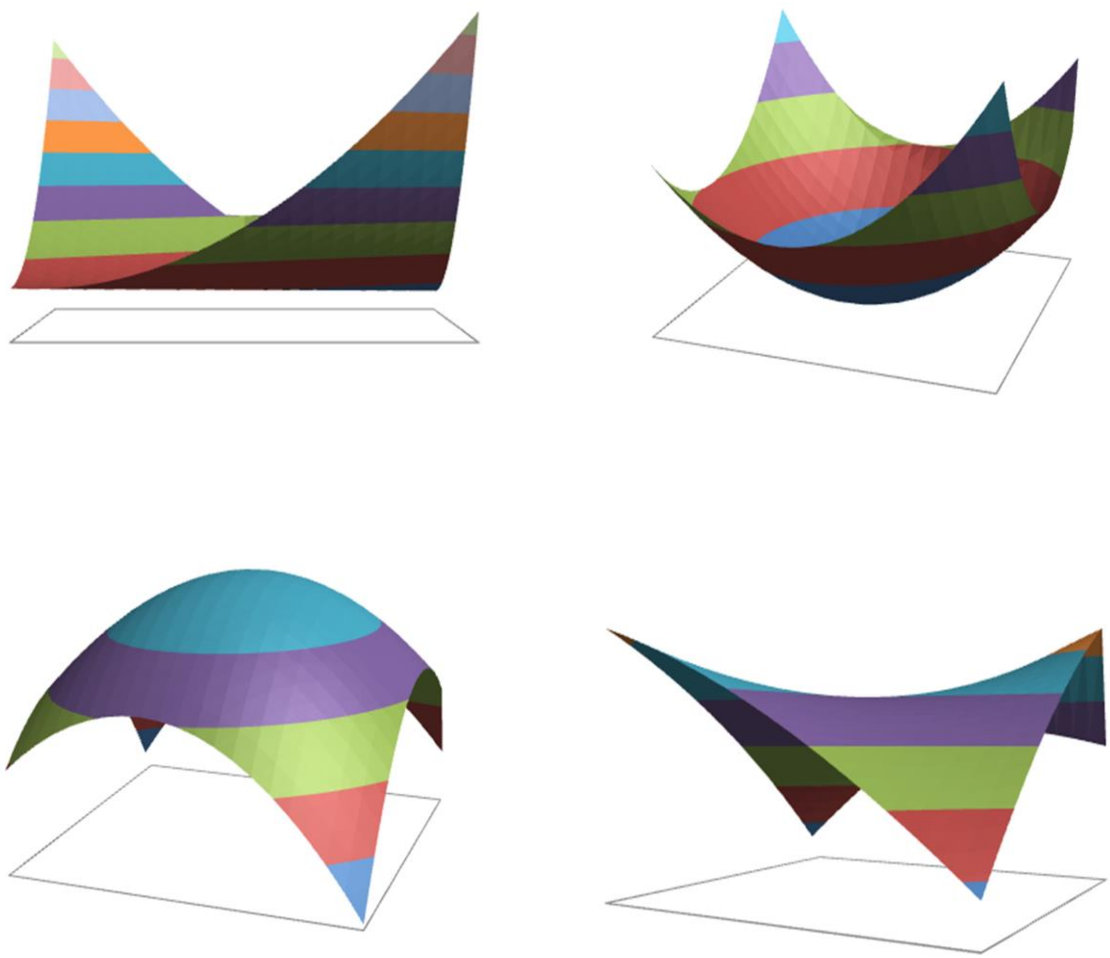


Figure B.2-3: Quadratic forms

Example: Quadratic form with a saddle point

Consider the quadratic form $q(x_1,x_2)=-x_1^2+4x_1x_2-x_2^2$. Then $q(x_1,0)=-x_1{}^2$ and $q(0,x_2)=-x_2{}^2$. Thus, taking one variable at a time, it appears that the quadratic form has a maximum at $x=0$. However, setting $x_1=x_2=z$ we have $q(z,z)=2z^2$.

Therefore the quadratic form has the saddle shape depicted in the bottom right of Figure B.2-3. ■

The properties of the quadratic form lie at the heart of the theory of optimization. The reason is that the graph of a function can be approximated by a quadratic function that takes on the same value and has the same first- and second-partial derivatives at x^0. For the two-variable case we have the following quadratic approximation

$$h(x) = f(x^0) + \sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(x^0)(x_i - x_i^0) + \tfrac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0)(x_i - x_i^0)(x_j - x_j^0) \; .$$

Note first that $h(x^0) = f(x^0)$. Differentiating by x_1,

$$\frac{\partial h}{\partial x_1}(x) = \frac{\partial f}{\partial x_1}(x^0) + \frac{\partial^2 f}{\partial x_1^{\,2}}(x^0)(x_1 - x_1^0) + \frac{\partial^2 f}{\partial x_1 \partial x_2}(x^0)(x_2 - x_2^0) \, .$$

Setting $x = x^0$, $\frac{\partial h}{\partial x_1}(x^0) = \frac{\partial f}{\partial x_1}(x^0)$. Note that $\frac{\partial h}{\partial x_1}$ is a linear function with partial derivatives of $\frac{\partial^2 f}{\partial x_1^{\,2}}(x^0)$ and $\frac{\partial^2 f}{\partial x_1 \partial x_2}(x^0)$. Thus the second-partial derivatives of *h* and *f* are also the same.

As a preliminary step we consider the two-variable case and seek conditions that ensure that a quadratic form is negative. If the quadratic form is negative for all *x*, (that is, $q(x) \le 0$), it is said to be negative semi-definite. If $q(x)$ is strictly negative for all $x \ne 0$ it is said to be negative definite.

Proposition B.2-1: Necessary and sufficient conditions for $q(x_1, x_2) \equiv \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij} x_i x_j$ **to be negative definite**[4]

If (i) $a_{11} < 0$ and (ii) $a_{11}a_{22} - a_{12}a_{21} > 0$ then $q(x_1, x_2) < 0$ for all $x \ne 0$.

[4] Note that if (i) and (ii) both hold, then $a_{22} < 0$ as well.

Proposition B.2-2: Necessary and sufficient conditions for $q(x_1,x_2)\equiv\sum_{i=1}^{2}\sum_{j=1}^{2}a_{ij}x_ix_j$ to be negative semi-definite

$q(x_1,x_2)\le 0$ if and only if (i) $a_{11},a_{22}\le 0$ and (ii) $a_{11}a_{22}-a_{12}a_{21}\ge 0$.

We will derive Proposition B.2-2[5]. Setting $x_2=0$ in

$$q(x_1,x_2)\equiv\sum_{i=1}^{2}\sum_{j=1}^{2}a_{ij}x_ix_j, \qquad \text{(B.2-1)}$$

then

$$q(x_1,x_2)=a_{11}x_1^2. \qquad \text{(B.2-2)}$$

Thus for $q(x_1,x_2)$ to be negative, a necessary condition is that $a_{11}\le 0$. A symmetric argument establishes that it is also necessary that $a_{22}\le 0$.

Suppose first that $a_{11}=0$. Then, for any $x_1\ne 0$,

$$q(x_1,x_2)=2a_{12}x_1x_2+a_{22}x_2{}^2=(2a_{12}+a_{22}(\frac{x_2}{x_1}))x_1x_2.$$

If $a_{12}>0$ we can choose x_2/x_1 sufficiently close to zero that the term in parentheses is strictly positive. If $x_1x_2>0$ then $q(x_1,x_2)>0$ so $q(x)$ is not negative semi-definite. An identical argument establishes that it cannot be the case that $a_{12}<0$ either. Then a necessary condition for $q(x)$ to be negative semi-definite is $a_{12}=0$. (Note that with $a_{11}=0$ the condition $a_{11}a_{22}-a_{12}a_{21}\ge 0$ is equivalent to the condition $a_{12}=0$.)

Next suppose that $a_{11}<0$. Completing the square, (B.2-1) can be rewritten as

$$q(x_1,x_2)=a_{11}(x_1+\frac{a_{12}}{a_{11}}x_2)^2+\frac{1}{a_{11}}(a_{11}a_{22}-a_{12}a_{12})x_2{}^2.$$

Actually, because $a_{12}=a_{21}$ we can write this more conveniently as

[5] The proof of Proposition B.2-1 is very similar.

$$q(x_1, x_2) = a_{11}(x_1 + \frac{a_{12}}{a_{11}})^2 + \frac{1}{a_{11}}(a_{11}a_{22} - a_{12}a_{21})x_2{}^2. \tag{B.2-3}$$

Suppose we choose x_1 so that the first term on the right-hand side of (B.2-3) is zero. Because $a_{11} < 0$, a necessary condition for the right-hand side to be negative is

$$a_{11}a_{22} - a_{12}a_{21} \geq 0. \tag{B.2-4}$$

We have therefore established the necessity of conditions (i) and (ii). Sufficiency follows almost immediately. If $a_{11} = 0$ it follows from (B.2-4) that $a_{12} = 0$ and hence that $q(x_1, x_2) = a_{22}x_2{}^2$. Appealing to (i) in the statement of the proposition, it follows that $q(x) \leq 0$. If $a_{11} < 0$ if follows directly from (B.2-3) that $q(x_1, x_2) \leq 0$.

Concave functions

In Chapter 1 a function was defined to be concave if, for any two points on the graph of *f*, the line joining these two points lies below the graph of the function. This definition generalizes naturally.

Definition: Concave function

The function f is concave on the convex set $X \subset \mathbb{R}^n$ if for any vectors x^0 and x^1 in *X*, and any convex combination $x^\lambda = (1-\lambda)x^0 + \lambda x^1,\ 0 < \lambda < 1$,

$$f(x^\lambda) \geq (1-\lambda) f(x^0) + \lambda f(x^1).$$

If the inequality is always strict, the function is strictly concave.

The two-variable case is depicted below. Starting from x^0, pick any other point x^1, then the curve must bend forward in the direction of x^1. In the second figure, the linear approximation of the function at x^0 has been added.

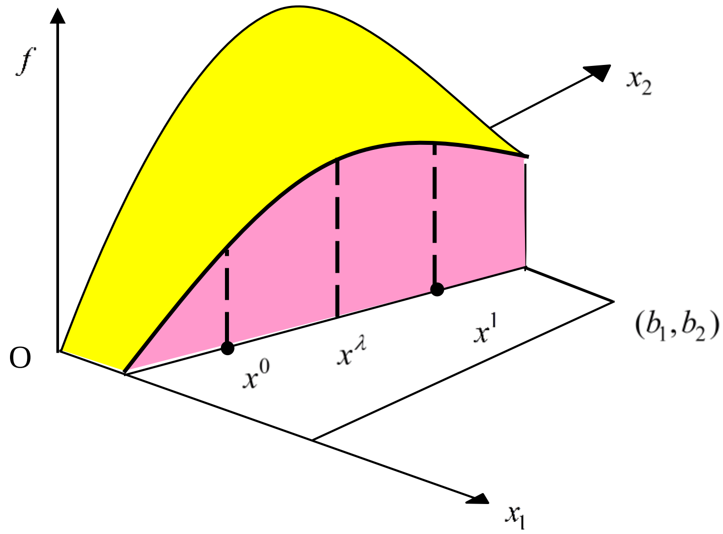


Figure B.2-4: Concave function

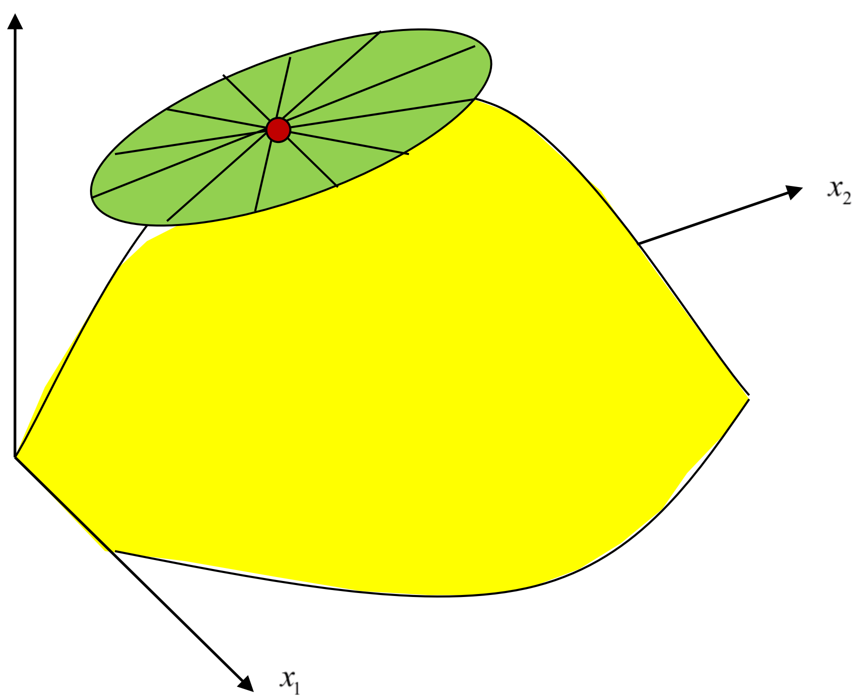


Fig B.2-5: Tangent plane

Just as the tangent line is a line tangential to a curve, this is called the tangent plane. In higher dimensions it is called the tangent hyperplane.

Tangent hyperplane

The set $H = \{x \mid \frac{\partial f}{\partial x}(x^0)\cdot(x - x^0) = 0\}$ is the tangent hyperplane of the function *f* at x^0

As is intuitively clear, a differentiable function is concave if and only if the graph of all the tangent hyperplanes lie above the graph of the function.

Proposition B.2-3: A differentiable function *f* is concave on the convex set $X \subset \mathbb{R}^n$ if and only if for any x^0 and $x^1 \in X$,

$$f(x^1) \le f(x^0) + \frac{\partial f}{\partial x}(x^0)\cdot(x^1 - x^0). \tag{B.2-5}$$

<u>Proof</u>:

The proof of sufficiency proceeds exactly as in the proof for the one-dimensional case (see Proposition A.5-3).

To prove necessity, define

$$g(\lambda) = f(x^\lambda) = f(x^0 + \lambda(x^1 - x^0)).$$

Note that

$$g'(\lambda) = \frac{\partial f}{\partial x}(x^\lambda)\cdot(x^1 - x^0). \tag{B.2-6}$$

Because *f* is concave,

$$g(\lambda) = f(x^\lambda) = f(x^0 + \lambda(x^1 - x^0)) \ge (1-\lambda)f(x^0) + \lambda f(x^1).$$

Because $g(0) = f(x^0)$, it follows that for all $\lambda \in (0,1)$,

$$g(\lambda) - g(0) \ge \lambda(f(x^1) - f(x^0)).$$

Rearranging,

$$\frac{g(\lambda) - g(0)}{\lambda - 0} \ge f(x^1) - f(x^0).$$

Taking the limit as $\lambda \to 0$,

$$g'(0) \geq f(x^1) - f(x^0).$$

Finally, appealing to (B.2-6) and setting $\lambda = 0$,

$$\frac{\partial f}{\partial x}(x^0) \cdot (x^1 - x^0) \geq f(x^1) - f(x^0).$$

QED

The second equivalence theorem for the one-variable case (Proposition A.5-4) also has its generalization.

Proposition B.2-4: A twice differentiable function *f* is concave on the convex set $X \subset \mathbb{R}^n$ if and only if, for all x^0 and $x \in X$, the quadratic form

$$q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0)(x_i - x_i^0)(x_j - x_j^0) \text{ is negative semi-definite.}$$

Proof: For any x^0 and $x^1 = x^0 + z$ define $x = x^0 + \theta z$ and consider the function $g(\theta) = f(x^0 + \theta z)$. For the two-dimensional case, this is the curve depicted in Figure B.2-4.

To establish necessity we begin by showing that, if *f* is concave then $g(\theta)$ must be concave, that is

$$g(\theta^\lambda) \geq (1-\lambda)g(\theta^0) + \lambda g(\theta^1) \text{ for any } \theta^0, \theta^1 \text{ and convex combination } \theta^\lambda.$$

From the definition of *g*,

$$g(\theta^\lambda) = f(x^0 + (1-\lambda)\theta^0 z + \lambda\theta^1 z)$$

$$= f((1-\lambda)(x^0 + \theta^0 z) + \lambda(x^0 + \theta^1 z))$$

Appealing to the concavity of *f* it follows that

$$g(\theta^\lambda) \geq (1-\lambda) f(x^0 + \theta^0 z) + \lambda f(x^0 + \theta^1 z)$$

$$= (1-\lambda)g(\theta^0) + \lambda g(\theta^1).$$

Thus g is concave and so $g''(0) \leq 0$.

From the definition of *g*,

$$g'(\theta) = \frac{\partial f}{\partial x}(x^0 + \theta z) \cdot z \text{ and } g''(\theta) = \sum_{i=1}^{n} \sum_{j}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0 + \theta z) z_i z_j.$$

Hence $\sum_{i=1}^{n}\sum_{j}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0) z_i z_j = \sum_{i=1}^{n}\sum_{j}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0)(x_i - x_i^0)(x_j - x_j^0) \le 0$.

To prove sufficiency, note that if the quadratic form is negative, then $g''(\theta) \le 0$, that is $g'(\theta)$ is decreasing on the interval $[0,1]$. Thus $g'(\theta) \le g'(0)$ for all $\theta \in [0,1]$

From the definition of g,

$$f(x^1) - f(x^0) = g(1) - g(0) = \int_0^1 g'(\theta) d\theta \le \int_0^1 g'(0) d\theta = g'(0)$$

$$= \frac{\partial f}{\partial x}(x^0) \cdot z = \frac{\partial f}{\partial x}(x^0) \cdot (x^1 - x^0)$$

QED

Appealing to Propositions B.2-2 and B.2-4, it follows that a twice differentiable function of two variables $f(x_1, x_2)$, is concave if and only if

(i) $\frac{\partial^2 f}{\partial x_j{}^2} \le 0, \quad j = 1, 2$ and (ii) $\frac{\partial^2 f}{\partial x_1{}^2} \frac{\partial^2 f}{\partial x_2{}^2} - (\frac{\partial^2 f}{\partial x_1 \partial x_2})^2 \ge 0.$

Example: Cobb-Douglas function

Because of its simplicity, economists make much use of the following "Cobb-Douglas" function:

$$f(x) = x_1{}^{\alpha} x_2{}^{\beta}, \quad \alpha, \beta \ge 0, \quad x \ge 0.$$

First note that $\frac{\partial^2 f}{\partial x_1{}^2} = \alpha(\alpha - 1) x_1{}^{\alpha-2} x_2{}^{\beta}$ and $\frac{\partial^2 f}{\partial x_2{}^2} = \beta(\beta - 1) x_1{}^{\alpha} x_2{}^{\beta-2}$. Appealing to (i), from above, it follows that for *f* to be concave, both α and β must lie on the interval $[0,1]$. Moreover $\frac{\partial^2 f}{\partial x_1 \partial x_2} = \alpha\beta x_1{}^{\alpha-1} x_2{}^{\beta-1}$. Appealing to (ii)

$$\frac{\partial^2 f}{\partial x_1{}^2} \frac{\partial^2 f}{\partial x_2{}^2} - (\frac{\partial^2 f}{\partial x_1 \partial x_2})^2 = [(\alpha - 1)(\beta - 1)\alpha\beta - \alpha^2 \beta^2] x_1{}^{2(\alpha-1)} x_1{}^{2(\beta-1)}$$

$$= \alpha\beta(1 - \alpha - \beta) x_1{}^{2(\alpha-1)} x_1{}^{2(\beta-1)}.$$

Thus for concavity $\alpha + \beta \le 1$. ■

The following propositions are often helpful in determining whether a function is concave.

Proposition B.2-5: Sum of concave functions

If f and g are concave then $f+g$ is concave.

Proposition B.2-6: Concave function of a function

The function $h(x)=g(f(x))$ is concave if g is concave and either

(a) g is increasing and f is concave or (b) f is linear.

The proofs are left as exercises.

Example 1: $f(x)=\sum_{j=1}^{n} \ln x_j$, $x \in \mathbb{R}_+^n$

This is concave because each term in the summation is a concave function. $\blacksquare$

Example 2: $f(x)=(x_1^{\alpha}+x_2^{\beta})^{\gamma}$, $\alpha,\beta,\gamma \in (0,1]$ and $x \gg 0$.

Note that $f(x)=y(x)^{\gamma}$, where $y(x)=x_1^{\alpha}+x_2^{\beta}$. Because $\alpha \in (0,1)$ x^{α} has a negative second derivative and so x_1^{α} is concave. By a symmetrical argument x_2^{β} is concave. Then $y(x)$ is concave because it is the sum of two concave functions. Also y^{γ} is an increasing concave function on $\mathbb{R}^+$ because it has a negative second derivative. Then f is concave because it is an increasing concave function of a concave function.

$\blacksquare$

Quasi-concave and Quasi-convex Functions

In Appendix A we also considered quasi-concave functions. Again the definition generalizes naturally.

Definition: Quasi-concave and Quasi-convex Functions

The function f is quasi-concave on the convex set X if for any vectors x^0 and x^1 in X and any convex combination x^λ, $0 < \lambda < 1$, $f(x^\lambda) \geq Min\{f(x^0), f(x^1)\}$.

The function is quasi-convex if, for any convex combination,

$f(x^\lambda) \leq Max\{f(x^0), f(x^1)\}$.

Note that we can always relabel the vectors x^0 and x^1 so that $f(x^1) \geq f(x^0)$. The following definition of a quasi-concave function is therefore equivalent.

Definition: Quasi-concave Function

The function f is quasi-concave on the convex set X if, for any vectors x^0 and x^1,

$f(x^1) \geq f(x^0) \Rightarrow \ f(x^\lambda) \geq f(x^0)$ for all convex combinations x^λ .

The following propositions are especially helpful in checking for quasi-concavity.[6]

Proposition B.2-7: A concave function is quasi-concave

Proposition B.2-8: An increasing function of a concave function is quasi-concave

If f is concave and g is increasing then $h(x) = g(f(x))$ is quasi-concave.

Example: Cobb-Douglas function $f(x) = x_1^{\alpha_1} x_2^{\alpha_2} .... x_n^{\alpha_n}$, $x \gg 0$

Note that $g(x) = \ln f(x) = \sum_{j=1}^{n} \alpha_j \ln x_j$. Because each term in the summation is a concave function $g(x)$ is concave. Inverting, $f(x) = \ln^{-1} g(x) = e^{g(x)}$. Because the exponential function is increasing, it follows from Proposition B.2-8 that f is quasi-concave.

∎

[6] The proofs are left as exercises.

The next proposition if especially helpful in production theory where it is often assumed that production functions are homogeneous degree 1, that is for all $x \in \mathbb{R}^n_+$ and $\mu > 0,\ \ f(\mu x) = \mu f(x)$.

Proposition B.2-9: A quasi-concave function that is homogeneous of degree 1 is concave

Proof: Because *f* is homogeneous of degree 1 it follows that for any $x^0,\ x^1$ and $\lambda \in (0,1)$

$$(1-\lambda) f(x^0) + \lambda f(x^1) = f((1-\lambda)x^0) + f(\lambda x^1),$$

Also, for some $\theta > 0$

$$f((1-\lambda)x^0) = \theta f(\lambda x^1). \tag{B.2-7}$$

Then

$$(1-\lambda) f(x^0) + \lambda f(x^1) = (1+\theta) f(\lambda x^1). \tag{B.2-8}$$

Because *f* is homogeneous of degree 1 it follows from (B.2-7) that $f((1-\lambda)x^0) = f(\theta\lambda x^1)$. Note that

$$\frac{\theta}{1+\theta}((1-\lambda)x^0 + \lambda x^1) = \frac{\theta}{1+\theta}(1-\lambda)x^0 + \frac{1}{1+\theta}\theta\lambda x^1$$

is a convex combination of $(1-\lambda)x^0$ and $\theta\lambda x^1$. Therefore, by the quasi-concavity of *f*

$$f(\frac{\theta}{1+\theta}((1-\lambda)x^0 + \lambda x^0)) \geq f(\theta\lambda x^1). \tag{B.2-9}$$

Because *f* is homogeneous of degree 1 it follows that

$$\frac{\theta}{1+\theta} f((1-\lambda)x^0 + \lambda x^0) \geq \theta f(\lambda x^1)$$

and hence that

$$f((1-\lambda)x^0 + \lambda x^0) \geq (1+\theta) f(\lambda x^1).$$

Appealing to (B.2-8),

$$f((1-\lambda)x^0 + \lambda x^0) \geq (1-\lambda) f(x^0) + \lambda f(x^1).$$

Q.E.D.

Suppose that x^0 and x^1 are linearly independent so that $(1-\lambda)x^0 \neq \theta\lambda x^1$. Then if f is strictly quasi-concave the inequality in (B.2-9) is strict. We therefore have the following corollary.

Corollary B.2-10: A strictly quasi-concave function that is homogeneous of degree 1 is concave, moreover, if x^0 and x^1 are linearly dependent (i.e. $x^1 \neq \mu x^0$), then

$$f((1-\lambda)x^0 + \lambda x^0) > (1-\lambda) f(x^0) + \lambda f(x^1), \quad 0 < \lambda < 1$$

Example: Cobb-Douglas function $f(x) = x_1^{\alpha_1} x_2^{\alpha_2} .... x_n^{\alpha_n}$, α, $x \gg 0$

We have already seen that for any α, f is quasi-concave. If $\sum_{i=1}^{n} \alpha_i = 1$,

$$f(\lambda x) = (\lambda x_1)^{\alpha_1} (\lambda x_2)^{\alpha_2} ....(\lambda x_n)^{\alpha_n} = \lambda x_1^{\alpha_1} x_2^{\alpha_2} .... x_n^{\alpha_n} = \lambda f(x)$$

Therefore f is homogeneous of degree 1. By Corollary B.2-10 it follows that f is concave. Suppose next that $\sum_{i=1}^{n} \alpha_i = \theta < 1$. Then define $\beta_i = \alpha_i / \theta$. Because $\sum_{i=1}^{n} \beta_i = 1$, the function $g(x) = x_1^{\beta_1} x_2^{\beta_2} .... x_n^{\beta_n}$ is concave. Also

$$f(x) = x_1^{\alpha_1} x_2^{\alpha_2} .... x_n^{\alpha_n} = x_1^{\theta\beta_1} x_2^{\theta\beta_2} .... x_n^{\theta\beta_n} = (x_1^{\beta_1} x_2^{\beta_2} .... x_n^{\beta_n})^{\theta} = g(x)^{\theta}$$

Thus f is an increasing concave function of g. By proposition B.2.6, f is concave.

∎

Contour sets

The function $y = f(x_1, x_2)$, depicted below, can be represented on the horizontal plane by mapping all the points for which the function takes on a particular value. This is called a contour set (or "level curve"). In mathematical shorthand,

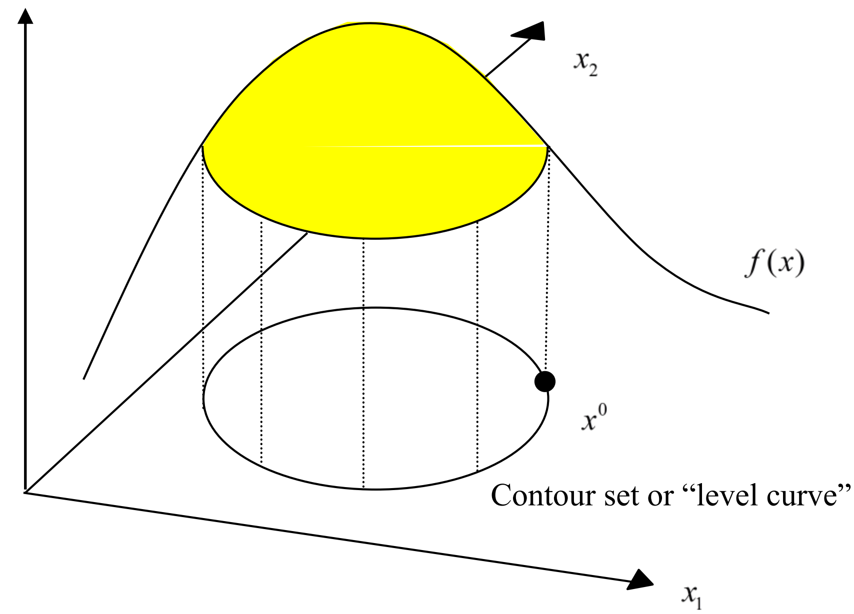
$$C(x^0) = \{x \mid f(x) = f(x^0)\}.$$

The set of vectors for which the function is above some value, that is,

$C_U(x^0) = \{x \mid f(x) \geq f(x^0)\}$

is called an upper contour set. In the figure this is the set of points inside the circle

The lower contour sets $C_L(x^0)$,are similarly defined.



FigureB.2-6: Contour set

Example: Consumer Choice

Suppose that x^* solves $\underset{x}{Max}\{U(x) \mid p \cdot x \leq I\}$. With $n = 2$ we can depict the solution in a standard indifference curve budget-line diagram.

∎

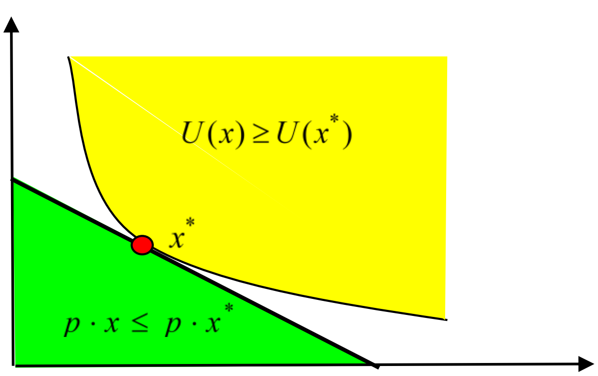


Figure B.2-7: Upper and lower contour sets

The set of consumption vectors to the right of the indifference curve is an upper contour set of the utility function U. The set of consumption vectors below the budget line is a lower contour set of the expenditure function $e(x) = p \cdot x$.

Often in economic analysis it is helpful (and natural) to assume that the upper contour sets of certain functions are convex. As we shall now see, this is equivalent to assuming that the function is quasi-concave.

Proposition B.2-11: A function is quasi-concave if and only if the upper contour sets of the function are convex.

Proof: Suppose f is quasi-concave and for any $\hat{x}$, consider vectors x^0 and x^1 which lie in the upper contour set $C_U(\hat{x})$ of this function. That is $f(x^0) \geq f(\hat{x})$ and $f(x^1) \geq f(\hat{x})$. Without loss of generality we may assume that $f(x^1) \geq f(x^0)$. From the definition of quasi-concavity, for any convex combination,

$$x^\lambda = (1-\lambda)x^0 + \lambda x^1, \quad f(x^\lambda) \geq f(x^0).$$

Because $f(x^0) \geq f(\hat{x})$, it follows that for all convex combinations $f(x^\lambda) \geq f(\hat{x})$. Hence all convex combinations lie in the upper contour set. Thus $C_U(\hat{x})$ is convex.

Conversely, if the upper contour sets are convex, then f is quasi-concave. To demonstrate this, suppose $f(x^1) \geq f(x^0)$. Then x^0 and x^1 are both in $C_U(x^0)$ therefore, by the convexity of $C_U(x^0)$, all convex combination lie in this set. That is $f(x^\lambda) \geq f(x^0)$ for all λ, $0 < \lambda < 1$.

QED

If f is differentiable, then it has the following linear approximation at x^0.

$$f^L(x) = f(x^0) + \frac{\partial f}{\partial x}(x^0) \cdot (x - x^0).$$

The contour set for $f^L(x)$ through x^0 is the set of points satisfying

$$f^L(x) = f^L(x^0)\text{, that is } \frac{\partial f}{\partial x}(x^0)\cdot(x - x^0) = 0\,.$$

This contour set is depicted in the figure below along with the upper contour set $X_U = \{x \mid f(x) \geq f(x^0)\}$. In the two-dimensional case the contour set for the linear approximation is the tangent line. With $n > 2$ this becomes the tangent hyperplane.

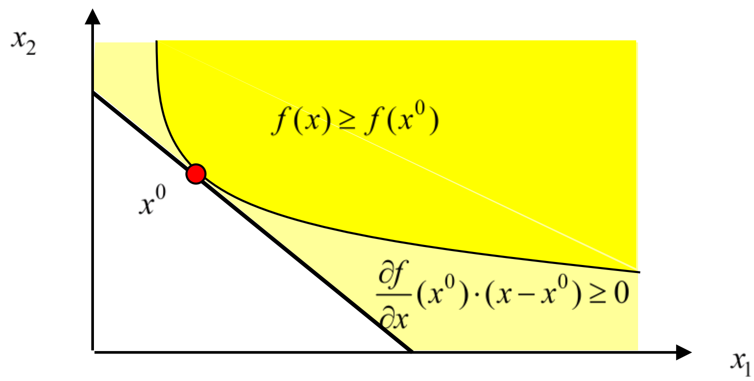


Figure B.2-8: Tangent Line

From the diagram it is clear that if *f* is quasi-concave, so that the boundary of the upper contour set is bowed out, then the upper contour set for *f* must be everywhere above the tangent hyperplane. We now show that this is true for the *n*-dimensional case and that the converse is also true.

Proposition B.2-12: Quasi-concavity of a differentiable function

The function $f \in \mathbb{C}^1$ is quasi-concave if and only if

$$f(x) \geq f(x^0) \;\Rightarrow\; \frac{\partial f}{\partial x}(x^0)\cdot(x - x^0) \geq 0.$$

(B.2-10)

Proof: We first show that quasi-concavity implies (B.2-10). The upper contour set $\{x \mid f(x) \geq f(x^0)\}$ and supporting hyperplane through x^0 are depicted below. Consider any point *x* in the upper contour set. Given that f is quasi-concave, if $f(x) \geq f(x^0)$ then for any convex combination $x^\lambda = (1-\lambda)x^0 + \lambda x = x^0 + \lambda(x - x^0)$, $f(x^\lambda) \geq f(x^0)$.

Define $g(\lambda) \equiv f(x^{\lambda}) = f(x^0 + \lambda(x - x^0))$. Differentiating by λ,

$$\frac{dg}{d\lambda}(\lambda) = \frac{\partial f}{\partial x}(x^{\lambda}) \cdot (x - x^0).$$

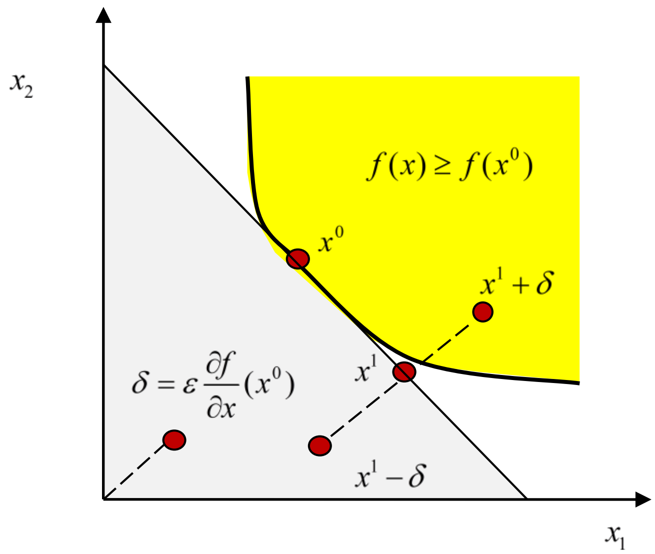


Figure B.2-9: Quasi-concave function

But, we have just argued that for all $\lambda \in (0,1)$, $g(\lambda) = f(x^{\lambda}) \geq f(x^0) = g(0)$. Hence $\frac{dg}{d\lambda}(0) \geq 0$, that is, $\frac{\partial f}{\partial x}(x^0) \cdot (x - x^0) \geq 0$.

To prove the converse, we suppose that the hypothesis is false and seek a contradiction. That is we assume that condition (B.2-10) holds and there exists some x^0, x^1 and convex combination $x^{\beta} = (1-\beta)x^0 + \beta x^1$ such that

$$f(x^1) \geq f(x^0) > f(x^{\beta}) \ .$$

Once again we define the function $g(\lambda) = f((1-\lambda)x^0 + \lambda x^1) = f(x^0 + \lambda(x^1 - x^0))$. Then

$$g(1) \geq g(0) > g(\beta).$$

It follows from the continuity of f that $g(\lambda)$ is continuous. Thus there exists some $\alpha \in [0, \beta)$ such that

$$g(\alpha) = g(0) \text{ and } g(\lambda) < g(0) \text{ for all } \lambda \in (\alpha, \beta]. \qquad \text{(B.2-11)}$$

This is depicted below.

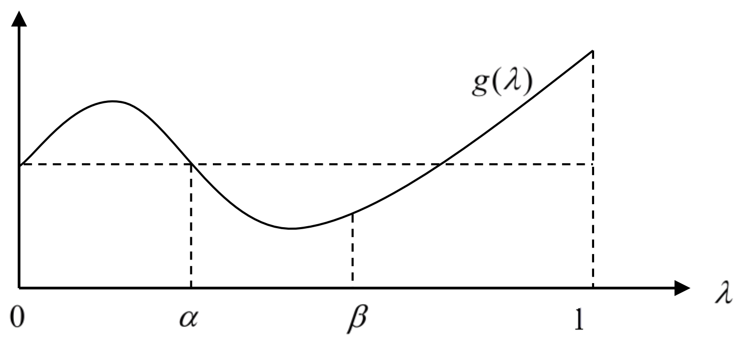


Figure B.2-10: Proof of the converse

Then for all $\lambda \in [\alpha, \beta]$

$$f(x^\lambda) = g(\lambda) \le g(1) = f(x^1).$$

Appealing to (B.2-10), it follows that for all $\lambda \in [\alpha, \beta]$,

$$\frac{\partial f}{\partial x}(x^\lambda) \cdot (x^1 - x^\lambda) \ge 0.$$

Also $x^1 - x^\lambda = (1-\lambda)(x^1 - x^0)$. Therefore

$$\frac{\partial f}{\partial x}(x^\lambda) \cdot (x^1 - x^0) \ge 0, \text{ for all } \lambda \in [\alpha, \beta].$$

(B.2-12)

Next note that

$$g'(\lambda) = \frac{d}{d\lambda} f(x^0 + \lambda(x^1 - x^0)) = \frac{\partial f}{\partial x}(x^\lambda) \cdot (x^1 - x^0).$$

Appealing to (B.2-12), it follows that $g'(\lambda) \ge 0$ for all $\lambda \in [\alpha, \beta]$. Then $g(\beta) \ge g(\alpha)$.

But this is impossible because it contradicts (B.2-11).

QED

We also have the following Corollary.

Corollary B.2-13: Upper contour sets and tangent hyperplanes of a quasi-concave function

For any quasi-concave and differentiable function f such that $\frac{\partial f}{\partial x}(x^0) \neq 0$,

$$f(x) > f(x^0) \ \Rightarrow \ \frac{\partial f}{\partial x}(x^0)\cdot(x - x^0) > 0 .$$

Proof: From Proposition B.2-12 we know that

$$f(x) > f(x^0) \ \Rightarrow \ \frac{\partial f}{\partial x}(x^0)\cdot(x - x^0) \geq 0. \qquad \text{(B.2-13)}$$

Either the corollary holds, or there is some $x^1 \neq x^0$ such that $\frac{\partial f}{\partial x}(x^0)\cdot(x^1 - x^0) = 0$. If so x^1 is on the supporting hyperplane and $f(x^1) > f(x^0)$. Consider x^1 in Figure B.2-9. Also shown is the gradient vector scaled down by the factor ε, that is $\delta = \varepsilon \frac{\partial f}{\partial x}(x^0)$. Consider $x^2 = x^1 - \delta$. Because $f(x^1) > f(x^0)$ it follows that if we choose ε positive and sufficiently small ε, we can make δ positive and sufficiently small so that $f(x^2) = f(x^1 - \delta) > f(x^0)$.

Also

$$\frac{\partial f}{\partial x}(x^0)\cdot(x^2 - x^0) = \frac{\partial f}{\partial x}(x^0)\cdot(x^2 - x^1) + \frac{\partial f}{\partial x}(x^0)\cdot(x^1 - x^0) .$$

But the second term on the right-hand side is zero and $x^2 - x^1 = -\varepsilon \frac{\partial f}{\partial x}(x^0)$. We have therefore established that

$$f(x^2) > f(x^0) \ \text{and} \ \frac{\partial f}{\partial x}(x^0)\cdot(x^2 - x^0) = -\varepsilon \frac{\partial f}{\partial x}(x^0)\cdot\frac{\partial f}{\partial x}(x^0) < 0 .$$

But this result contradicts (B.2-13) so there can be no such x^2.

QED

Exercises

Exercise B.2-1: Positive semi-definite quadratic form

What are the necessary and sufficient conditions for $q(x)=\sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}x_ix_j$ to be everywhere non-negative?

Hint: If $q(x)$ is non-negative, $-q(x)$ must be non-positive.

Exercise B.2-2: Positive definite quadratic form

What are the necessary and sufficient conditions for $q(x)=\sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}x_ix_j$ to be strictly positive for all $x \neq 0$?

Exercise B.2-3: Concave and quasi-concave functions

Sketch proofs of Propositions B.2-5 and B.2-6.

Exercise B.2-4: Convex lower contour sets

(a) Show that the lower contour sets of a function are convex if and only if the function is quasi-convex.

(b) The output of commodity *j* has labor input requirements $f_j(q_j)$, $j=1,...,n$ where $f_j(\cdot)$ is convex. Show that the total labor requirements function $L(q)=\sum_{j=1}^{n} f_j(q_j)$ is quasi-convex.

(c) Hence show that if the supply of labor is fixed, the set of feasible outputs is convex.

Exercise B.2-5: Contour Sets

In each case depict the contour sets $f(x)=0,\ \ f(x)=k,\ \ f(x)=-k$ where $k>0$.

(a) $f(x)=p_1x_1+p_2x_2$ where $p>0$

(b) $f(x)=x_1x_2$

(c) $f(x)=x_2^{\,2}-4x_1^{\,2}=(x_2-2x_1)(x_2+2x_1)$

*(d) $f(x)=-x_1^{\,2}+6x_1x_2+7x_2^{\,2}$

B.3 TRANSFORMATIONS OF VECTORS

Key Ideas: linear transformations, matrix, quadratic form, inverse

A function transforms an *n*-dimensional vector into a real number (a one-dimensional vector). More generally "transformations" map *n*-dimensional vectors into *m*-dimensional vectors. Here we consider the family of linear transformations. Consider the following $m \times n$ "matrix", an array of *m* rows and *n* columns

$$\mathbf{A} = \begin{bmatrix} a_{11} & . & . & . & a_{1n} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{m1} & . & . & . & a_{mn} \end{bmatrix}$$

Each of the rows is an *n*-dimensional vector. Denote the *i*th row as $a_i = (a_{i1},...,a_{in})$. Then we transform the *n*-dimensional vector *x* into an *m*-dimensional vector *y* by taking the product of each row and *x*:

$$\begin{aligned} &y_1 = a_1 \cdot x \\ &. \\ &. \qquad\qquad . \\ &. \\ &y_m = a_m \cdot x \end{aligned}$$

It proves useful to write the vectors *x* and *y* as special matrices with one column ("column vectors"). Then we write

$$y = \mathbf{A}x \quad \text{where} \quad y_i = (i\text{th row of } \mathbf{A}) \cdot x .$$

Next suppose we make two such linear transformations, $y = \mathbf{A}x$ and $z = \mathbf{B}y$. These are depicted below. Combining the two transformations we may write

$$z = \mathbf{B}(\mathbf{A}x).$$

An important property of linear transformations is that the product of two such transformations is itself a linear transformation, that is

$$z = \mathbf{B}(\mathbf{A}x) = \mathbf{C}x .$$

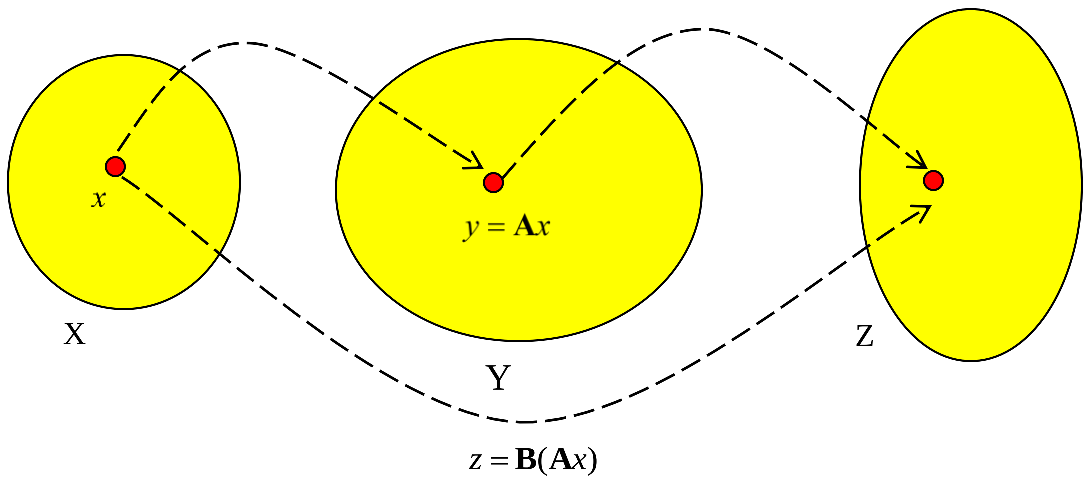


Figure B.3-1: Product of linear transformations

Moreover there is a simple rule for computing the elements of the new matrix $\mathbf{C}$.

$$c_{ij} = (i\text{th row of } \mathbf{B}) \cdot (j\text{th column of } \mathbf{A}).$$

We demonstrate this for the 2×2 case. Consider two linear transformations of two-dimensional vectors:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Multiplying each row of the matrix **A** by the column vector x,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$$

and hence

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$$

$$= \begin{bmatrix} (b_{11}a_{11} + b_{12}a_{12})x_1 + (b_{11}a_{12} + b_{12}a_{22})x_2 \\ (b_{21}a_{11} + b_{22}a_{21})x_1 + (b_{21}a_{12} + b_{22}a_{22})x_2 \end{bmatrix}$$

$$= \begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \quad \text{where } c_{ij} = (i\text{th row of } \mathbf{B}) \cdot (j\text{th column of } \mathbf{A}).$$

Transpose of a Matrix

A matrix is "transposed" by rotating it around its leading diagonal. Let $\mathbf{A}'$ be the transpose of $\mathbf{A}$, that is $a'_{ij} = a_{ji}$. Then the ith row and jth column of $\mathbf{A}$ become, respectively, the ith column and jth row of $\mathbf{A}'$. From the rules of matrix multiplication,

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'.$$

To confirm this, let $\mathbf{C} = \mathbf{AB}$. We need to establish that $\mathbf{C}' = \mathbf{B}'\mathbf{A}'$, that is, the ijth element of $\mathbf{C}'$ is the product of the ith row of $\mathbf{B}'$ and the jth column of $\mathbf{A}'$.

$$c'_{ij} = c_{ji} = (j\text{th row of } \mathbf{A}) \cdot (i\text{th column of } \mathbf{B}) = (i\text{th column of } \mathbf{B}) \cdot (j\text{th row of } \mathbf{A})$$

$$= (i\text{th row of } \mathbf{B}') \cdot (j\text{th column of } \mathbf{A}').$$

Note that the transpose of a $n \times 1$ column vector x is a $1 \times n$ row vector x'. Thus the square of the distance between two vectors x and y can be written either as an inner product or a matrix product

$$\|x - y\|^2 = (x - y) \cdot (x - y) = (x - y)'(x - y).$$

Quadratic Form

Consider the matrix product $x'\mathbf{A}x$. In the two- dimensional case,

$$x'\mathbf{A}x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{21}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}.$$

$$= a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2.$$

This is the general quadratic function of a vector $x \in \mathbb{R}^2$. Generally, with an n vector x and $n \times n$ matrix $\mathbf{A}$, the general quadratic function (or "quadratic form" of the matrix $\mathbf{A}$) can be expressed as follows:

$$q(x) = x'\mathbf{A}x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \,.$$

The $x_i x_j$ term is $a_{ij} + a_{ji}$. Thus there is no loss in generality in assuming that the matrix **A** is symmetric, that is, $a_{ij} = a_{ji}$.

Consider the partial derivatives of the quadratic form $q(x)$.

Collecting all the terms in x_1,

$$\begin{aligned} q(x) = a_{11}x_1^{\,2} + a_{12}x_1x_2 + ... + a_{1n}x_1x_n \\ + a_{21}x_2x_1 + ... + a_{n1}x_nx_1 + \text{other terms independent of } x_1. \end{aligned}$$

Given the symmetry of **A,**

$$q(x) = a_{11}x_1^{\,2} + 2(a_{12}x_1x_2 + ... + a_{1n}x_1x_n) + \text{terms independent of } x_1.$$

Then

$$\frac{\partial q}{\partial x_1} = 2(a_{11},...,a_{1n}) \cdot (x_1,...,x_n) = 2(\text{first row of } \mathbf{A}) \cdot x \,.$$

Similarly,

$$\frac{\partial q}{\partial x_i} = 2(i\text{th row of } \mathbf{A}) \cdot x \,. \qquad \text{(B.3-1)}$$

Thus the vector of partial derivatives,

$$\frac{\partial q}{\partial x} = 2\mathbf{A}x \,.$$

Also, from (B.3-1),

$$\frac{\partial^2 q}{\partial x_j \partial x_i} = 2a_{ij} \,.$$

Thus the matrix of second partial derivatives,

$$\left[\frac{\partial^2 q}{\partial x_i \partial x_j} \right] = 2\mathbf{A} \,.$$

We have therefore proved the following proposition.

Proposition B.3-1: Derivatives of a quadratic form

If $q(x) = x'\mathbf{A}x$ then $\dfrac{\partial q}{\partial x} = 2\mathbf{A}x$ and $\left[\dfrac{\partial^2 q}{\partial x_i \partial x_j}\right] = 2\mathbf{A}$

Square Matrices

One important class of linear transformations maps *n*-dimensional vectors into other *n*-dimensional vectors, that is $y = \mathbf{A}x$, where both *x* and *y* are *n*-dimensional vectors. To multiply **A** and *x* the matrix must have *n* columns. For *y* to have *n* elements, there must be *n* rows of **A** so the matrix **A** is an *n*-dimensional square matrix. Consider the simplest case *n* = 2.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} x_1 + \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} x_2 \equiv x_1 a_{(1)} + x_2 a_{(2)} .$$

Thus the vector *y* is a linear combination of the two column vectors $a_{(1)}$ and $a_{(2)}$. In the *n*-dimensional case the vector *y* is the linear combination of the *n* columns of **A**. Geometrically, we can draw this linear combination by first drawing each column vector, then scaling each vector and summing the scaled vectors.

With $n = 2$, two possibilities are depicted below .

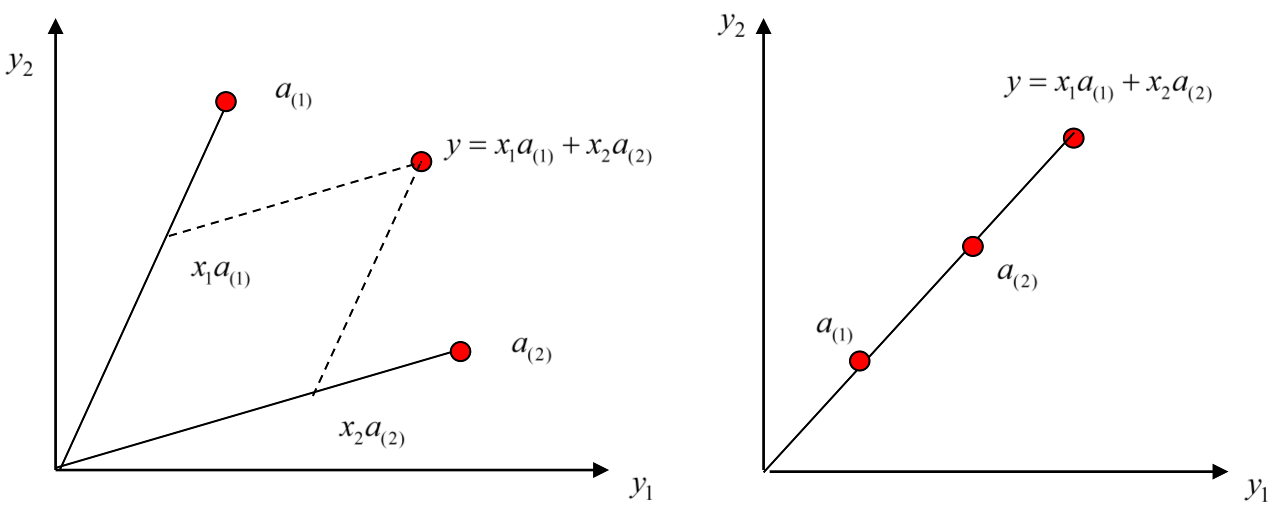


Figure B.3-2: Linearly dependent and independent vectors

In the first case the two column vectors do not differ simply by scale so the mapping is onto the entire two-dimensional space. In this case the columns of **A** are said to be linearly independent. In the second case $a_{(2)} = \theta a_{(1)}$ and the mapping is from the two-dimensional space of vectors $x = (x_1, x_2)$ onto a one-dimensional line. In this case the column vectors are said to be linearly dependent.

Similarly, with 3 dimensions,

$$y = Ax = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} x_2 + \begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \end{bmatrix} x_3 \equiv x_1 a_{(1)} + x_2 a_{(2)} + x_3 a_{(3)} .$$

In the figure below, the linear combinations of the first two vectors map out a plane in 3-dimensional space. If the third vector is in this plane which is a linear combination of the other two vectors, then all the linear combinations of the three vectors also lie in the plane.

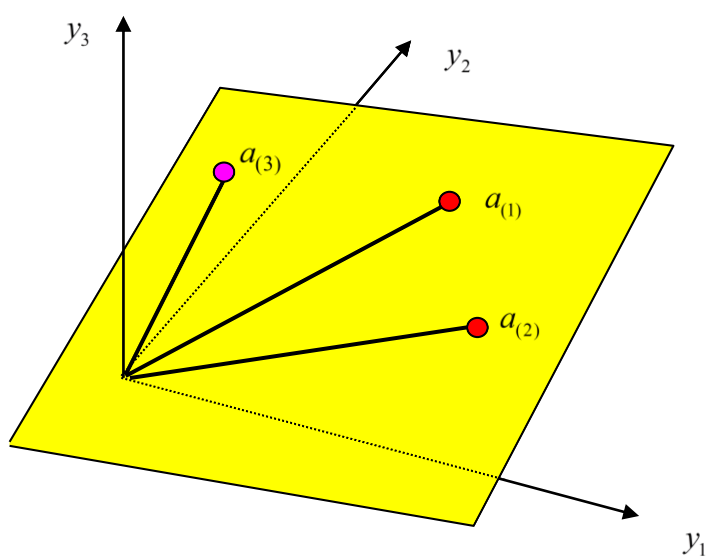


Figure B.3-3: Linear dependence

Generally, the columns of a square *n*-matrix are independent if no column can be expressed as a linear combination of the other *n*-1 columns. In this case the mapping is onto the entire n-dimensional space. For any n-dimensional matrix $\mathbf{A}$ we can determine whether the columns are independent by computing the determinant of $\mathbf{A}$ denoted $|\mathbf{A}|$.

Definition: Determinant of a Square Matrix

$n = 2$

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

n > 2

the determinant is defined recursively as follows:

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \left|\mathbf{A}_{(ij)}\right|$$ where $\mathbf{A}_{(ij)}$ is the matrix obtained after deleting the ith row and jth column of $\mathbf{A}$.

Example 1: Linearly independent columns

Expanding along the first row,

$$|\mathbf{A}| = \begin{vmatrix} 2 & 3 & 0 \\ 4 & 1 & 3 \\ 1 & 1 & 2 \end{vmatrix} = (-1)^2(2)\begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} + (-1)^3(3)\begin{vmatrix} 4 & 3 \\ 1 & 2 \end{vmatrix} + (-1)^4(0)\begin{vmatrix} 4 & 1 \\ 1 & 1 \end{vmatrix}$$

$$= (-1)^2(2)(-1) + (-1)^3(3)(5) = -17$$

Example 2: Linearly dependent columns

Expanding along the third row,

$$|\mathbf{A}| = \begin{vmatrix} 2 & 0 & 4 \\ 1 & 1 & 4 \\ 0 & 2 & 4 \end{vmatrix} = (-1)^4(0)\begin{vmatrix} 0 & 4 \\ 1 & 4 \end{vmatrix} + (-1)^5(2)\begin{vmatrix} 2 & 4 \\ 1 & 4 \end{vmatrix} + (-1)^6(4)\begin{vmatrix} 2 & 0 \\ 1 & 1 \end{vmatrix} = 0$$

Note that in the second example, the third column vector is twice the sum of the first two columns. Therefore the vectors are not independent.

Identity Matrix

The identity matrix **I** has all off-diagonal elements of zero and all on-diagonal elements of 1. Following the rules of matrix multiplication, the identity matrix maps every vector onto itself, that is, $\mathbf{I}x = x$.

Inverse Matrix

Let **A** be a square matrix with non-zero determinant. Let **B** be a second matrix that maps the vector *y* back to *x*. Then

$$x = \mathbf{B}y = \mathbf{B}(\mathbf{A}x) = (\mathbf{BA})x = \mathbf{I}x\,.$$

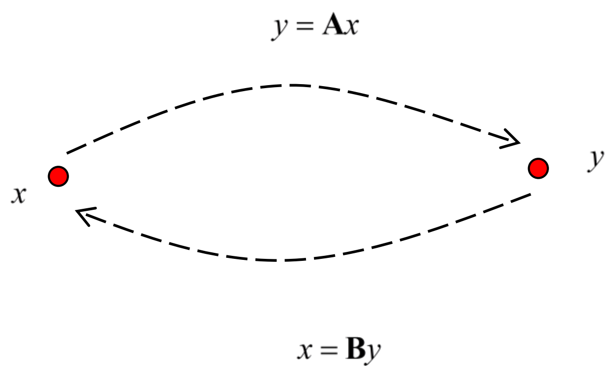


Figure B.3-4: Transformation and its inverse

Thus the "inverse" mapping **B** times the matrix **A** must equal the identity matrix. Also, starting from a vector *y* and mapping to *x* and back again,

$$y = \mathbf{A}x = \mathbf{A}(\mathbf{B}y) = (\mathbf{AB})y = \mathbf{I}y\,.$$

Thus regardless of the order of multiplication, the product of the matrix **A** and its inverse is the identity matrix. We write the inverse as $\mathbf{A}^{-1}$. Then

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}\,.$$

To compute the inverse we first define the "minor determinants" of a matrix **A**. For each element a_{ij} of **A,** the minor matrix $\mathbf{M}_{ij}$ is obtained by eliminating the *i*th row and *j*th column of **A**. The minor determinant $\left|\mathbf{M}_{ij}\right|$ is the determinant of this minor matrix.

Computation of the Inverse

If $\left|\mathbf{A}\right| \neq 0$ the square matrix A has an inverse $\mathbf{A}^{-1} = \left[a_{ij}^{-1} \right]$ where

$$a_{ij}^{-1} = \frac{(-1)^{i+j}\left|\mathbf{M}_{ji}\right|}{\left|\mathbf{A}\right|}.$$

Example: $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

$\left|\mathbf{M}_{11}\right| = a_{22}$, $\left|\mathbf{M}_{12}\right| = a_{21}$, $\left|\mathbf{M}_{21}\right| = a_{12}$, $\left|\mathbf{M}_{22}\right| = a_{11}$

Then

$$\mathbf{A}^{-1} = \begin{bmatrix} \dfrac{a_{22}}{\left|\mathbf{A}\right|} & -\dfrac{a_{12}}{\left|\mathbf{A}\right|} \\ -\dfrac{a_{21}}{\left|\mathbf{A}\right|} & \dfrac{a_{11}}{\left|\mathbf{A}\right|} \end{bmatrix} = \frac{1}{\left|\mathbf{A}\right|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

If the columns of **A** are independent, we have argued that the mapping is onto the entire *n*-dimensional space. Thus for any vector *b* there must be some vector *x* such that

$$\mathbf{A}x = b.$$

Mathematically, because the inverse exists,

$$\mathbf{A}^{-1}(\mathbf{A}x) = (\mathbf{A}^{-1}\mathbf{A})x) = \mathbf{I}x = x = \mathbf{A}^{-1}b.$$

Cramer's Rule

Having computed an inverse, it is simply a mechanical exercise to solve a system of linear equations. Cramer's Rule provides a convenient way of summarizing this computation. Define $\mathbf{A}_{j,b}$ to be the matrix A with the *j*th column replaced by *b*. Then

$$x_j = \frac{\left|\mathbf{A}_{j,b}\right|}{\left|\mathbf{A}\right|}.$$

Example:

$$\begin{aligned} 3x_1 + 4x_2 &= 11 \\ 2x_1 + 3x_2 &= 8 \end{aligned} \quad \text{that is} \quad \begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 11 \\ 8 \end{bmatrix}$$

Appealing to Cramer's Rule,

$$x_1 = \frac{\begin{vmatrix} \mathbf{11} & 4 \\ \mathbf{8} & 3 \end{vmatrix}}{\begin{vmatrix} 3 & 4 \\ 2 & 3 \end{vmatrix}} = 1, \quad x_2 = \frac{\begin{vmatrix} 3 & \mathbf{11} \\ 2 & \mathbf{8} \end{vmatrix}}{\begin{vmatrix} 3 & 4 \\ 2 & 3 \end{vmatrix}} = 2 .$$

You can plug back into the equations to confirm that this is the solution.

Negative semi-definite matrix

In the last section we examined the two-variable case and derived necessary conditions under which a quadratic form is negative semi-definite. We now state an important generalization of this result.

Definition: Principal minor determinant

The ith principal minor of the square matrix **A** is the matrix obtained by deleting all but the first *i* rows and *i* columns of **A**. The principal minor determinant is the determinant of this matrix.

Proposition B.3-2: Necessary and sufficient conditions for a quadratic form to the negative semi-definite.

The quadratic form $x'\mathbf{A}x$ is negative semi-definite if and only if the *i*th principal minor determinant of **A** is of sign $(-1)^i$, $i = 1,...,n$.

Consider the two-variable case. The principal minors of **A** are $|a_{11}|$ and $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$. Thus the necessary and sufficient conditions are $a_{11} \le 0$ and $a_{11}a_{22} - a_{12}a_{21} \ge 0$.

Finally consider the quadratic form $x'\mathbf{A}x$ where the *n* variables satisfy the linear constraint $b'x = 0$ where $b \neq 0$. Suppose that $b_i \neq 0$. Then

$$x_i = \frac{1}{b_i} \sum_{\substack{j=1 \\ j \neq i}}^{n} b_j x_j .$$

We can therefore substitute for x_i in the quadratic form and obtain a new quadratic form with $n-1$ variables. Tedious algebra then yields the following proposition.

Proposition B.3-3:

The quadratic form $x'\mathbf{A}x$ is negative for all x satisfying the linear constraint $b'x=0$ if and only if the *i*th principal minor of the matrix

$$\begin{bmatrix} 0 & b' \\ b & \mathbf{A} \end{bmatrix}$$

is of sign $(-1)^{1+i}$, $i=1,...,n+1$

Application: Necessary condition for quasi-concavity

An appeal to this last result makes it relatively easy to determine the restrictions that quasi-concavity places on a twice continuously differentiable function. From the previous section we know that for a quasi-concave function

$$f(x) > f(x^0) \Rightarrow \frac{\partial f}{\partial x}(x^0)\cdot(x-x^0) > 0 .$$

It follows that if $\frac{\partial f}{\partial x}(x^0)\cdot(x-x^0) \le 0$ then $f(x) \le f(x^0)$.

Consider any x^1 on the tangent hyperplane at x^0. That is $\frac{\partial f}{\partial x}(x^0)\cdot(x^1-x^0)=0$.

Then all convex combinations $x^\lambda = x^0 + \lambda(x^1-x^0)$ lie on this tangent hyperplane, that is

$$\frac{\partial f}{\partial x}(x^0)\cdot(x^1-x^0)=0 \quad . \qquad \text{(B.3-2)}$$

Thus for any x^λ, $f(x^\lambda) \le f(x^0)$.

Define $g(\lambda) = f(x^\lambda) = f(x^0+\lambda\Delta)$, where $\Delta = x^1 - x^0$. Given the above arguments $g(\lambda) \le g(0)$. But

$$g'(\lambda) = \frac{d}{d\lambda} f(x^0+\lambda\Delta) = \frac{\partial f}{\partial x}(x^\lambda)\cdot\Delta = \frac{\partial f}{\partial x}(x^\lambda)\cdot(x^1-x^0) .$$

Appealing to (B.3-2), it follows that $g'(0)=0$.

We wish to show that $g''(0) \le 0$. To do so we suppose instead that $g''(0) > 0$ and seek a contradiction. If $g''(0) > 0$, then in some neighborhood of 0, $g'(\lambda)$ is strictly increasing and, because $g'(0)=0$, it follows that $g'(\lambda)>0$ over this neighborhood and so for some $\lambda>0$, $g(\lambda)>0$. But we have already argued that $g(\lambda)\le g(0)$. Thus we have a contradiction. It follows that $g''(0)\le 0$. But

$$g''(\lambda)=\frac{d^2}{d\lambda^2}f(x^0+\lambda\Delta)=\frac{d}{d\lambda}\frac{\partial f}{\partial x}(x^0+\lambda\Delta)\cdot\Delta .$$

Then $g''(0)=\sum_j\sum_i f_{x_i x_j}(x^0)\Delta_i\Delta_j$ where we use subscripts to indicate partial derivatives.

Thus if f is quasi-concave,

$$\sum_i f_{x_i}(x^0)\Delta_i=0 \Rightarrow \sum_j\sum_i f_{x_i x_j}(x^0)\Delta_i\Delta_j\le 0 .$$

Appealing to Proposition B.3-3 we have the following result.

Proposition B.3-4: Necessary conditions for quasi-concavity

If U is quasi-concave, then the ith principal minor of

$$\begin{bmatrix} 0 & U_{x_1} & . & . & U_{x_n} \\ U_{x_1} & U_{x_1x_1} & . & . & U_{x_1x_n} \\ . & . & & & . \\ . & . & & & . \\ U_{x_n} & U_{x_nx_1} & . & . & U_{x_nx_n} \end{bmatrix}$$

has sign $(-1)^{1+i}$.

Exercise B.3-1: Linear Economy

Production of x_j units of commodity j $(j=1,2)$ uses up a_{j1} units of commodity 1 and a_{j2} units of commodity 2. Each unit of gross output of commodity j also requires a_{0j} units of labor. The supply of labor is L.

(a) If the gross output vector is x, confirm that the net output vector is $y=(I-A)x$.

(b) Suppose that

$$a_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix} \quad L = 80$$

Show that if only commodity 1 is produced, the net output vector is $y^0 = (40, -60)$ and solve for the net output vector y^1 if only commodity 2 is produced.

(c) Depict these vectors in a neat figure and depict also the net output if half the labor is used to produce commodity 1 and half to produce commodity 2. Which of the three net output vectors is feasible?

(d) Show that net output must satisfy the following constraint

$$b \cdot y = a_0{}'(I - A)^{-1} y \le L .$$

and solve for the vector b given the data of the problem.

(e) What is the maximum feasible net output of commodity 1?

Exercise B.3-2: Simple Regression

A data analyst wishes to estimate the relationship between X and Y. She has n observations of each variable. Define the vectors $X^0 = (1,...,1)$, $X^1 = (X_1,...,X_n)$ and $Y = (Y_1,...,Y_n)$. The estimated relationship is to be a linear combination of X^0 and X^1. That is, $\hat{Y} = aX^0 + bX^1$. This is depicted below for the case of 3 observations.

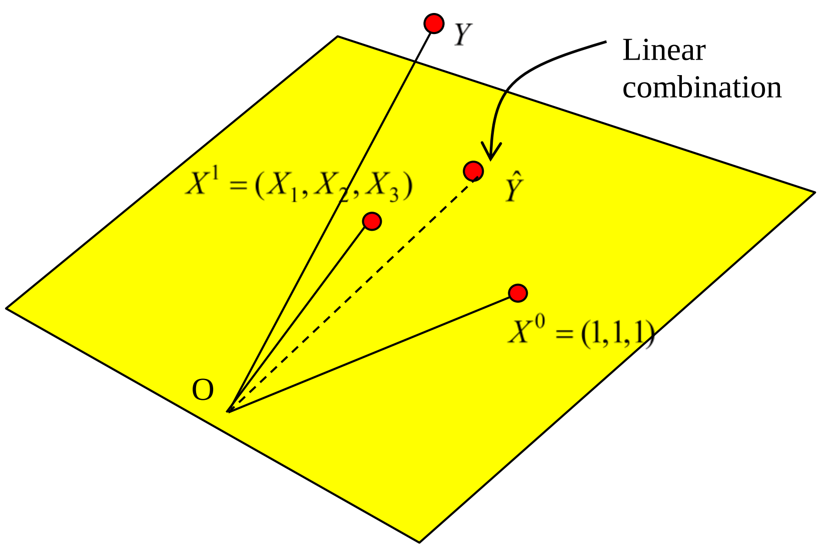


Figure B.3-5: Simple regression

(a) Explain why the square of the distance between Y and $\hat{Y}$ is

$$\left\|Y-\hat{Y}\right\|^2=\sum_{j=1}^{n}(Y_j-a-bX_j)^2 .$$

(b) Show that the distance minimizing parameter is $a=\bar{Y}-b\bar{X}$ where $\bar{X}$ and $\bar{Y}$ are the sample means.

(c) Define $x=X-\bar{X}$ and $y=Y-\bar{Y}$ and appeal to (b) to show that

$$\left\|Y-\hat{Y}\right\|^2=\sum_{j=1}^{n}(y_j-bx_j)^2 .$$

(d) Solve for the distance minimizing value of *b*.

Exercise B.3-3: Least squared error

The $t\times n$ matrix **X** is a matrix composed of *n* column vectors each of dimension *t*.

(a) If *a* is a *n* dimensional column vector, explain why $\mathbf{X}a$ is a linearly weighted combination of the column vectors $x_{(1)},...,x_{(n)}$ of **X**.

(b) Let *y* be another *t*-dimensional column vector. Show that the distance between this vector and the linear combination can be written as follows:

$$e'e=y'y-2a'\mathbf{X}'y+a'\mathbf{X}'\mathbf{X}a .$$

(c) Show that the gradient of this difference is

$$\frac{\partial}{\partial a}e'e=-2\mathbf{X}'y+2\mathbf{X}'\mathbf{X}a .$$

(d) Choosing *a* to minimize $e'e$ is the same as choosing *a* to maximize $-e'e$. Show that $-e'e$ is a concave function of *a*. Thus the FOC are both necessary and sufficient for a maximum.

HINT: Appeal to Proposition B.3-1 to obtain an expression for the matrix of second partial derivatives and then show that this matrix is negative definite.

(e) Hence confirm that to minimize this difference, the weights chosen must satisfy

$$a=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y .$$

B.4: SYSTEMS OF LINEAR DIFFERENCE EQUATIONS

Key Ideas: phase diagram, eigenvalues, eigenvectors, Walrasian price dynamics

Economists often study the evolution of a vector of economic variables over time. Consider the following difference equation system.

$$x(t+1) = \mathbf{A}x(t) \tag{B.4.1}$$

where $x(t)$ is an n-dimensional vector and $\mathbf{A}$ is an $n \times n$ matrix. We will refer to $x(t)$ as the "state" variable at t.

We focus on the special two-dimensional case. The general principles apply also for higher n.

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} .$$

Note that $x(t+1) - x(t) = \mathbf{A}x(t) - \mathbf{I}x(t) = (\mathbf{A} - \mathbf{I})x(t)$. Thus if there is a stationary point $\bar{x}$, $(\mathbf{A} - \mathbf{I})\bar{x} = \mathbf{0}$. Therefore as long as the determinant of $\mathbf{A} - \mathbf{I}$ is non zero, the unique stationary point is $\bar{x} = \mathbf{0}$. We make this assumption throughout.

An example

We can readily characterize the dynamics. Suppose that

$$\mathbf{A} = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}.$$

Then

$$\begin{bmatrix} \Delta x_1(t) \\ \Delta x_2(t) \end{bmatrix} \equiv x(t+1) - x(t) = (\mathbf{A} - \mathbf{I})x(t) = \begin{bmatrix} 4x_1(t) - x_2(t) \\ 3x_1(t) \end{bmatrix}. \tag{B.4.2}$$

Thus $\Delta x_2(t) = 3x_1(t)$, so that $x_2(t)$ is increasing if and only if $x_1(t) \geq 0$. Also, from (B.4.2),

$$\Delta x_1(t) = x_1(t+1) - x_1(t) = 4x_1(t) - x_2(t) .$$

Thus $x_1(t)$ is increasing if and only if $x_2(t) \leq 4x_1(t)$.

Consider the figure below. The two lines $\Delta x_1 = 0$ and $\Delta x_2 = 0$ separate the plane into four regions. In each region the state moves in a particular direction (NE, SE, NW or

SW). Thus the dynamic paths are said to be in a particular " phase." For example in phase I, shaded in the figure, $x_2 \geq 0$ and $x_2 \leq 4x_1$ so that state variables are increasing.

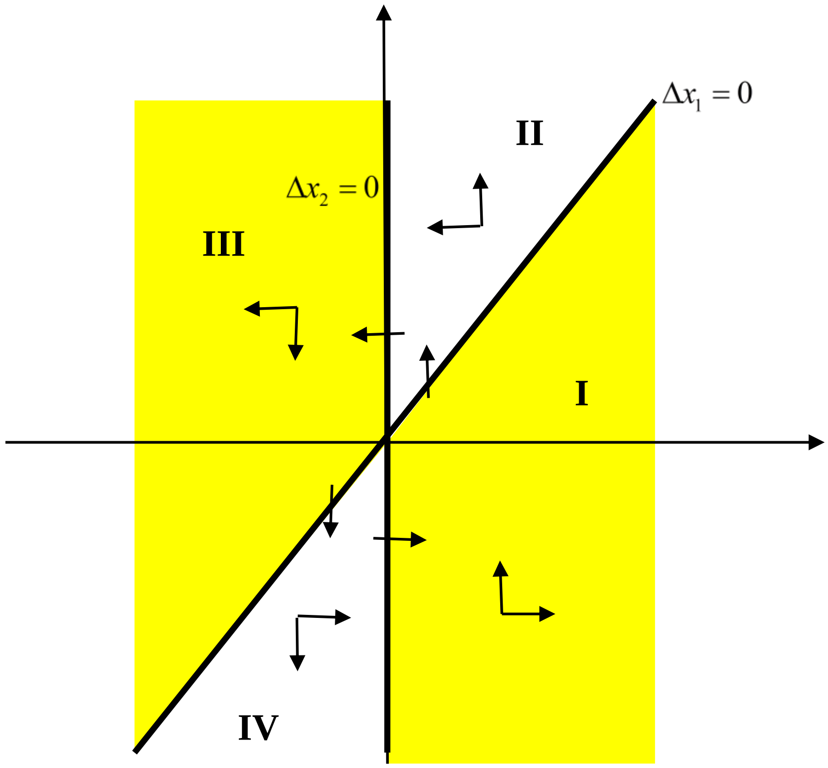


Fig B.4-1: The four phases

We can learn a lot more about the dynamics by seeking a subspace of the state space in which the two variables grow at the same rate. From (B.4.2)

$$\frac{\Delta x_2}{\Delta x_1} = \frac{a_{21}x_1 + (a_{22}-1)x_2}{(a_{11}-1)x_1 + a_{12}x_2} = \frac{3x_1}{4x_1 - x_2}.$$

Suppose that $x(t)$ grows proportionally, that is $x(t+1) = \lambda x(t)$. If this is the case then

$$\frac{\Delta x_2}{\Delta x_1} = \frac{x_2(t+1) - x_2(t)}{x_1(t+1) - x_1(t)} = \frac{(\lambda - 1)x_2(t)}{(\lambda - 1)x_1(t)} = \frac{x_2(t)}{x_1(t)} \quad .$$

(Given our assumption that $|\mathbf{A} \text{ - } \mathbf{I}| \neq 0$, $\lambda \neq 1$.)

Then, for the example,

$$\frac{\Delta x_2}{\Delta x_1} = \frac{3x_1}{4x_1 - x_2} = \frac{x_2}{x_1}.$$

Hence $3x_1^{\,2} - 4x_1x_2 + x_2^{\,2} = (3x_1 - x_2)(x_1 - x_2) = 0.$

Solving this quadratic equation, either $x_2 = x_1$ or $x_2 = 3x_1$. In the first case, by substituting back into the difference equation we obtain

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_1(t) \end{bmatrix} = 4 \begin{bmatrix} x_1(t) \\ x_1(t) \end{bmatrix}.$$

Thus the growth factor $\lambda_1 = 4$. Similarly it can be shown that the growth factor in the second case is $\lambda_2 = 2$. The constant growth lines are also shown as the dashed lines in the "phase diagram" below. Along the dashed line $x_2 = 3x_1$, $x(t)$ doubles from period to period. Along the dashed line $x_2 = x_1$, $x(t)$ quadruples form period to period.

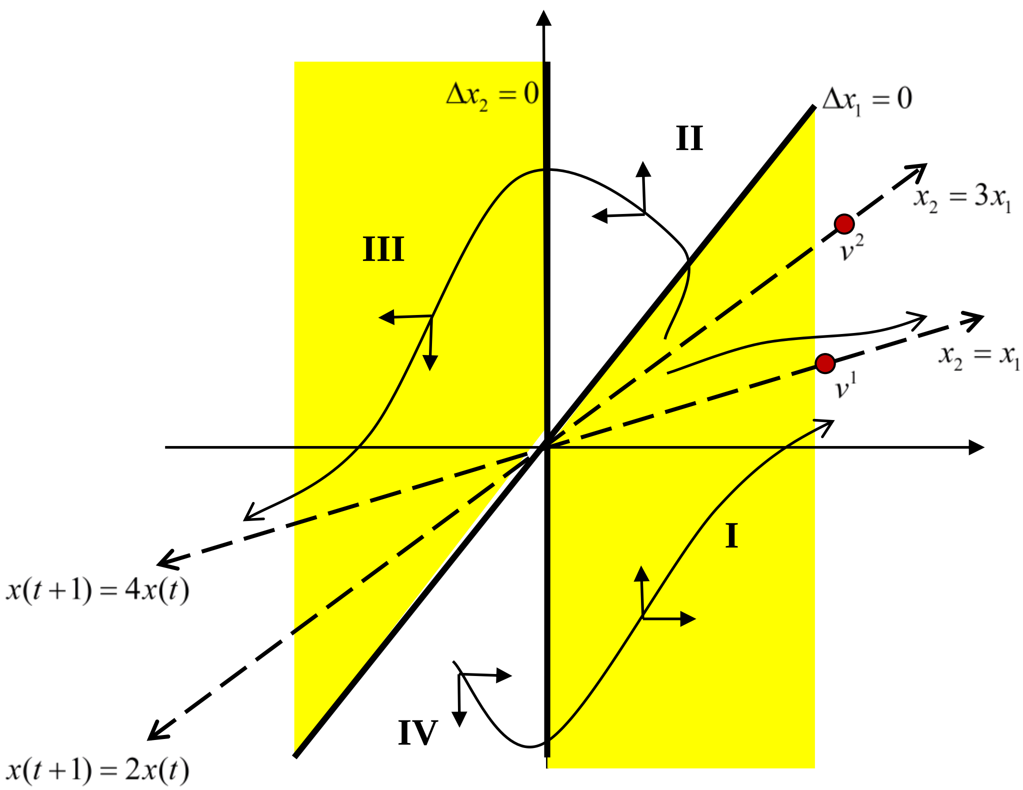


Fig B.4-2: Phase diagram

Pick the vector $v^1 >> 0$ on the unit circle and on the line $x_2 = x_1$ and vector $v^2 >> 0$ on the unit circle and line $x_2 = 3x_1$. Any initial vector $x(1)$ can be expressed as a linear combination of the vectors v^1 and v^2 because they are linearly independent, that is

$$x(1) = \alpha_1 v^1 + \alpha_2 v^2.$$

Then

$$\begin{aligned} x(2) &= \mathbf{A}x(1) = \mathbf{A}(\alpha_1 v^1 + \alpha_2 v^2) \\ &= \alpha_1 \mathbf{A}v^1 + \alpha_2 \mathbf{A}v^2 \\ &= \alpha_1 \lambda_1 v^1 + \alpha_2 \lambda_2 v^2. \end{aligned}$$

Repeating this argument, $x(t+1) = \alpha_1 \lambda_1^{\,t} v^1 + \alpha_2 \lambda_2^{\,t} v^2$. Because $|\lambda_1| > |\lambda_2|$, it follows that for large *t*

$$x(t+1) = \lambda_1^{\,t}(\alpha_1 v^1 + \alpha_2 (\frac{\lambda_2}{\lambda_1})^t v^2) \approx \lambda_1^{\,t} \alpha_1 v^1 . \qquad \text{(B.4.3)}$$

Thus for large *t*, the growth rate is approximately equal to λ_1. Moreover

$$\frac{x_2(t+1)}{x_1(t+1)} \to \frac{v_2^1}{v_2^1}.$$

Thus the ratio of x_2 to x_1 approaches the ratio along the high constant growth line.

Consider the phase diagram once more. Suppose that $x(1) > 0$. The parameters (α_1, α_2) are chosen so that $x(1) = \alpha_1 v^1 + \alpha_2 v^2$. Then if $x(1)$ lies between the two constant growth lines both α_1 and α_2 are positive. Then for all *t*,

$$x(t+1) = \alpha_1 \lambda_1^{\,t} v^1 + \alpha_2 \lambda_2^{\,t} v^2.$$

Therefore $x(t+1)$ is a weighted average of v^1 and v^2 where both weights are positive. Thus the path lies between $x_2 = x_1$ and $x_2 = 3x_1$ for all t. Moreover, as *t* grows large, $x_2(t+1)/x_1(t+1) \to 1$ and the growth rate approaches $\lambda^1 = 4$.

Suppose next that $x(1)$ lies below the high constant growth line $x_2 = x_1$. Then $\alpha_2 < 0 < \alpha_1$ and so $x(t+1)$ lies below the high constant growth line. In the limit $x_2(t+1)/x_1(t+1) \to 1$ and again the growth factor approaches $\lambda_1 = 4$.

The third possibility is that $x(1)$ lies above the low constant growth rate line $x_2 = 3x_1$. Then $\alpha_1 < 0 < \alpha_2$ and it follows that for sufficiently large *t*, $x(t+1)$ is negative. As depicted, $x(1)$ is in phase I. The path moves through phase II and ends in phase III.

The general two-variable model

In the example, the long run outcome is for both state variables to grow at approximately the same rate. We now show that for a big class of linear difference equation systems this will be the case. We begin by seeking initial state $x(1)$ such that

$$x(t+1) = \mathbf{A}x(t) = \lambda x(t) = \lambda \mathbf{I} x(t) .$$

Rearranging,

$$x(t+1) - \lambda x(t) = \mathbf{A}x(t) - \lambda \mathbf{I} x(t) = (\mathbf{A} - \lambda\, \mathbf{I})x(t) = 0 .$$

For a stationary state this must hold with $\lambda = 1$, that is $(\mathbf{A} - \mathbf{I})x(t) = 0$. We assume that $|\mathbf{A} - \mathbf{I}| \neq 0$, so that the matrix $\mathbf{A}$ - $\mathbf{I}$ is invertible. Then the unique solution to the equation system $(\mathbf{A} - \mathbf{I})x(t) = 0$ is $x(t) = 0$.

Next consider $\lambda \neq 1$. For a constant growth path the equation system $(\mathbf{A} - \lambda\, \mathbf{I})x(t) = 0$ must have a non-zero solution. Arguing as above, this is not possible if $\mathbf{A} - \lambda\, \mathbf{I}$ is invertible. Hence for constant growth,

$$|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = \lambda^2 - (a_{11} + a_{22})\lambda + |\mathbf{A}| = 0 . \qquad \text{(B.4.4)}$$

This equation is known as the characteristic equation of the matrix **A**. The two roots λ_1, λ_2 of the quadratic equation are known as the characteristic roots or eigenvalues. Associated with each eigenvalue is a value of the state vector v^i on the unit circle. These are known as eigenvectors. Suppose that the two eigenvalues differ. To show that the eigenvectors must be independent, suppose instead that $v^1 = \theta v^2 = v$. Because both eigenvectors lie on the unit circle, $(\mathbf{A} - \lambda_1 \mathbf{I})v = 0 = (\mathbf{A} - \lambda_2 \mathbf{I})v$ and therefore $(\lambda_2 - \lambda_1)\mathbf{I}\, v = (\lambda_2 - \lambda_1)v = 0$. But this is impossible because $v \neq 0$. Therefore the two eigenvectors are indeed independent.

Real eigenvectors

We can also write the characteristic equation

$$\lambda^2 - (a_{11} + a_{22})\lambda + |\mathbf{A}| = 0$$

as follows, where λ_1 and λ_2 are the characteristic roots

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda - \lambda_1\lambda_2 = 0. \tag{B.4.5}$$

Note that the sum of the roots is the sum of the terms in the leading diagonal of $\mathbf{A}$ and the product of the roots is the determinant of $\mathbf{A}$.

Solving the quadratic characteristic equation yields the two roots

$$\begin{aligned} \lambda_1 &= \tfrac{1}{2}(a_{11} + a_{22}) + \sqrt{\tfrac{1}{4}(a_{11} + a_{22})^2 - |A|} \\ \lambda_2 &= \tfrac{1}{2}(a_{11} + a_{22}) - \sqrt{\tfrac{1}{4}(a_{11} + a_{22})^2 - |A|} \end{aligned}. \tag{B.4.6}$$

Thus there are two distinct real roots if and only if

$$(a_{11} + a_{22})^2 - 4|A| = (a_{11} - a_{22})^2 + 4a_{21}a_{12} > 0.$$

Thus a sufficient condition for two distinct real roots is that $a_{12}a_{21} > 0$.

Suppose this condition holds. Arguing exactly as in the example, any initial state vector $x(1)$ can be expressed as a linear combination of v^1 and v^2, that is

$$x(1) = \alpha_1 v^1 + \alpha_2 v^2.$$

Because $Av^i = \lambda_i v^i$, $i = 1, 2$, it follows that $A^t v^i = \lambda_i^t v^i$, $i = 1, 2$. Then

$$x(t+1) = \alpha_1 \lambda_1^t v^1 + \alpha_2 \lambda_2^t v^2.$$

Thus the long run dynamics are determined by the eigenvalue with the larger absolute value.

Example: Walrasian price dynamics

Suppose that the Walrasian auctioneer adjusts prices in proportion to excess demand. Let $\bar{p}$ be the equilibrium price and define $x(t) = p(t) - \bar{p}$. Then the excess demand vector can be written as follows:

$$\hat{e}(x(t)) = e(\bar{p} + x(t)).$$

Note that $\hat{e}(0) = \mathbf{0}$. Consider the linear approximation of $\hat{e}(x)$ in the neighborhood of the equilibrium

$$\hat{e}(x) = \mathbf{E}x = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Then the dynamic adjustment process can be written as follows:

$$x(t+1) = \mathbf{A}x(t) \text{ where } \mathbf{A} = k\mathbf{E} \text{ and } k > 0.$$

Assumptions:

(a) Own price effects are strictly negative $e_{11}, e_{22} < 0$.

(b) Cross price effects are strictly positive $e_{12}, e_{21} > 0$.

(c) Own price effects dominate cross price effects $|\mathbf{E}| = e_{11}e_{22} - e_{21}e_{12} > 0$.

(d) The rate of adjustment is slow $k(e_{11} + e_{22}) > -1$.

By assumption (b) the cross effects are of the same sign and it follows that $a_{12}a_{21} = k^2 e_{12}e_{21} > 0$ so that there are two distinct real roots. By assumption (c) the product of the roots is positive so both have the same sign. By assumption (a) the sum of the roots, $\lambda_1 + \lambda_2 = a_{11} + a_{22}$ is negative. Because λ_1 and λ_2 have the same sign it follows that both must be negative. Finally, by assumption (d), $a_{11} + a_{22} = k(e_{11} + e_{22}) > -1$.

Because $|\mathbf{A}| > 0$, $\sqrt{\tfrac{1}{4}(a_{11} + a_{22})^2 - |A|} < \tfrac{1}{2}|a_{11} + a_{22}|$. Then, from (B.4.6) $\lambda_2 = \tfrac{1}{2}(a_{11} + a_{22}) - \sqrt{\tfrac{1}{4}(a_{11} + a_{22})^2 - |A|} > a_{11} + a_{22} > -1$. Because both eigenvalues are negative both therefore lie in the interval (0,1). Then the dynamic system is stable and oscillates towards the equilibrium point.[7]

Complex eigenvalues

Consider the solution to the characteristic equation. Suppose that the expression under the square root in (B.4.6) is negative. Then define $\alpha = \tfrac{1}{2}(a_{11} + a_{22})$ and $\beta^2 = |A| - \tfrac{1}{4}(a_{11} + a_{22})^2$. Then (B.4.6) can be rewritten as follows:

$$\begin{aligned} \lambda_1 &= \alpha + \sqrt{-\beta^2} \\ \lambda_2 &= \alpha - \sqrt{-\beta^2} \end{aligned}.$$

[7] If the rate of adjustment parameter k is large enough, $\lambda_2 < \tfrac{1}{2}(a_{11} + a_{22}) = \tfrac{1}{2}k(e_{11} + e_{22}) < -1$, thus making the Walrasian adjustment system unstable.

Thus there is no solution in terms of real numbers and so there can be no constant growth paths. We employ a remarkable mathematical slight of hand and introduce complex numbers. We define $i = \sqrt{-1}$. Then

$$\lambda_1 = \alpha + i\beta \text{ and } \lambda_2 = \alpha - i\beta.$$

Example:

Suppose that $\mathbf{A} = \begin{bmatrix} 1 & -\frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$.

An example of an explosive cycle is depicted in Figure B.4-3 below. The initial values $x(1) = (x_1(1), x_2(1))$ are indicated by the dot.

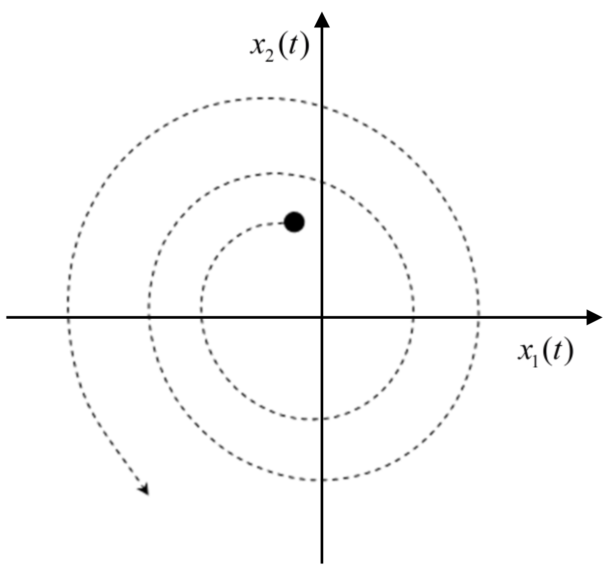


Figure B.4-3: Solution with complex eigenvalues

The characteristic equation is

$$|\mathbf{A} - \lambda\mathbf{I}| = \begin{vmatrix} 1-\lambda & -1/2 \\ 1/2 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 + \tfrac{1}{4} = 0.$$

While there is no real root we note that $i^2 = -1$ and so $(1-\lambda)^2 = \frac{1}{4} i^2$.

Taking the square root,

$$\lambda = 1 \pm \tfrac{1}{2} i. \qquad \blacksquare$$

As we shall see, the example illustrates a general result. Both $x_1(t)$ and $x_2(t)$ must cycle indefinitely. As depicted the state moves further and further from the stationary point so the system is unstable. However, for other parameter values, the state vector oscillates towards the stationary point. The third possibility is that the system has a cycle that is neither damped not explosive. This is the case, for example if

$$\mathbf{A} = \begin{bmatrix} \cos \pi / n & -\sin \pi / n \\ \sin \pi / n & \cos \pi / n \end{bmatrix}.$$

The solution is depicted in Figure B-4-4.

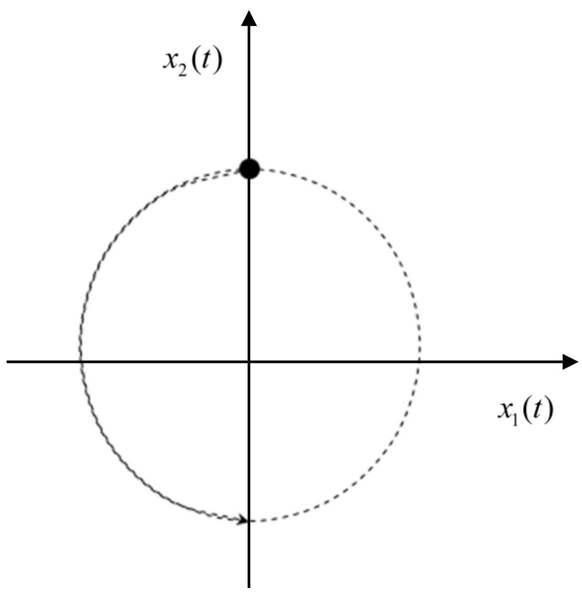


Figure B.4-4: Cycle is neither damped nor explosive.

As the first step in solving for the general solution, note that the eigenvalue λ_1 and associated eigenvector v must satisfy the constant growth condition:

$$(\mathbf{A} - \lambda_1 \mathbf{I})v = 0.$$

That is $(a_{11}-\lambda_1)v_1 + a_{12}v_2 = 0$. Without loss of generality we may choose $v_1 = 1$ then $v_2 = -(a_{11}-\lambda_1)/a_{12}$. Because λ_1 is a complex number, so is v_2. We will write this more succinctly as $v_2 = k_1 + ik_2$.

It proves extremely useful to express the vector of parameters (α,β) in polar coordinates. See Figure B.4-5 below. Note that $r = \sqrt{\alpha^2+\beta^2}$ and $\tan\theta = \beta/\alpha$.

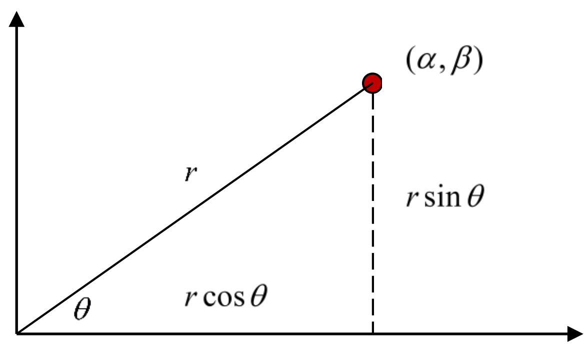


Figure B.4-5: Polar coordinates

Then we can rewrite the eigenvalue as $\lambda_1 = r(\cos\theta + i\sin\theta)$. Next we note[8] that for any z,

$$\cos z + i\sin z = e^{iz}.$$

Then

$$(\cos\theta + i\sin\theta)^t = e^{i\theta t} = \cos\theta t + i\sin\theta t.$$

And so if $x(1) = v$,

$$x(t+1) = Ax(t) = A^t v = \lambda_1^t v$$

$$= r^t e^{i\theta t}\begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = r^t e^{i\theta t}\begin{bmatrix} 1 \\ k_1 + ik_2 \end{bmatrix}$$

[8] In Exercise B.4-1you are asked to confirm this by appealing to Taylor's expansion.

$$= r^t(\cos\theta t + i\sin\theta t)\begin{bmatrix} 1 \\ k_1 + ik_2 \end{bmatrix}.$$

Collecting real and complex terms,

$$x(t+1) = r^t\begin{bmatrix} \cos\theta t \\ k_1\cos\theta t - k_2\sin\theta t \end{bmatrix} + ir^t\begin{bmatrix} \sin\theta t \\ k_1\sin\theta t + k_2\cos\theta t \end{bmatrix}. \qquad \text{(B.4.7)}$$

Define

$$v^1(t+1) = r^t\begin{bmatrix} \cos\theta t \\ k_1\cos\theta t - k_2\sin\theta t \end{bmatrix} \text{ and } v^2(t+1) = r^t\begin{bmatrix} \sin\theta t \\ k_1\sin\theta t + k_2\cos\theta t \end{bmatrix}.$$

We now argue that both $v^1(t)$ and $v^2(t)$ are solutions to the difference equation system.

Because $v^1(t) + iv^2(t)$ is a solution,

$$x(t+1) = v^1(t+1) + iv^2(t+1) = \mathbf{A}x(t) = \mathbf{A}v^1(t) + i\mathbf{A}v^2(t).$$

Collecting real and complex terms,

$$v^1(t+1) = \mathbf{A}v^1(t) \quad \text{and} \quad v^2(t+1) = \mathbf{A}v^2(t).$$

Choose α_1 and α_2 so that $\alpha_1 v^1(1) + \alpha_2 v^2(1) = x(1)$. Because $v^1(t)$ and $v^2(t)$ are solutions, so is any linear combination. Then

$$x(t) = \alpha_1 v^1(t) + \alpha_2 v^2(t)$$

is the general solution, given the initial state $x(1)$.

Stable and unstable systems

From (B.4.6)

$$\lambda_1 = \alpha + i\beta = \tfrac{1}{2}(a_{11} + a_{22}) + \sqrt{\tfrac{1}{4}(a_{11} + a_{22})^2 - |A|}$$
$$\lambda_2 = \alpha - i\beta = \tfrac{1}{2}(a_{11} + a_{22}) - \sqrt{\tfrac{1}{4}(a_{11} + a_{22})^2 - |A|}\ .$$

From Figure B.4-5, we can rewrite these eigenvalues in polar coordinates as follows:

$$\lambda_1 = r(\cos\theta + i\sin\theta), \quad \lambda_2 = r(\cos\theta - i\sin\theta)\ ,$$

where $r = \sqrt{\alpha^2 + \beta^2}$ and $(\alpha, \beta) = (\tfrac{1}{2}(a_{11} + a_{22}), \sqrt{|A| - \tfrac{1}{4}(a_{11} + a_{22})^2})$.

Substituting for α and β, $r = \sqrt{|A|}$.

From (B.4.7) the amplitude of the oscillation is r^t at time *t*. Thus the amplitude of the oscillations is increasing if $r > 1$ and decreasing if $r < 1$. Therefore the cycles of an oscillating system with complex eigenvalues are damped if and only if $|A| < 1$.

Example (continued)

If $\mathbf{A} = \begin{bmatrix} 1 & -\frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$, we have seen that the eigenvalues are $\lambda_1 = 1 + \frac{1}{2}i$ and $\lambda_2 = 1 - \frac{1}{2}i$.

Transforming the eigenvalues into polar coordinates,

$\lambda_1 = \frac{\sqrt{5}}{2}(\cos\bar{\theta}t + i\sin\bar{\theta}t)$ and $\lambda_2 = \frac{\sqrt{5}}{2}(\cos\bar{\theta}t - i\sin\bar{\theta}t)$, where $\tan\bar{\theta} = \frac{1}{2}$.

Note that because $r = |A| > 1$, the amplitude of the oscillations is increasing. Thus the dynamic system is unstable.

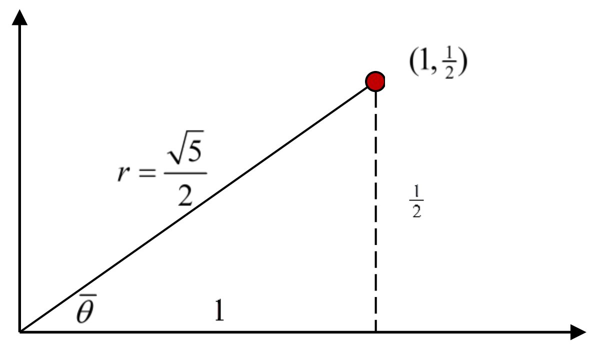


Figure B.4-6: Polar coordinates

Exercise B.4-1: Polar representation of a complex number

(a) Appeal to Taylor's expansion around $z = 0$ to show that

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} \text{ and } \sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!}.$$

Therefore $\cos z + i\sin z = 1 + iz - \frac{z^2}{2!} + i\frac{z^3}{3!} + \frac{z^4}{4!} + .$

(b) Confirm that this summation can be rewritten as follows.

$$\cos z + i\sin z = 1 + iz + \frac{i^2 z^2}{2!} + \frac{i^3 z^3}{3!} + \frac{i^4 z^4}{4!} + \ldots$$

(c) Appeal to Taylor's Expansion around $z = 0$ to show that

$$e^{iz} = 1 + \frac{iz}{1!} + \frac{i^2 z^2}{2!} + \frac{i^3 z^3}{3!} + \frac{i^4 z^4}{4!} + \ldots.$$

Hence $e^{iz} = \cos z + i\sin z$

Exercise B.4-2: Saddle point dynamics

Suppose that $x(t+1) = \mathbf{A}x(t)$ where $\mathbf{A} = \begin{bmatrix} 2 & \frac{3}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$.

(a) Show that the eigenvalues are $\lambda_1 = \frac{5}{2}$ and $\lambda_2 = \frac{1}{2}$ and solve for the corresponding eigenvectors.

(b) Draw the phase portrait.

(c) For what values of $x(1)$ does the solution converge to zero?

Exercise B.4-3: Oscillations

Suppose that $x(t+1) = \mathbf{A}x(t)$ where $\mathbf{A} = \begin{bmatrix} -1 & 1 \\ -5 & 1 \end{bmatrix}$.

(a) Solve for the eigenvalues.

(b) Obtain expressions for the two eigenvalues in polar coordinates.

(c) Hence show that $x(t)$ oscillates explosively with cycles of 4 periods.

Exercise B.4-4: Walrasian dynamics

The difference between the period t Walrasian price vector and the equilibrium price vector is $x(t)$. The price adjustment is proportional to excess demand, that is

$$x(t+1) = k\mathbf{E}x(t) \text{ where } \mathbf{E} = \begin{bmatrix} -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & -\frac{3}{4} \end{bmatrix}.$$

(a) If $k = 1$ show that both of the eigenvectors $\hat{\lambda}_1$ and $\hat{\lambda}_2$ lie on the interval $(-1, 0)$ so the dynamic adjustment process is stable.

(b) If $k \neq 1$ show that the new eigenvectors $(\lambda_1, \lambda_2) = k(\hat{\lambda}_1, \hat{\lambda}_2)$.

(c) For what values of k is the system stable?

INDEX

B

C

D

E

F

G

H

I

L

M

N

O

P

Q

S

T

V