Optimization and Multivariate Calculus

Joe Patten

August 9, 2018

1 Homogeneity

We say that a function $f(x_1, x_2, ..., x_n)$ is homogeneous of degree k (commonly referred to as HDk) if:

$$f(\alpha x_1, \alpha x_2, ..., \alpha x_n) = \alpha^k f(x_1, x_2, ..., x_n) \text{ for all } \mathbf{x} \text{ and all } \alpha > 0 \tag{1}$$

In Economics, when a production is homogeneous of degree 1, it is said to have constant returns to scale (CRS). If $k > 1$, the production function has increasing returns to scale, and if $k < 1$, the production function has decreasing returns to scale.

Example

Consider the function: $f(x, y) = 5x^2y^3 + 6x^6y^{-1}$. To determine if this function if homogeneous, we need to multiply each input by α:

$$\begin{aligned} f(\alpha x, \alpha y) &= 5(\alpha x)^2(\alpha y)^3 + 6(\alpha x)^6(\alpha y)^{-1} \\ &= \alpha^{2+3}5x^2y^3 + \alpha^{6-1}6x^6y^{-1} \\ &= \alpha^5(5x^2y^3 + 6x^6y^{-1}) \\ &= \alpha^5(f(x, y)) \end{aligned}$$

This function is homogeneous of degree 5 (HD5).

1.1 Euler's Theorem

If we take the derivative of both sides of equation (1) by x_i, we get the following:

$$\begin{aligned} \frac{\partial f(\alpha x_1, \alpha x_2, ..., \alpha x_n)}{\partial x_i} \cdot \alpha &= \alpha^k \frac{\partial f(x_1, x_2, ..., x_n)}{\partial x_i} \\ \frac{\partial f(\alpha x_1, \alpha x_2, ..., \alpha x_n)}{\partial x_i} &= \alpha^{k-1} \frac{\partial f(x_1, x_2, ..., x_n)}{\partial x_i} \end{aligned} \tag{2}$$

We can use the result from equation (2) to derive Euler's Thereom:

Theorem 1. If f is a C^1, homogeneous of degree k function on $\mathbb{R}^n_+$, then it follows:

$$x_1 \frac{\partial f(x)}{\partial x_1} + x_2 \frac{\partial f(x)}{\partial x_2} + \ldots + x_n \frac{\partial f(x)}{\partial x_n} = kf(x)$$

2 Definiteness of Matrix

2.1 Quadratic Form

A function $Q : \mathbb{R}^n \to \mathbb{R}$ is a quadratic form if it is a homogeneous polynomial of degree two. Thus, a quadratic form can be written as:

$$Q(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \tag{3}$$

where $a_{ij} \in \mathbb{R}$

Notice that equation (4) can be written using vectors and matrices:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & (a_{12}+a_{21})/2 & \dots & (a_{1n}+a_{n1})/2 \\ (a_{12}+a_{21})/2 & a_{22} & \dots & (a_{2n}+a_{n2})/2 \\ \vdots & \vdots & \ddots & \vdots \\ (a_{1n}+a_{n1})/2 & (a_{2n}+a_{n2})/2 & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (4)$$

$$= \mathbf{x}^T A\mathbf{x} \quad (5)$$

Notice that the coefficient matrix, which we will call A, is square and symmetric. There is an infinite amount of coefficient matrices that would yield the same quadratic form, however, it is convenient to define A in such a way that it is symmetric.

Example

Let $Q(\mathbf{x}) = 2x_1^2 + 3x_1x_2$. If we put this in matrix form, we would get the following result:

$$Q(\mathbf{x}) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & \frac{3}{2} \\ \frac{3}{2} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Practice

Express the quadratic form $Q(\mathbf{x}) = -3x_1^2 + 2x_1x_2 + 4x_1x_3 - 2x_2^2 + 5x_2x_3$ in matrix form.

2.2 Definiteness

Consider an $n \times n$ symmetric matrix A. A is:

positive definite	if $\mathbf{x}^T A\mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^n - \mathbf{0}$
negative definite	if $\mathbf{x}^T A\mathbf{x} < 0 \quad \forall \mathbf{x} \in \mathbb{R}^n - \mathbf{0}$
positive semidefinite	if $\mathbf{x}^T A\mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$
negative semidefinite	if $\mathbf{x}^T A\mathbf{x} \leq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$
indefinite	if $\mathbf{x}^T A\mathbf{x} > 0$ for some $x \in \mathbb{R}^n$, and $\mathbf{x}^T A\mathbf{x} < 0$ for some $x \in \mathbb{R}^n$

Table 1: Definiteness

2.3 Principal Minors

We can also evaluate the principal minors of A to determine the definiteness of A.

Let A be an $n \times n$ matrix. A kth order principal submatrix is $k \times k$ and is formed by deleting $n-k$ rows, and the same $n-k$ columns. Taking the determinant of a kth order principal submatrix yields a kth order principal minor.

The kth order leading principal submatrix of A, usually written as $|A_k|$, is the left most submatrix in A that is $k \times k$. The determinant of the kth order leading principal submatrix is called the kth order leading principal determinant.

Let A be an $n \times n$ matrix. Then,

- A is positive definite iff all of its leading principal minors are positive.
- A is negative definite iff leading principal minors alternate in sign, and the 1st order princiapl minor is negative.
- A is positive semidefinite iff every principal minor of is is nonnegative.

- A is negative semidefinite iff every principal minor of odd order is nonpositive, and every principal minors of even order is nonnegative.
- A is indefinite iff A does not have any of these patterns.

3 Derivatives

Recall from single-variable calculus, the derivative of a function f with respect to x at point x_0 is defined as:

$$\frac{df(x_0)}{dx} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

If this limit exists, then we say that f is differentiable at x_0. We can extend this definition to talk about derivatives of multivariate functions.

3.1 Partial Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$. The partial derivative of f with respect to variable x_i at $\mathbf{x}^0$ is given by:

$$\frac{\partial f(\mathbf{x}^0)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, x_1, ..., x_i + h, ..., x_n) - f(x_1, x_2, ..., x_i, ..., x_n)}{h}$$

Notice that in this definition, the *ith* variable is affected. To take the partial derivative of variable x_i, we treat all the other variables as constants.

Example

Consider the function: $f(x, y) = 4x^2y^5 + 3x^3y^2 + 6y + 10$.

$\frac{\partial f(x,y)}{\partial x} = 8xy^5 + 9x^2y^2$

$\frac{\partial f(x,y)}{\partial y} = 20x^2y^4 + 6x^3y + 6$

3.2 Gradient Vector

We can put all of the partials of the function $F : \mathbb{R}^n \to \mathbb{R}$ at x^* (which we call the derivative of F) in a row vector:

$$DF_{x^*} = \begin{bmatrix} \frac{\partial F(x^*)}{\partial x_1} & \dots & \frac{\partial F(x^*)}{\partial x_n} \end{bmatrix}$$

This can also be referred to as the Jacobian derivative of F.

We can express the derivative in a column vector:

$$\nabla F_{x^*} = \begin{bmatrix} \frac{\partial F(x^*)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x^*)}{\partial x_n} \end{bmatrix}$$

This representation is usually referred to as the gradient vector.

Example

The gradient vector of our previous example would be:

$$\nabla F = \begin{bmatrix} 8xy^5 + 9x^2y^2 \\ 20x^2y^4 + 6x^3y + 6 \end{bmatrix}$$

3.3 Jacobian Matrix

We won't always be working with functions of the form $F : \mathbb{R}^n \to \mathbb{R}$. We might work with functions of the form $F : \mathbb{R}^n \to \mathbb{R}^m$. A common example example in economics is a production function that has n inputs and m outputs. Considering the production function example, notice that we can write this function as m functions:

$$\begin{aligned} q_1 =& f_1(x_1, x_2, ..., x_n) \\ q_2 =& f_2(x_1, x_2, ..., x_n) \\ &\vdots \\ q_m =& f_1(x_1, x_2, ..., x_n) \end{aligned}$$

We can put the functions and their respective partials in a matrix in order to get the Jacobian Matrix:

$$DF(x^*) = \begin{bmatrix} \frac{\partial f_1(x^*)}{\partial x_1} & \frac{\partial f_1(x^*)}{\partial x_2} & \cdots & \frac{\partial f_1(x^*)}{\partial x_n} \\ \frac{\partial f_2(x^*)}{\partial x_1} & \frac{\partial f_2(x^*)}{\partial x_2} & \cdots & \frac{\partial f_2(x^*)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(x^*)}{\partial x_1} & \frac{\partial f_m(x^*)}{\partial x_2} & \cdots & \frac{\partial f_m(x^*)}{\partial x_n} \end{bmatrix}$$

3.4 Hessian Matrix

Recall that for an function of n variables, there are n partial derivatives. We can take partial derivatives of each partial derivative. The partial derivative of a partial derivative is called the second order partial derivative.

Example

The second order partial derivatives for the example above are defined as:

$\frac{\partial^2 f(x,y)}{\partial x^2} = 8y^5 + 18xy^2$

$\frac{\partial^2 f(x,y)}{\partial y^2} = 80x^2y^3 + 6x^3$

$\frac{\partial^2 f(x,y)}{\partial y \partial x} = 40xy^4 + 18x^2y$

$\frac{\partial^2 f(x,y)}{\partial x \partial y} = 40xy^4 + 18x^2y$

The second order partial derivatives of the form $\frac{\partial^2 f(x,y)}{\partial x \partial y}$ where $x \neq y$ are called the cross partial derivatives. Notice from our example, that $\frac{\partial^2 f(x,y)}{\partial x \partial y} = \frac{\partial^2 f(x,y)}{\partial y \partial x}$. This is always the case with cross partials. We see that:

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}$$

We can put all of these second order partials into a matrix, which is referred to as the Hessian Matrix:

$$\begin{bmatrix} \frac{\partial^2 f(x^*)}{\partial x_1^2} & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x^*)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x^*)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n^2} \end{bmatrix}$$

Let the function $f : A \to \mathbb{R}$ be a C^2 function, where A is a convex and open set in $\mathbb{R}^n$.

- f is strictly concave iff its Hessian matrix is positive definite for any $x \in A$.
- f is strictly convex iff its Hessian matrix is negative definite for any $x \in A$.

- f is (weakly) concave iff its Hessian matrix is positive semidefinite for any $x \in A$.
- f is (weakly) convex iff its Hessian matrix is negative semidefinite for any $x \in A$.

4 Convexity and Concavity

4.1 Convex Sets

A set A, in a real vector space V, is convex iff:

$$\lambda x_1 + (1 - \lambda)x_2 \in A$$

for any $\lambda \in [0, 1]$ and any $x_1, x_2 \in A$.

4.2 Function Concavity and Convexity

Let A be a convex set in vector space V. Consider the function $f : A \to \mathbb{R}$.

1. f is concave iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{6}$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

2. f is convex iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{7}$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

3. f is strictly concave iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{8}$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

4. f is strictly convex iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) > \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{9}$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

Note

If a function is not convex, it does not mean that it is concave. Likewise, if a function is not concave, it does not mean that it is convex.

Practice

Consider $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ where A is a convex set in a vector space. If f and g are concave functions show that:

1. $f + g$ is a concave function.
2. cf is a concave function if $c > 0$, and a convex function if $c < 0$.

4.3 Jensen's Inequality

Let the function $f : A \Rightarrow \mathbb{R}$ where A is a convex set in a vector space, then:

- f is concave iff

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq \sum_{i=1}^{n} \lambda_i f(x_i)$$

- f is convex iff

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i)$$

  for any $\lambda_1, ..., \lambda_n \in \mathbb{R}_+$ such that $\sum_{i=1}^{n} \lambda_i = 1$ and $x_1, ..., x_n \in A$

4.4 Quasiconcave and Quasiconvex

Let A be a convex set in vector space V. Consider the function $f : A \to \mathbb{R}$.

1. f is quasiconcave iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \max\{f(x_1), f(x_2)\} \quad (10)$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

2. f is quasiconvex iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) \geq \max\{f(x_1), f(x_2)\} \quad (11)$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

3. f is strictly quasiconcave iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) < \max\{f(x_1), f(x_2)\} \quad (12)$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

4. f is strictly quasiconvex iff:

$$f(\lambda x_1 + (1 - \lambda)x_2) > \max\{f(x_1), f(x_2)\} \quad (13)$$

   for any $x_1, x_2 \in A$ and $\lambda \in [0, 1]$.

Practice

1. Show that if a function f is concave, then f is also quasiconcave.
2. Show that if a function f is convex, then f is also quasiconvex.

4.5 Contour Sets

Let A be a convex set in vector space V. Consider the function $f : A \to \mathbb{R}$. An upper contour set of $a \in A$ is defined as:

$$\{x \in A : f(x) \geq a\}$$

A lower contour set of $a \in A$ is defined similarly:

$$\{x \in A : f(x) \leq a\}$$

Let A be a convex set in vector space V. Consider the function $f : A \to \mathbb{R}$. Then,

1. f is quasiconcave iff its upper contour set is convex for any $a \in \mathbb{R}$
2. f is quasiconvex iff its lower contour set is convex for any $a \in \mathbb{R}$

4.6 Graphs

Let the function $f : A \to \mathbb{R}$. The graph of f is defined as the following set:

$$G(f) = \{(x, y) \in A \times \mathbb{R} : y = f(x)\}$$

The epigraph is the set above the graph, and is defined as:

$$G^+(f) = \{(x, y) \in A \times \mathbb{R} : y \geq f(x)\}$$

The epigraph is the set below the graph, and is defined as:

$$G^-(f) = \{(x, y) \in A \times \mathbb{R} : y \leq f(x)\}$$

The following theorem follows:

1. $G^-(f)$ is a convex set iff f is convex.
2. $G^+(f)$ is a convex set iff f is concave.

5 Multivariate Calculus

5.1 Derivatives

Let $f(x)$ and $g(x)$ be differentiable functions, and $a, n \in \mathbb{R}$. Derivatives have following properties:

1. $(af)' = af'(x)$
2. $(f + g) = f'(x) + g'(x)$
3. $(fg)' = f'g + fg'$
4. $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$
5. $\frac{d}{dx}(c) = 0$
6. $\frac{d}{dx}(f(g(x)) = f'(g(x))g'(x)$

5.2 Integrals

Integrals have the following properties:

1. $\int af(x)dx = a \int f(x)dx$
2. $\int (f + g)dx = \int f(x)dx + \int g(x)dx$

5.3 Integration by Parts

We can use integration by parts to integrate some more complex expressions. The formula for integration by parts is:

$$\int u(x) \cdot v'(x)dx = u(x) \cdot v(x) - \int u'(x) \cdot v(x)dx$$

Example

Using integration by parts, we can integrate the expression xe^x:

Let $u(x) = x$, and $v'(x) = e^{2x}$. Thus $u'(x) = 1$ and $v(x) = \frac{1}{2}e^{2x}$. Using the integration by parts, we see that:

$$\begin{aligned} \int xe^{2x}dx &= x\frac{1}{2}e^{2x} - \int 1 \cdot \frac{1}{2}e^{2x}dx \\ &= \frac{1}{2}\left(xe^{2x} - \int e^{2x}dx\right) \\ &= \frac{1}{2}xe^{2x} - \frac{1}{4}e^{2x} + C \end{aligned}$$

where $C \in \mathbb{R}$.

5.4 Chain Rule

Let $w = f(x, y)$ where f is a differentiable function of x and y. Let $x = g(t)$ and $y = h(t)$ where g and h are differentiable functions of t. Then by the chain rule:

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt}$$

Example

Let $w = x^3y^2 - x^2$ and $x = e^t$ and $y = cos(t)$.

$$\begin{aligned} \frac{dw}{dt} &= \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} \\ &= \left(3x^2y^2 - 2x\right)\left(e^t\right) + \left(2x^3y\right)\left(-sin(t)\right) \\ &= \left(3e^{2t}cos^2(t) - 2e^t\right)\left(e^t\right) - \left(2e^{3t}cos(t)\right)\left(sin(t)\right) \end{aligned}$$

5.5 Total Differential

Recall that when we take a partial derivative, we measure a variable's direct effect on a function (as we keep all other variables constant). If we also want to take into account a variable's indirect effect on a function (i.e. the effect that it has on other variables, which in turn affect the function), then we need to take a total differential.

Consider $z = f(x, y)$. The total differential of z is given by:

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

Example

Find the total differential for: $z = 2x\sin(y) - 3x^2y^2$.

$$\begin{aligned} dz &= \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy \\ &= \left(2\sin(y) - 3x^2y^2\right)dx + \left(2x\cos(y) - 6x^2y\right)dy \end{aligned}$$

5.6 Implicit Differentiation

Consider the equation $F(x, y) = 0$ where y is defined implicitly as a differentiable function of x. Then,

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}$$

Example

Consider $xy^2 + x^3y + 5y - 4 = 0$. Find $\frac{dy}{dx}$:

$$\begin{aligned} \frac{dy}{dx} &= -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}} \\ &= -\frac{y^2 + 3x^2y}{2xy + x^3 + 5} \\ &= \frac{-y^2 - 3x^2y}{2xy + x^3 + 5} \end{aligned}$$

Practice

Use the chain rule to derive the implicit differentiation problem above.

5.7 Taylor Polynomial

If f is differentiable of order $n + 1$ on interval I, then there exists z between points x and c, which are on in the interval I, such that:

$$f(x) = f(c) + f'(c)(x - c) + \frac{f''(c)}{2!}(x - c)^2 + \frac{f'''(c)}{3!}(x - c)^3 + \ldots + \frac{f^n(c)}{n!}(x - c)^n + R_n(c)$$

where $R_n(c) = \frac{f^{n+1}(z)}{(n+1)!}(x - c)^{n+1}$.

$R_n(c)$ is commonly referred to as the remainder or error. There are many uses of the Taylor polynomial. One use is to approximate the value of a function at a certain point, x, given that you know the value of the function at a close point, c. The higher the degree of polynomial we use, the closer we will get the the actual value of $f(x)$. You will notice that in each equation below, I have left out the remainder term, thus, we get an approximate value for $f(x)$

First order Taylor polynomial: $f(x) \approx f(c) + f'(c)(x - c)$
Second order Taylor polynomial: $f(x) \approx f(c) + f'(c)(x - c) + \frac{f''(c)}{2!}(x - c)^2$
Third order Taylor polynomial: $f(x) \approx f(c) + f'(c)(x - c) + \frac{f''(c)}{2!}(x - c)^2 + \frac{f'''(c)}{3!}(x - c)^3$

Practice

Given a function is strictly concave, and $x > c$ (i.e. we are given $f(c)$ and approximating $f(x)$), show that the approximate value for $f(x)$ using a first order Taylor polynomial is greater than the actual value of $f(x)$.