Calculus and optimization

These notes essentially correspond to mathematical appendix 2 in the text.

1 Functions of a single variable

Now that we have defined functions we turn our attention to calculus. A function f is differentiable if it is continuous and "smooth" with no breaks or kinks. A derivative, $f'(x)$, gives the slope or instantaneous rate of change in $f(x)$. If $f'(x)$ is differentiable, then we can find the second derivative, $f''(x)$, too. The second derivative is related to the curvature of the original function as it tells how the first derivative is changing. Table 1 provides some common rules of differentiation.

We now state some equivalency relations for concave functions.

Theorem 1 *Let D be a nondegenerate interval of real numbers on which f is twice continuously differentiable. The following statements are equivalent:*

1. *f is concave*
2. *$f''(x) \leq 0$ for all $x \in D$*
3. *For all $x^0 \in D : f(x) \leq f\left(x^0\right) + f'\left(x^0\right)\left(x - x^0\right)$*
4. *If $f''(x) < 0$ for all $x \in D$, then f is strictly concave.*

From our prior theorem relating convex and concave functions we also have a theorem for convex functions by reversing the inequalities and changing the word "concave" to "convex".

2 Functions of several variables

Many times we will work with functions of several variables. In consumer theory firms have utility functions which are functions of the quantities consumed of various goods; in producer theory firms have cost functions which are functions of the quantities of inputs used to produce a good. Since there are multiple variables, it is sometimes easier to think in terms of the slope of one particular variable, holding the other variables constant.

Definition 2 *Let $y = f(x_1, ..., x_n)$. The partial derivative of f with respect to x_i is defined as:*

$$\frac{\partial f(x)}{\partial x_i} \equiv \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h} \tag{1}$$

Constants, α	$\frac{d}{dx}(\alpha) = 0$
Sums	$\frac{d}{dx}(f(x) \pm g(x)) = f'(x) \pm g'(x)$
Power rule	$\frac{d}{dx}(\alpha x^n) = n\alpha x^{n-1}$
Product rule	$\frac{d}{dx}(f(x)g(x)) = f(x)g'(x) + f'(x)g(x)$
Quotient rule	$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}$
Chain rule	$\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$
ln rule	$\frac{d(\alpha \ln x)}{dx} = \frac{\alpha}{x}$

Table 1: Common rules of differentiation

The first thing to notice is that there are n partial derivatives. In essence, to find the n^{th} partial derivative of a particular function simply treat the other $n-1$ variables as constants. If $f(x_1, x_2) = x_1^2 + 3x_1x_2 - x_2^2$, then we have:

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 2x_1 + 3x_2 \\ \frac{\partial f}{\partial x_2} &= 3x_1 - 2x_2 \end{aligned} \tag{2}$$

If we collect the n partial derivatives into a row vector, then this will be the *gradient* of the function. So for our example the gradient, defined as $\nabla f(x)$ is:

$$\left[\begin{array}{cc} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{array}\right] = \left[\begin{array}{cc} 2x_1 + 3x_2 & 3x_1 - 2x_2 \end{array}\right] \tag{3}$$

We can also consider the partial derivatives of the gradient, as the second partials of the original function should provide information about the curvature of the original function. Note that if one takes the second partial derivatives of a row vector of length n then the result will be an $n \times n$ matrix of second partials. This matrix of second partial derivatives is called the Hessian matrix. Considering our example of $f(x_1, x_2) = x_1^2 + 3x_1x_2 - x_2^2$, the Hessian matrix, $H(x)$, is:

$$H(x) = \left[\begin{array}{cc} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{array}\right] = \left[\begin{array}{cc} 2 & 3 \\ 3 & -2 \end{array}\right] \tag{4}$$

A theorem which we will be able to invoke later is Young's Theorem, which basically states that the order in which the partial derivatives are taken does not affect the resulting second partial derivative:

Theorem 3 *(Young's Theorem) For any twice differentiable function $f(x)$*

$$\frac{\partial f^2}{\partial x_i \partial x_j} = \frac{\partial f^2}{\partial x_j \partial x_i} \tag{5}$$

We can see in our example that $\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 3$.

While we will not get into the details of curvature it should be helpful to state some results:

Theorem 4 *Let D be a convex subset of $\mathbb{R}^n$ with a nonempty interior on which f is twice continuously differentiable. The following statements are equivalent:*

1. *f is concave*
2. *$H(x)$ is negative semidefinite for all x in D.*
3. *For all $x^0 \in D : f(x) \leq f\left(x^0\right) + \nabla f\left(x^0\right)\left(x - x^0\right)$*

2.1 Homogeneous functions

A homogeneous function are real-valued functions which behave in particular ways as all variables are increased simultaneously and in the same proportion.

Definition 5 *A real-valued function is called homogeneous of degree k if*

$$f(tx) \equiv t^k f(x) \tag{6}$$

If a function is homogeneous of degree 0, then increasing the variables in the same proportion will leave the value of the function unchanged. If the function is homogeneous of degree 1, then increasing the variables in the same proportion will increase the value of the function in the same proportion (if all the variables are doubled, the value of the function doubles, etc.).

3 Optimization

Suppose we have a single variable function $f(x)$. Assume that $f(x)$ is differentiable. The function $f(x)$ achieves a local maximum at x^* if $f(x^*) \geq f(x)$ for all x in some neighborhood of x^*. The function $f(x)$ achieves a global maximum at x^* if $f(x^*) \geq f(x)$ for all x in the domain of the function. The local maximum is unique if $f(x^*) > f(x)$ for all x in some neighborhood of x^* and the global maximum is unique if $f(x^*) > f(x)$ for all x in the domain of the function.

Similarly, the function achieves a local minimum at $\widetilde{x}$ if $f(\widetilde{x}) \geq f(x)$ for all x in some neighborhood of $\widetilde{x}$ and a global mininum at $\widetilde{x}$ if $f(\widetilde{x}) \geq f(x)$ for all x in the domain of the function. If the inequalities are strict then the minimum is unique. What follows are the first-order necessary conditions and second-order necessary conditions for the optima of a general twice continuously differentiable function of one variable:

Theorem 6 *Let $f(x)$ be a twice continuously differentiable function of one variable. Then $f(x)$ reaches a local interior:*

1. *maximum at x^** $\Longrightarrow f'(x^*) = 0$
   $\Longrightarrow f''(x^*) \leq 0$
2. *minimum at $\widetilde{x}$* $\Longrightarrow f'(x^*) = 0$
   $\Longrightarrow f''(\widetilde{x}) \geq 0$

3.1 Optima for real-valued functions of multiple variables

Suppose that $D \subset \mathbb{R}^n$ and let $f : D \to R$ be a twice continuously differentiable real-valued function of n variables. Intuitively, a local maximum is achieved at x^* if no small movement away from x^* results in the function increasing. The first-order necessary conditions for a local interior optima at x^* are as follows:

Theorem 7 *If the differentiable function $f(x)$ reaches a local interior maximum or minimum at x^* then x^* solves the system of simultaneous equations:*

$$\begin{array}{c} \frac{\partial f(x^*)}{\partial x_1} = 0 \\ \frac{\partial f(x^*)}{\partial x_2} = 0 \\ \ldots \\ \frac{\partial f(x^*)}{\partial x_n} = 0 \end{array} \tag{7}$$

The second-order necessary conditions are as follows:

Theorem 8 *Let $f(x)$ be twice continuously differentiable.*

1. *If $f(x)$ reaches a local interior maximum at x^*, then $H(x^*)$ is negative semidefinite.*
2. *If $f(x)$ reaches a local interior minimum at $\widetilde{x}$, then $H(\widetilde{x})$ is positive semidefinite.*

Again, we will not get into the details of negative and positive semidefiniteness, but the function will be strictly concave if the principal minors of the Hessian matrix alternate in sign, beginning with a negative sign. It will be strictly convex if the principal minors of the Hessian matrix are all positive.

3.2 Constrained optimization

Thus far we have focused on unconstrained optimization. However, in many problems there are constraints which must be met. They may be equality constraints, such that we are optimizing a particular function $f(x^1, x^2)$ subject to the equality constraint $g(x^1, x^2) = 0$. They may be constraints as simple as nonnegativity constraints, so that $x^1 \geq 0$ and $x^2 \geq 0$. They may be more complex inequality constraints, such that we are optimizing the function $f(x^1, x^2)$ subject to $g(x^1, x^2) \geq 0$. The function $g(x^1, x^2)$ may be linear or nonlinear.

3.2.1 Equality constraints

Formally, an optimization with an equality constraint is set up as:

$$\max_{x_1,x_2} \; f(x_1,x_2) \;\text{ s.t. }\; g(x_1,x_2)=0 \tag{8}$$

The function which we are optimizing, f in this instance, is called the objective function. The variables that are being chosen, x_1 and x_2 in this problem, are the choice variables. The function $g(x_1,x_2)$ is called the constraint function.

With equality constrained optimization problems the problem can be solved by substitution. If we solve for one of the variables in the constraint function, say x_2, we would have a new function:

$$x_2=\widetilde{g}(x_1) \tag{9}$$

Substitute this directly into the objective function and the problem becomes:

$$\max_{x_1} \; f(x_1,\widetilde{g}(x_1)) \tag{10}$$

and then we are maximizing a function of a single variable. We then set the first derivative equal to zero and find the critical value for x_1. This first order condition is:

$$\frac{\partial f(x_1^*,\widetilde{g}(x_1^*))}{\partial x_1}+\frac{\partial f(x_1^*,\widetilde{g}(x_1^*))}{\partial x_2}\frac{d\widetilde{g}(x_1^*)}{dx_1}=0 \tag{11}$$

Once x_1^* is known we find x_2^* by using $x_2^*=\widetilde{g}(x_1^*)$. For simple problems this substitution method works well; when the constraint functions are complex or there are multiple choice variables (or constraints) this method becomes tedious.

3.2.2 Lagrange's method

Solving unconstrained optimization problems is (relatively) easy. The idea that Lagrange had was to turn constrained optimization problems into unconstrained optimization problems. Our problem from before is:

$$\max_{x_1,x_2} \; f(x_1,x_2) \;\text{ s.t. }\; g(x_1,x_2)=0 \tag{12}$$

Now, multiply the constraint by the variable λ (why? that comes later). Add that product to the objective function and we have created a new function called the Lagrangian function (or Lagrangian). The Lagrangian is:

$$\mathcal{L}(x_1,x_2,\lambda)=f(x_1,x_2)+\lambda g(x_1,x_2) \tag{13}$$

The first-order necessary conditions for optimizing the Lagrangian are the set of partial derivatives set equal to zero:

$$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial f(x_1^*,x_2^*)}{\partial x_1}+\lambda^*\frac{\partial g(x_1^*,x_2^*)}{\partial x_1}=0 \tag{14}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = \frac{\partial f(x_1^*,x_2^*)}{\partial x_2}+\lambda^*\frac{\partial g(x_1^*,x_2^*)}{\partial x_2}=0 \tag{15}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = g(x_1^*,x_2^*)=0 \tag{16}$$

The idea of Lagrange's method is that if we solve these three equations simultaneously for x_1^*, x_2^*, and λ^* we will have found a critical point of $f(x_1,x_2)$ along the constraint $g(x_1,x_2)=0$. While we will not go through all of the details to prove why this method works, the book provides a nice discussion. Also, as of now we do not know whether or not these critical values are maxima or minima, but some comments on this topic will be made shortly.

Note that Lagrange's method can be used for any number of variables (n) and any number of constraints (m) so long as the number of constraints is less than the number of variables ($m<n$). There is still the question of whether or not a solution actually exists for a problem and whether or not the λ variable exists. While these are important questions a general discussion of these concepts is more than is needed here. When we discuss particular economic optimization problems we will discuss these concepts in slightly more detail. The formal statement of Lagrange's theorem is here (note that Λ^* is a vector of the λ variables):

Theorem 9 *Let $f(x)$ and $g^j(x)$, $j = 1, ..., m$, be continuously differentiable real-valued functions over some domain $D \subset \mathbb{R}^n$. Let x^* be an interior point of D and suppose that x^* is an optimum (maximum or minimum) of f subject to the constraints, $g^j(x^*) = 0$, $j = 1, ..., m$. If the gradient vectors $\nabla g^j(x^*)$, $j = 1, ..., m$ are linearly independent, then there exist m unique numbers λ_j^*, $j = 1, ..., m$ such that*

$$\frac{\partial \mathcal{L}(x^*, \Lambda^*)}{\partial x_i} = \frac{\partial f(x^*)}{\partial x_i} + \sum_{j=1}^{m} \lambda_j^* \frac{\partial g^j(x^*)}{\partial x_i} = 0 \text{ for } i = 1, ..., n \tag{17}$$

So this theorem guarantees us the existence of the λ variables.

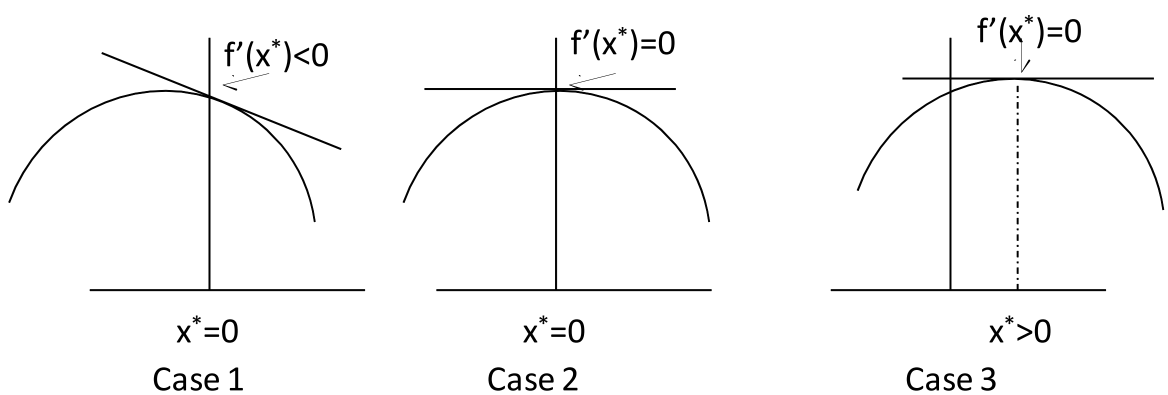
As for determining whether or not the constrained function is attaining a maximum or minimum, one can check the determinant of the bordered Hessian of the Lagrangian function. The bordered Hessian is simply the Hessian matrix of the Lagrangian function bordered by the first-order partial derivatives of the constraint equation and a zero. The bordered Hessian for a problem with two choice variables and one constraint is:

$$\overline{H} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & g_1 \\ \mathcal{L}_{21} & \mathcal{L}_{22} & g_2 \\ g_1 & g_2 & 0 \end{bmatrix}$$

where $\mathcal{L}_{ij} = \frac{\partial^2 \mathcal{L}}{\partial x_i \partial x_j}$ and $g_i = \frac{\partial g}{\partial x_i}$. As a practical matter in this course we will not be concerned with determining the sign of the determinant of the bordered Hessian matrix, but you are encouraged to read the text for more detail.

3.2.3 Inequality constraints

In some problems we will have inequality constraints to contend with. Many times we assume that the choice variables must be nonnegative, or $x_i \geq 0$. Look at the following pictures and attempt to characterize the solution to the problem:



In case 1 we have $x^* = 0$ and $f'(x^*) < 0$.
In case 2 we have $x^* = 0$ and $f'(x^*) = 0$.
In case 3 we have $x^* > 0$ and $f'(x^*) = 0$.

In all three of these cases we have $x^*[f'(x^*)] = 0$. But this is not the only condition because in case 3 when $\widetilde{x} = 0$ we have $\widetilde{x}[f'(\widetilde{x})] = 0$ but $\widetilde{x}$ is NOT a maximum. So the necessary conditions we need to have hold are:

$$f'(x^*) \leq 0 \tag{18}$$
$$x^*[f'(x^*)] = 0 \tag{19}$$
$$x^* \geq 0 \tag{20}$$

The necessary conditions for a minimum are fairly similar to those of a maximum. We have:

$$f'(x^*) \geq 0 \tag{21}$$
$$x^*[f'(x^*)] = 0 \tag{22}$$
$$x^* \geq 0 \tag{23}$$

3.2.4 Kuhn-Tucker conditions

A general problem that we might encounter is:

$$\max_{x_1,x_2} f(x_1,x_2) \text{ s.t. } g(x_1,x_2) \geq 0 \tag{24}$$

This is a nonlinear programming problem, as a linear programming problem we have a linear function which is optimized subject to linear equality constraints. We make no such restrictions here. We can construct the Lagrangian to find:

$$\mathcal{L}(x_1,x_2,\lambda) = f(x_1,x_2) + \lambda[g(x_1,x_2)]. \tag{25}$$

The Kuhn-Tucker necessary conditions for this problem are:

$$\frac{\partial f}{\partial x_1} + \lambda\frac{\partial g}{\partial x_1} = 0 \tag{26}$$
$$\frac{\partial f}{\partial x_2} + \lambda\frac{\partial g}{\partial x_2} = 0 \tag{27}$$
$$\lambda g(x_1,x_2) = 0 \tag{28}$$
$$\lambda \geq 0, g(x_1,x_2) \geq 0 \tag{29}$$

Now, there are many times in which we might restrict our choice variables x_1 and x_2 to be greater than or equal to 0. In those instances, our problem would be:

$$\max_{x_1\geq 0, x_2 \geq 0} f(x_1,x_2) \text{ s.t. } g(x_1,x_2) \geq 0 \tag{30}$$

Technically, the Lagrangian is:

$$\mathcal{L}(x_1,x_2,\lambda_1,\lambda_2,\lambda_3) = f(x_1,x_2) + \lambda_1[g(x_1,x_2)] + \lambda_2 x_1 + \lambda_3 x_2 \tag{31}$$

The Kuhn-Tucker necessary conditions for this type of problem are:

$$\begin{array}{lll} \frac{\partial\mathcal{L}}{\partial x_1} \leq 0, & x_1 \geq 0, & x_1 * \frac{\partial\mathcal{L}}{\partial x_1} = 0 \\ \frac{\partial\mathcal{L}}{\partial x_2} \leq 0, & x_2 \geq 0, & x_2 * \frac{\partial\mathcal{L}}{\partial x_2} = 0 \\ \frac{\partial\mathcal{L}}{\partial \lambda_1} \geq 0, & \lambda_1 \geq 0, & \lambda_1 * \frac{\partial\mathcal{L}}{\partial \lambda_1} = 0 \\ \frac{\partial\mathcal{L}}{\partial \lambda_2} \geq 0, & \lambda_2 \geq 0, & \lambda_2 * \frac{\partial\mathcal{L}}{\partial \lambda_2} = 0 \\ \frac{\partial\mathcal{L}}{\partial \lambda_3} \geq 0, & \lambda_3 \geq 0, & \lambda_3 * \frac{\partial\mathcal{L}}{\partial \lambda_3} = 0 \end{array}.$$

Technically one would have to check all these conditions to see which constraints are binding and which are not. For most consumer and firm problems that we will solve we will be able to "know" that $x_1 > 0$ and $x_2 > 0$ by looking at the objective function.