14.102, Math for Economists
Fall 2005
Lecture Notes, 10/13/2005

These notes are primarily based on those written by Andrei Bremzen for 14.102 in 2002/3, and by Marek Pycia for the MIT Math Camp in 2003/4. I have made only minor changes to the order of presentation, and added a few short examples, mostly from Rudin. The usual disclaimer applies; questions and comments are welcome.

Nathan Barczi
nab@mit.edu

10 Static Optimization I

10.1 Unconstrained Optimization.

We have already stated the first optimization result, the Bolzano-Weierstraß theorem. Remember, the only property that we assumed[9] of $f : \mathbb{R}^n \to \mathbb{R}$ was continuity and that the theorem, although asserting that a maximum exists (over a compact set), gave no clue as to how to find it. Differentiability is a stronger property than continuity; and yes, as it is very often the case in math, stronger assumptions allow us to come to stronger conclusions. We may indeed locate optima of f by looking at its derivative (or gradient).

The very definition of differentiability states that locally a differentiable function is well approximated by a linear function. But optimizing a linear function is easy: it never reaches an interior maximum or a minimum except if all its coefficients are zero. That immediately gives us the following necessary condition:

Theorem 244 *(First Order Conditions) If $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ reaches its (local) maximum at some interior point $x^* \in intZ$ (by interior we mean that x^* belongs to Z together with a small enough open ball $B_r(x^*)$, for some $r > 0$) and f is differentiable at x^* then $D_{x^*}f = 0$ (points at which $Df = 0$ are called critical points of f).*

Corollary 245 *The same result holds if x^* is instead a local minimum.*

This theorem is the theoretical ground behind the mechanical differentiation used by many college students. Three points should be made about using this theorem.

First, this theorem gives you a *necessary* condition which is by no means *sufficient*. If a function has a zero gradient at some particular interior point,

[9] We do not study oprtimization of more general functions $\mathbb{R}^n \to \mathbb{R}^m$ simply because there maximum or minimum value is not defined: remember, there is no natural way to order $\mathbb{R}^m$.

than it does not have to be (even a local) maximum or minimum (think about $f(x) = x^3$ at point 0).

Second, the theorem gives you a necessary condition only for an *interior* optimum. If a local optimum is reached by f at a point on the boundary of D, its gradient does not have to equal zero at this point (think about $f(x) = x$ on $D = [0, 1]$). You have to consider boundary points separately.

Third, the theorem tells you nothing about *global* optima. You have to employ other (not first order) considerations to figure out at which of the suspicious points (which include all critical points of f and all boundary points of D) the function actually attains a global maximum. Good news is, though, that typically there will not be too many of those.

That is all there is to say about first order condition and unconstrained optimization. Again, assuming more structure (in our case, existence of the second derivative) allows one to come to more definite conclusions.

Theorem 246 *(Second Order Conditions) Suppose f is a C^2 function on $Z \subset \mathbb{R}^n$, and x^* is an interior point of Z. If f has a local maximum (respectively, minimum) at x^*, then $D_{x^*}(f)$ is zero and $H_f(x^*)$ is negative (respectively, positive) semidefinite. Conversely, if $D_{x^*}(f)$ is zero and $H_f(x^*)$ is negative (respectively, positive) definite, then f has a strict local maximum (respectively, minimum) at x^*.*

The above theorem gives almost necessary and sufficient conditions for an interior optimum. Almost – because no conclusions can be drawn if the Hessian is semidefinite but not definite.

In the one-dimentional case, the Hessian of f is simply one number: f''. Therefore, second order conditions do not give a definite answer for points at which both the first and second derivatives are zero. A natural next move is then to consider the third derivative – whether $f'''(x^*) \neq 0$. If so, then locally the function looks like x^3 around zero, i.e., it is not an optimum. If $f'''(x^*) = 0$, then consider the fourth derivative. If $f''''(x_0) < 0$, then locally f looks like x^4, i.e., x^* is a local minimum. If $f''''(x^*) > 0$, then x^* is a local maximum. If $f''''(x^*) = 0$, you have to consider further derivatives. Unfortunately, as the last example shows, this process may never come to the end: you may be evaluating higher and higher order derivatives and they all may turn out to be zero even though the function itself is nonzero around x^*, and you will never find out whether x^* is a local maximum, a local minimum, or neither.

Besides working out quadratic forms, there is another simple algorithm for testing the definiteness of a symmetric matrix like the Hessian. First, we need some definitions:

Definition 247 *Let A be an $n \times n$ matrix. A $k \times k$ submatrix of A formed by deleting $n - k$ rows of A, and the same $n - k$ columns of A, is called **principal submatrix** of A. The determinant of a principal submatrix of A is called a **principal minor** of A.*

Note that the definition does not specify which $n - k$ rows and columns to delete, only that their indices must be the same.

Example 248 *For a general* 3×3 *matrix,*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

there is one third order principal minor, namely $|A|$. *There are three second order principal minors:*

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \text{ formed by deleting column 3 and row 3;}$$

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \text{ formed by deleting column 2 and row 2;}$$

$$\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \text{ formed by deleting column 1 and row 1}$$

And there are three first order principal minors:

$$\left| a_{11} \right|, \text{ formed by deleting the last two rows and columns}$$

$$\left| a_{22} \right|, \text{ formed by deleting the first and third rows and columns}$$

$$\left| a_{33} \right|, \text{ formed by deleting the first two rows and columns}$$

Definition 249 *Let* A *by an* $n \times n$ *matrix. The* k*th order principal submatrix of* A *obtained by deleting the last* $n - k$ *rows and columns of* A *is called the* k*th order* ***leading principal submatrix*** *of* A*, and its determinant is called the* k*th order* ***leading principal minor*** *of* A.

We will denoted the kth order leading principal submatrix of A by A_k, and its kth order leading principal minor by $|A_k|$. Now, the algorithm for testing the definiteness of a symmetric matrix:

Theorem 250 *Let* A *be an* $n \times n$ *symmetric matrix. Then,*

1. (a) *A is positive definite if and only if all its n leading principal minors are (strictly) positive.*

   (b) *A is negative definite if and only if its n leading princiapl minors alternate in sign as follows:*

   $$|A_1| < 0, \ |A_2| > 0, \ |A_3| < 0, \text{ etc.}$$

   (c) *If some kth order leading principal minor of A is nonzero but does not fit either of the above sign patterns, then A is indefinite.*

One particular failure of this algorithm occurs when some leading principal minor is zero, but the others fit one of the patterns above. In this case, the matrix is not definite, but may or may not be semidefinite. In this case, we must unfortunately check not only the principal leading minors, but *every* principal minor.

Theorem 251 *Let A be an $n \times n$ symmetric matrix. Then, A is positive semidefinite if and only if every principal minor of A is ≥ 0. A is negative semidefinite if and only if every principal minor of odd order is ≤ 0 and every principal minor of even order is ≥ 0.*

10.2 Equality Constrained Optimization

Suppose we want to maximize (or minimize) a smooth function $f(x) : \mathbb{R}^2 \to \mathbb{R}$ on the set $\{x \in \mathbb{R}^2 : g(x) = 0\}$, where $g(x) : \mathbb{R}^2 \to \mathbb{R}$ is another smooth function. By the implicit function theorem, around any point x^* at which $\nabla g(x^*) \neq 0$ this set looks like a smooth curve and from x^* we can move along it in either direction. What would be the first order condition for x^* if it is a solution to the optimization problem? Well, if $\nabla f(x^*)$ is not orthogonal to the curve at x^*, then x^* is clearly not an optimum: of the two opposite directions at x^* one will give a strictly positive dot product with ∇f and moving in this direction will strictly (first order) improve the value of f. Therefore, at the optimum it is necessary that ∇f is orthogonal to the curve. But so is ∇g (at any point). Hence at the optimum it should be the case that $\nabla f \parallel \nabla g$, which implies (provided $\nabla g \neq 0$) that $\nabla f(x^*) = \lambda \nabla g(x^*)$.

Exercise 252 *Suppose $f(x), g_1(x), g_2(x) : \mathbb{R}^3 \to \mathbb{R}$ are C^1 functions and we want to maximize $f(x)$ subject to $g_1(x) = g_2(x) = 0$. Assume that at some point x^* in the feasible set $\nabla g_1(x^*) \nparallel \nabla g_2(x)$ (which means that around x^* the feasible set is a smooth curve). Persuade yourself that a necessary condition for a local optimum is that $\nabla f(x^*)$ lie in the span of $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$, i.e., that $\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*)$ for some λ_1, λ_2.*

The intuition conveyed by the two above examples is generalized in the following

Theorem 253 *(The Theorem of Lagrange) Let $f, g_i : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions, $i = 1, ..., k$.Suppose x^* is a local maximum or minimum of f on the set*

$$Z = U \cap \{x | g_i(x) = 0, i = 1, ..., k\},$$

where $U \subset \mathbb{R}^n$ is open. Suppose also that $rank[\nabla g_1(x^*), ..., \nabla g_2(x^*)] = k$. Then, there exist real numbers $\lambda_1^*, ..., \lambda_k^*$ such that $\nabla f(x^*) = \sum_{i=1}^{k} \lambda_i^* \nabla g_i(x^*)$.

Note that, similar to unconstrained first order conditions, this theorem gives only *necessary* conditions for an optimum, which are by no means sufficient. For sufficiency one must consider second order conditions, which are messy and which I am not going to consider here.

The *constraint qualification* condition (that $rank[\nabla g_1(x^*), ..., \nabla g_2(x^*)] = k$, or, equivalently, that gradients of all the constraints at x^* are linearly independent) is important, as the following example suggests: without it the theorem may fail, i.e., at a local optimum the corresponding λ_i (which are called *Lagrange multipliers*) need not exist. However, in all my time in economics I have never seen such a problem. Therefore it might be optimal timewise to forget about them and solve the problem using the Lagrange method (discussed in the next paragraph), unless you are absolute risk averse.

Example 254 *Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = -y$ and $g(x, y) = y^3 - x^2$. Show that the maximum of f subject to $g = 0$ is attained at the origin, but that the constraint qualification is violated there. Show that the conclusion of the theorem fails and the required λ does not exist.*

Lagrange's main contribution to the study of constrained optimization was to associate a function L (called the *Lagrangean*) with an equality-constrained optimization problem, in such a way that the problem is then reduced to the unconstrained optimization of L. Following Lagrange, let us define

$$L(x_1, ..., x_n, \lambda_1, ..., \lambda_n) = f(x_1, ..., x_n) - \sum_{i=1}^{k} \lambda_i g_i(x_1, ..., x_n).$$

Exercise 255 *Deduce from Theorem of Lagrange that if $x^* \in \mathbb{R}^n$ is a (local) optimum of f subject to $g_i = 0, i = 1, ..., k$, then there exists $\lambda^* \in \mathbb{R}^k$ such that (x^*, λ^*) is a critical point of $L(x, \lambda)$.*

10.2.1 Level Curves and the Theorem of Lagrange

Recall the geometric interpretation of the Implicit Function Theorem: a level curve $F(x, y) = c$ defines a curve in the plane, and the theorem gives conditions under which we can think of this curve as defining y as a function of x. The following restatement of the theorem formalizes this intuition:

Theorem 256 *Let (x_0, y_0) be a point on the locus of points $F(x, y) = c$ in the plane, where F is a C^1 function of two variables. If $(\partial F/\partial y)(x_0, y_0) \neq 0$, then $F(x, y) = c$ defines a smooth curve around (x_0, y_0) which can be thought of as the graph of a C^1 function $y = f(x)$. Furthermore, the slope of this curve is*

$$-\frac{\frac{\partial F}{\partial x}(x_0, y_0)}{\frac{\partial F}{\partial y}(x_0, y_0)}. \tag{19}$$

If $(\partial F/\partial y)(x_0, y_0) = 0$, but $(\partial F/\partial x)(x_0, y_0) \neq 0$, then the Implicit Function Theorem tells us that the locus of points $F(x, y) = c$ is a smooth curve about (x_0, y_0) which we can consider as defining x as a function of y. It also tells us that the tangent line to the curve at (x_0, y_0) is parallel to the y-axis, i.e., vertical.

If either $(\partial F/\partial y)(x_0, y_0) \neq 0$ or $(\partial F/\partial x)(x_0, y_0) \neq 0$ holds, we call(x_0, y_0) a **regular point** of the function $F(x, y)$; the theorem tells us that if every point on a particular level set is regular, then that level set defines y as a function of x (or x as a function of y) everywhere on the curve, and that there is a well-defined tangent line to each point on the curve.