Calculus

Daniel N. Hauser

This note summarizes some of the key concepts from the second set of lectures.

Derivatives

We have a class of functions that are easy to work with and we understand very well (linear functions) and another larger class of functions that are at least in principle "well-behaved" (continuous functions). Can we leverage the tractability of the former to help us analyze the latter? The *derivative*, when it exists, gives us a tool that does exactly that. The *gradient* of a function, denoted

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)^T$$

linearizes f at x, in the sense that for any $\hat{x}$

$$f(\hat{x}) = f(x) + \nabla f(x) \cdot (\hat{x} - x) + o(\hat{x} - x).$$

So for values close to x, there is a linear function that approximates f that has "slope" $\nabla f(x)$.[1] [2] For a twice differentiable $f : \mathbb{R}^m \to \mathbb{R}$ we can write the *second derivative matrix/hessian*

$$D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_m \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_m \partial x_m} \end{pmatrix}.$$

For most functions this matrix is symmetric.[3] We can use this to construct even better approximations using the *Taylor expansion*, for any twice continuously differentiable function

$$f(\hat{x}) = f(x) + (\hat{x} - x)^T \nabla f(x) + \frac{1}{2}(\hat{x} - x)^T D^2 f(x)(\hat{x} - x) + o((\hat{x} - x)^2).$$

Economic "applications" of derivatives (beyond finding maxes)

1. Comparative statics: How do endogenous choices respond to small changes in exogenous parameters. How much does a small change in price change demand?
2. Dynamics: Allows us to describe dynamics. How does an economy change over time? Is an economy going to approach a specific steady state?

[1] Similarly the Taylor expansion let's us approximate a continuous function with a polynomials by using higher order derivative. The higher the degree of the polynomial approximation, the faster the approximation error vanishes.

[2] For a function $f : \mathbb{R}^m \to \mathbb{R}^n$, we denote the derivative *matrix* as

$$Df(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_m} \end{pmatrix}$$

Note that for a function that maps to $\mathbb{R}$, $Df(x)^T = \nabla f(x)$. In the next section, I use D_x to denote the derivative matrix with respect to x (which may be a vector).

[3] Young's theorem.

Unconstrained Optimization

The derivative gives us a tool to find local maximizers and minimizers.

Theorem 1 (First order conditions)**.** *The gradient at an interior local maximum is* 0*.*

So, for any maximum x^* that occurs on the interior of the domain, a necessary condition is that $\nabla f(x^*) = 0$. This provides us with a set of candidate maximizers, called *critical points*.

First order conditions are necessary, but not sufficient for a maximizer. [4] The second derivative gives us a second test to help narrow our search:

[4] Clearly, this must also hold at minimizers. This can also hold at points that are neither, e.g. $x = 0$ for the function $f(x) = x^3$.

Theorem 2 (Second order conditions)**.** *At any critical point x^* of f if $D^2 f(x^*)$ is negative definite then x^* is a local maximum. If x^* is a local maximum then $D^2 f(x^*)$ is negative semidefinite.*

This gives us a good way to look for maximizers of a function f if it is twice differentiable.

1. First, find all points where $\nabla f(x) = 0$.
2. Then look at $D^2 f$ at those points:
   (a) If $D^2 f(x)$ is not negative semidefinite, it's not a max.
   (b) If $D^2 f(x)$ is negative definite, it is a (local) max.
   (c) If $D^2 f(x)$ is negative semidefinite, it could be a local max.
3. To find the global max, compare the values of f at the points you found in step 2b, c and the points on the boundary/where the function wasn't differentiable.

The Implicit Function Theorem

Consider functions $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$. We are often going be confronted with problems where for any given vector of exogenous variables $x \in \mathbb{R}^m$, we know the corresponding endogenous variables $y \in \mathbb{R}^n$ solve an equation of the form[5]

$$f(y, x) = 0.$$

[5] For instance, the first order conditions of the consumer problem tell us that if there are two goods x_1, x_2 being sold for prices p_1, p_1 then at the optimum the consumer sets the marginal rate of substitution equal to the ratio of the prices, e.g.

$$\frac{u_1(x_1, x_2)}{u_2(x_1, x_2)} - \frac{p_1}{p_2} = 0.$$

The prices are exogenous, while the amount of goods 1 and 2 consumed are endogenous.

A natural question to ask would be "How does y change as x varies?" If there was a nice, differentiable $y(x)$ that solved this equation at each x, then we'd know exactly what to do. Using the chain rule

$$\begin{aligned} D_x f(y(x), x) &= D_x 0 \\ D_y f Dy + D_x f &= 0 \\ Dy &= -(D_y f)^{-1} D_x f. \end{aligned}$$

But, for most problems it's not clear that such a y exists, much less is differentiable. The implicit function theorem tells us when we can do something like this:

Theorem 3 (Implicit Function Theorem)**.** *Let (x^*, y^*) solve $f(y^*, x^*) = 0$. Then if $D_y f(y^*, x^*)$ has full-rank, in a neighborhood of x^* there exists a unique differentiable y s.t. $f(y(x), x) = 0$ and $y(x^*) = y^*$. Moreover, in this neighborhood*

$$Dy = -(D_y f)^{-1} D_x f.$$

Optimization with Equality Constraints

Consider a maximization problem of the form

$$\max_{x \in \mathbb{R}^m} f(x)$$
$$\text{s.t. } g(x) = 0$$

for $f : \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^m \to \mathbb{R}^n$. The extreme value theorem tells us that for continuous f and nice enough g, this has a solution. It would be nice to have something like first order conditions to help us find it. In class, we used the implicit function theorem to show that at any maximum x^*, if Dg has full rank then

Maximizing utility subject to spending your entire budget

$$\max_{x \in \mathbb{R}^m} u(x)$$
$$\text{s.t. } p \cdot x - m = 0$$

is probably the most familiar of these.

$$\nabla f(x^*) = \lambda^T Dg(x^*)$$

for some $\lambda \in \mathbb{R}^n$. The λ's are called Lagrange multipliers.[6] So, as with unconstrained optimization problems, we can turn the problem of finding a maximum into the problem of solving a system of (potentially non-linear) equations, with the additional headache that we've added an "extra" variable for each constraint (the λ).

[6] There are second order conditions for these, but they are annoying so we aren't going to talk about them. Appealing to the concavity of the objective and the convexity of the feasible set is mostly what you'll do in the first year. More on this in the next sets of lectures.