Math Camp - Calculus

Calculus

It's great that we know a max exists, but it would be nice if we could also solve for one.

Our existence proof isn't super useful for this, we constructed a sequence that converged to the max, but not really something we could operationalize

Calculus

So let's narrow down the set of points we need to check.

We know at a maximum a function can't be increasing in any direction.

- How can we find where a function is increasing?
- Easy to do with linear functions.

  $$y = a \cdot x + b$$

  This is strictly increasing in direction d iff $a \cdot d > 0$.
- Can we do the same for non-linear functions?

Derivatives

If $f : \mathbb{R} \to \mathbb{R}$ is linear, we can use the slope to determine whether it is increasing or decreasing.

But slope doesn't really make sense for non-linear functions. So we define the derivative:

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x}.$$

Clearly cannot exist if f is not continuous at x. May not exist even if f is continuous at x (e.g. $f(x) = |x|$). We say a function with a derivative everywhere is differentiable.

Derivatives

As I'm sure you recall from your calculus course, there are a million derivatives that had to memorize. Fortunately, there are a lot fewer that you'll see a lot in the first year. Some common ones

- $f(x) = x^k$, $f'(x) = kx^{k-1}$.
- $f(x) = e^x$, $f'(x) = e^x$
- $f(x) = \ln x$, $f'(x) = \frac{1}{x}$.
- $f(x) = k$, $f'(x) = 0$

And some helpful rules

- Linearity$\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$, $\frac{d}{dx} af(x) = af'(x)$.
- $\frac{d}{dx} - f(x) = -f'(x)$.
- Chain rule $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$.
- Product rule $\frac{d}{dx} f(x)g(x) = f'(x)g(x) + g'(x)f(x)$.
- Quotient rule: $\frac{d}{dx} f(x)/g(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$.
- Fundamental theorem of calculus: $f(a) - f(b) = \int_a^b f'(x)\, dx$

Increasing functions

Derivative lets us see if a function is increasing or decreasing

- $f(x)$ inc at $x \Rightarrow f'(x) \geq 0$.
- $f(x)$ dec at $x \Rightarrow f'(x) \leq 0$.
- $f'(x) > 0 \Rightarrow f(x)$ is strictly inc at x
- $f'(x) < 0 \Rightarrow f(x)$ is strictly dec at x.

At a maximum a function can't be strictly decreasing or increasing so if a function is differentiable.

- Maxima can only occur at interior points were $f'(x) = 0$ or points that are not interior.

Local Maxima

The derivative is a local tool. Tells us how the function approximately behaves at close by points.

Definition

$x \in X$ is a **local maximum** of $f : X \to \mathbb{R}$ if there exists an $\epsilon > 0$ s.t. for all $x' \in X$ such that $||x - x'|| < \epsilon$, $f(x) \geq f(x')$.
$x \in X$ is a **global maximum** if for all $x' \in X$, $f(x) \geq f(x')$.

Example

$$\max_{x \in [0,1]} -(x-a)^2$$

for $a \in (0,1)$. This is clearly maximized at $x = a$. The first order condition is

$$-2(x-a) = 0$$

so $x = a$ is indeed a critical point.

Example

Consider

$$\max_{x \in [0,1]} (x - a)^2$$

This has first order condition

$$2(x - a) = 0$$

so $x = a$ is the only point where $f'(x) = 0$. But, we also need to worry about the boundary. Three points to check

$$x = a,\, x = 0\, x = 1.$$

$x = a$ is local min, $x = 0$ and 1 are local maxima.

The derivative lets us approximate our function with other functions we know a lot about

Theorem (Mean value theorem)

Suppose f is differentiable. Take $a, b \in \mathbb{R}$. There exists a c s.t. $c \in (a, b)$ and

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Theorem (Taylor's theorem)

Assume f is k times differentiable at x. Then there exists some function r s.t.

$$f(a) = \sum_{i=0}^{k} f^{(k)}(x)(a-x)^k/k! + r(a)(a-x)^k$$

where $\lim_{a \to x} r(a) = 0$.

We can use the derivative to locally approximate our function with a polynomial.

Maximizers

What else can we say about a maximizer? Well let's plug a maximizer x^* into the second order taylor expansion.

Around a critical point,

$$f(x) - f(x^*) = f''(x^*)(x - x^*)^2/2 + r(x)(x - x^*)^2$$

If $f''(x^*) > 0$, then we could find a x' close enough to x s.t. $f(x') - f(x^*) > 0$, which is a contradiction.

So at a maximizer $f''(x^*) \leq 0$. Is this enough?

Multivariate Calculus

Now let's think about $f : \mathbb{R}^m \to \mathbb{R}$. The idea of a slope isn't as clean here.

Other interpretation still makes sense, the derivative is a linear function that f behaves like locally, i.e. $\exists \nabla f \in \mathbb{R}^m$ s.t

$$\lim_{y \to x} \frac{||f(y) - f(x) - \nabla f \cdot (y - x)||}{||y - x||} \to 0.$$

∇f is called the gradient.

- If $f : \mathbb{R}^m \to \mathbb{R}^n$, we replace the vector with a matrix, usually denoted Df (called the Jacobian), otherwise this definition is the same.
- Does this ∇f exist? Obviously not always.
- Is this unique?
- How do we find it?

Gradient

We can answer uniqueness and how to solve for this at the same time.

Let e_i be the vector with a 1 in the *ith*-component and 0's in all other components. If the gradient exists then

$$\nabla f(x)_i = \lim_{h \to 0} \frac{f(x + e_i \cdot h) - f(x)}{h}$$

We call the right hand side the partial derivative wrt to x_i, denoted $\frac{\partial f}{\partial x_i}$.

- So ∇f is the vector of partial derivatives.
- Even if all partial derivatives exist at x, $f(\cdot)$ may not be differentiable, or even continuous at x.
- Fortunately, as long as all partial derivatives exist in some nbhd of x, x is differentiable.

Partial Derivatives

Partial derivatives are easy to find. Just treat all other variables as-if they were constants and take the derivative.

Examples:

- $\frac{\partial}{\partial x} xy = y$
- $\frac{\partial}{\partial y} (x^2 + y^2)^{1/2} = y(x^2 + y^2)^{-1/2}$.

Higher order derivatives

Just like in $\mathbb{R}$ we can take higher order derivatives.

The Hessian matrix is

$$D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

This is a symmetric matrix if the second derivatives are continuous (Young's theorem), i.e. $\frac{\partial f}{\partial x_i x_j} = \frac{\partial f}{\partial x_j x_i}$.

Taylor Expansion

We can do a Taylor expansion just like before. The second order Taylor expansion now looks like

$$f(x) \approx f(a) + (x-a)\nabla f(a) + \frac{1}{2!}(x-a)' D^2 f(a)(x-a).$$

First-order conditions

The gradiant tells us how $f(x)$ behaves locally in any direction.

So at any maximizer, $\nabla f(x) = 0$.

Second order conditions

Once again, this is not enough. Taylor expansion gives us second order condition: At any interior local max a, for $x \neq a$ we need

$$(x-a)' D^2 f(a)(x-a) \leq 0$$

and $\nabla f(x) = 0$. If $\nabla f(x) = 0$ and

$$(x-a)' D^2 f(a)(x-a) < 0$$

then a is a local max.

Definiteness

All homogeneous, degree 2 polynomials from $\mathbb{R}^m \to \mathbb{R}$ can be written using a symmetric matrix in $\mathbb{R}^{m \times m}$.

$$x'Ax$$

This is called a quadratic form.

We say a A is positive (semi)definite if for all x

$$x'Ax > (\geq) 0$$

for all x, i.e. the first term in the polynomial is always positive. We define negative (semi)definiteness analogously.

Definiteness

So to verify second order conditions, we need to make sure the Hessian is negative definite.

If $x \in \mathbb{R}^2$, this is equivalent to checking

$$det(D^2 f) > 0$$

and

$$\frac{\partial^2 f}{\partial x_1^2} < 0.$$

(A straightforward change of basis shows us a symmetric matrix is negative definite iff all eigenvalues are negative. Since the determinant is the product of the eigenvalues, this then follows immediately)

Definiteness

We can do a similar check for larger matricies, the determinants of the leading principal minors must alternate signs, and the first must be negative.

- You probably won't have to check this to often outside of the $\mathbb{R}^2$ case in the first year :)
- A matrix is positive definite iff all the leading principle minors have positive determinants, and A is neg def iff $-A$ is pos def. This is maybe easier to remember.
- To check whether a matrix is semi-definite, it is not enough to replace the strict inequalities with weak inequalities.

Implicit Function Theorem

We now have a tool that lets us treat our non-linear functions sort of like linear ones. Can we do other things with this?

Let $p \in \mathbb{R}^n$ be a (column) vector of prices, and suppose demand

$$q = Ap + b$$

for some $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$. Similarly supply is

$$q = Cp + d$$

Implicit Function Theorem

We now have a tool that lets us treat our non-linear functions sort of like linear ones. Can we do other things with this?

Let $p \in \mathbb{R}^n$ be a (column) vector of prices, and suppose demand

$$q = Ap + b$$

for some $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$. Similarly supply is

$$q = Cp + d$$

The point where demand and supply are equal is the solution to

$$\begin{aligned} Ap + b &= Cp + d \\ p &= (A - C)^{-1}(d - b) \end{aligned}$$

Not only can we solve for the equilibrium, we can also see what happens to prices as we vary that parameters.

Implicit function theorem

More generally, for $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, suppose

$$Ax + By = 0$$

Implicit function theorem

More generally, for $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, suppose

$$Ax + By = 0$$

If we change x to $x + dx$, $dx \in \mathbb{R}^n$, what can we say about how much y must change to maintain this equality.

$$\begin{aligned} A(x + dx) + B(y + dy) &= 0 \\ -Bdy &= Ax + By + Adx \\ dy &= -B^{-1}Adx \end{aligned}$$

Works as long as B has full rank.

Implicit Function Theorem

Turns out we can do this for non-linear functions too

Example: $x^2 + y^2 = 1$. Fix initial point (x_0, y_0)

- This is the unit circle.
- Can we define a function $g(x) = y$, $x^2 + g(x)^2 = 1$ and $g(x_0) = y_0$. Is this function unique? differentiable?
- Sort of: One of $y = \pm\sqrt{1 - x^2}$ works.
- Not unique at $(x_0, y_0) = (1, 0)$. Also not differentiable there.

Clearly this also becomes pretty tough to do for more complicated functions.

Implicit function theorem

Theorem (Implicit Function Theorem)

Consider a function $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$*. Denote elements of* $\mathbb{R}^{n+m}$ *as* (x, y) *where* $x \in \mathbb{R}^n$*,* $y \in \mathbb{R}^m$ *and fix a point* (a, b) *where* $f(a, b) = 0$*. If* $D_y f(a, b)$ *has full rank, then there exists a open set* $U \subseteq \mathbb{R}^n$ *and a unique continuously differentiable* $g : U \to \mathbb{R}^m$ *such that*

$$f(x, g(x)) = 0$$

for all $x \in U$ *and* $g(a) = b$*. Finally*

$$Dg(x) = -D_y f(x, g(x))^{-1} D_x f(x, g(x))'$$

Implicit Function Theorem

To see where the differential equation comes from, consider $f : \mathbb{R}^2 \to \mathbb{R}^1$. Then if $g(\cdot)$ exists

$$\begin{aligned} f(x, g(x)) &= 0 \\ \frac{\partial}{\partial x}[f(x, g(x))] &= 0 \\ f_x(x, g(x)) + f_y(x, g(x))g'(x) &= 0 \\ g'(x) &= -\frac{f_x(x, g(x))}{f_y(x, g(x))} \end{aligned}$$

Full rank condition makes sure we aren't dividing by 0

Proving this isn't too terribly difficult, but it requires enough of a detour into the theory of ODEs that we aren't going to do.

Implicit function theorem: Examples

Applying the implicit function theorem to the example, we get:

$$D_y f(x, y) = 2y$$
$$D_x f(x, y) = 2x$$
$$\text{So } g'(x) = -x/g(x)$$

For an initial condition (x_0, y_0), $y_0 \neq 0$ the unique solution to this differential equation is exactly the equation we found 2 slides ago.

Implicit function theorem: Examples

Applying the implicit function theorem to the example, we get:

$$D_y f(x,y) = 2y$$
$$D_x f(x,y) = 2x$$
$$\text{So } g'(x) = -x/g(x)$$

For an initial condition (x_0, y_0), $y_0 \neq 0$ the unique solution to this differential equation is exactly the equation we found 2 slides ago.

Just eyeballing this we see that at x_0, y_0 we see

$$\frac{dy}{dx} = \frac{-x_0}{y_0}$$

Gives us the"slope" of the circle

Implicit Function Theorem: Example

We can also apply this to more abstract things.

Suppose $f(x)$ is differentiable. What can we say about $f(x) = y$.

- Fix some x_0, $y_0 = f(x_0)$
- IFT tells us when we can construct a unique (local) inverse function, $f(g(y)) = y$ if $f'(x_0) \neq 0$

Implicit Function Theorem: Example

We can also apply this to more abstract things.

Suppose $f(x)$ is differentiable. What can we say about $f(x) = y$.

- Fix some x_0, $y_0 = f(x_0)$
- IFT tells us when we can construct a unique (local) inverse function, $f(g(y)) = y$ if $f'(x_0) \neq 0$
- Also tells us the derivative is

$$g'(y) = \frac{1}{f'(g(y))}$$

Implicit Function Theorem: Example

We can also apply this to more abstract things.

Suppose $f(x)$ is differentiable. What can we say about $f(x) = y$.

- Fix some x_0, $y_0 = f(x_0)$
- IFT tells us when we can construct a unique (local) inverse function, $f(g(y)) = y$ if $f'(x_0) \neq 0$
- Also tells us the derivative is

$$g'(y) = \frac{1}{f'(g(y))}$$

This is the inverse function theorem. If $f(\cdot)$ is an invertible function with non-zero derivative then,

$$f^{-1'}(y) = \frac{1}{f'(f^{-1}(y))}$$

Lagrange Multipliers

Now let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$

$$\max f(x)$$
$$\text{s.t. } g(x) = 0$$

Under some simple conditions, if x^* solves this, then there exists a (column vector) $\lambda \in \mathbb{R}^m$ s.t.

$$Df = \lambda' Dg$$

If $g(\cdot)$ was linear, we could just plug the constraints into the objective function and solve that way.

Example

Consumer problem:

$$\max xy$$
$$\text{s.t. } p_1 x + p_2 y = m$$

We can rewrite as unconstrained problem

$$\max x(m - p_1 x)/p_2$$

Which has FOC:

$$m - 2p_1 x = 0$$
$$x = \frac{m}{2p_1}$$
$$y = \frac{m}{2p_2}$$

Example

What if the constraints are non-linear

$$\begin{aligned} &\max xy \\ &\text{s.t. } x^2 + y^2 = 1 \end{aligned}$$

- Can't just invert the constraint and plug it back in

Example

What if the constraints are non-linear

$$\begin{aligned} &\max xy \\ &\text{s.t. } x^2 + y^2 = 1 \end{aligned}$$

- Can't just invert the constraint and plug it back in
- We know a maximizer must also be a maximizer of either

$$\max_{x \in [-1,1]} x\sqrt{1 - x^2}$$

or

$$\max_{x \in [-1,1]} -x\sqrt{1 - x^2}$$

FOCs:

$$0 = \pm \frac{1 - 2x^2}{\sqrt{1 - x^2}}$$

So have to check

$$(\pm 1/\sqrt{2}, \pm 1/\sqrt{2}), (\pm 1/\sqrt{2}, \mp 1/\sqrt{2}), (\pm 1, 0).$$

How much further can we go with plugging the constraint into the objective?

Lagrange Multipliers

Much harder to do with a bunch of general non-linear constraints. But the implicit function theorem gives us a way to do this!

To keep notation simple: $f : \mathbb{R}^2 \to \mathbb{R}$, $g : \mathbb{R}^2 \to \mathbb{R}$

$$\max f(x)$$
$$\text{s.t. } g(x) = 0$$

Let x^* be a solution to the problem. Then $g(x^*) = 0$.

- Decompose $x = (y, z)$.

Lagrange Multipliers

Much harder to do with a bunch of general non-linear constraints. But the implicit function theorem gives us a way to do this!

To keep notation simple: $f : \mathbb{R}^2 \to \mathbb{R}$, $g : \mathbb{R}^2 \to \mathbb{R}$

$$\max f(x)$$
$$\text{s.t. } g(x) = 0$$

Let x^* be a solution to the problem. Then $g(x^*) = 0$.

- Decompose $x = (y, z)$.
- By IFT, exists an $h(y)$ and a open set $U \subseteq \mathbb{R}$ s.t. $g(y, h(y)) = 0$ for all $y \in U$ and $h(y^*) = z^*$.

Lagrange Multipliers

Much harder to do with a bunch of general non-linear constraints. But the implicit function theorem gives us a way to do this!

To keep notation simple: $f : \mathbb{R}^2 \to \mathbb{R}$, $g : \mathbb{R}^2 \to \mathbb{R}$

$$\max f(x)$$
$$\text{s.t. } g(x) = 0$$

Let x^* be a solution to the problem. Then $g(x^*) = 0$.

- Decompose $x = (y, z)$.
- By IFT, exists an $h(y)$ and a open set $U \subseteq \mathbb{R}$ s.t. $g(y, h(y)) = 0$ for all $y \in U$ and $h(y^*) = z^*$.
- So x^* solves

  $$\max_{x \in U} f(y, h(z))$$

  which has first order condition

  $$f_y(y^*, h(y^*)) + f_z(y^*, h(y^*))h'(y^*) = 0$$

Lagrange Multiplier

$$f_y(y^*, h(y^*)) + f_z(y^*, h(y^*))h'(y^*) = 0$$

Applying IFT

$$f_y(y^*, h(y^*)) + f_z(y^*, h(y^*))h'(y^*) = 0$$
$$f_y(y^*, z^*) - f_z(y^*, z^*)\frac{g_y(y^*, z^*)}{g_z(y^*, z^*)} = 0$$

Lagrange Multiplier

$$f_y(y^*, h(y^*)) + f_z(y^*, h(y^*))h'(y^*) = 0$$

Applying IFT

$$\begin{aligned} f_y(y^*, h(y^*)) + f_z(y^*, h(y^*))h'(y^*) &= 0 \\ f_y(y^*, z^*) - f_z(y^*, z^*)\frac{g_y(y^*, z^*)}{g_z(y^*, z^*)} &= 0 \end{aligned}$$

So if λ exists, it must be $\lambda = f_z(y^*, z^*)/g_z(y^*, z^*)$. Need to make sure

$$\begin{aligned} f_z(y^*, z^*) &= \lambda g_z(y^*, z^*) \\ f_z(y^*, z^*) &= \frac{g_z(y^*, z^*)f_z(y^*, z^*)}{g_z(y^*, z^*)} \end{aligned}$$

Lagrange Multipliers

What did we need here?

- Note: We need $g_z(y^*, z^*) \neq 0$ to apply IFT. So at any max were $\nabla g = 0$, we have a problem
- This argument generalizes directly for arbitrary n, m, just need to keep track of a lot of annoying matricies

Lagrange Multipliers

In general:

Theorem (Lagrange Multipliers)

Let $f : \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$*. Let* x^* *be a solution to the following optimization problem such that rank* $Dg(x^*) = m < n$*:*

$$\max f(x)$$
$$s.t. \;\; g(x) = 0$$

Then there exists a unique Lagrange multiplier $\lambda \in \mathbb{R}^m$ *such that* $Df(x^*) = \lambda' Dg(x^*)$

Now we have a way to solve these problems:

- We now narrow down our set of possible maximizers to any (λ, x) that solves either

$$Df(x) = \lambda' Dg(x)$$
$$g(x) = 0$$

  or be a point where rank $Dg(x) \neq m$.
- Note that these are the first order conditions WRT to x and λ of

$$f(x) - \lambda g(x),$$

  this is called the lagrangian function.
- We could find second order conditions, etc. They exist (involve checking the definiteness of the bordered hessian) but are really annoying to check. We mostly will make assumptions to simplify this part.

Example

$$\begin{aligned} &\max xy \\ &\text{s.t. } x^2 + y^2 = 1 \end{aligned}$$

Example

$$\max xy$$
$$\text{s.t. } x^2 + y^2 = 1$$

FOCs:

$$y = \lambda 2x$$
$$x = \lambda 2y$$

So

$$y^2 = x^2$$

Plugging this back into the constraint we get $x = \pm 1/\sqrt{2}$ and $y = \pm x$ as our critical points. Plugging these into the objective gives us the maximizer.

Example - Cost minimization

A firm wants to reach production level $\bar{q}$, chooses amount of capital (k) and labor (l) to minimize the cost of reaching target given respective prices (r, w).

$$\min_{(k,l)\in\mathbb{R}^2_+} rk + wl$$
$$\text{s.t. } f(k, l) = \bar{q}$$

Example - Cost minimization

A firm wants to reach production level $\bar{q}$, chooses amount of capital (k) and labor (l) to minimize the cost of reaching target given respective prices (r, w).

$$\min_{(k,l)\in\mathbb{R}^2_+} rk + wl$$
$$\text{s.t. } f(k,l) = \bar{q}$$

We get first order conditions

$$r = \lambda f_k(k,l)$$
$$w = \lambda f_l(k,l)$$

So cost minimizing production plan sets MRS=MRT

$$\frac{r}{w} = \frac{f_k(k,l)}{f_l(k,l)}$$

What's next

We still have a few big problems:

1. When is the max unique? When is a local max a global max?
2. What if the constraints can't be expressed in the form $g(x) = 0$.