Lecture IV: Differentiation, IFT, Unconstrained Optimization

1. Differentiation

Definition 1. Let $I \in \mathbb{R}$ be an open interval; a function $f : I \to \mathbb{R}$ is ***differentiable*** at $a \in I$ if

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = L$$

for some L (that is, the limit exists). We write $f'(a) = L$ or $\frac{df}{dx}(a) = L$. If f is differentiable $\forall a \in I$ then we say f is differentiable in I.

Intuitively, differentiation yields the slope of a function at a point:

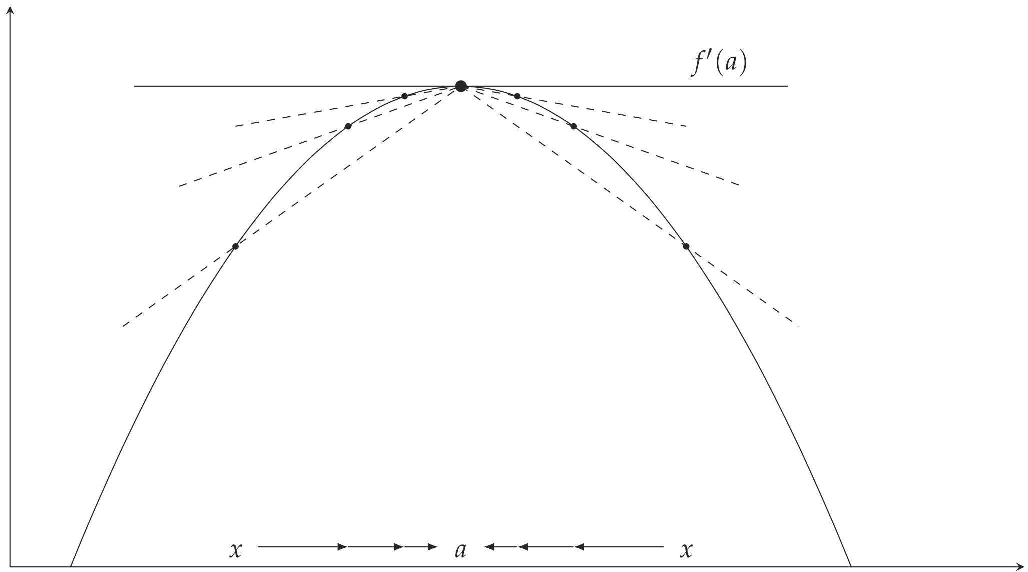


Figure 1: Graphical representation of a derivative

Theorem 1. *If $f : I \to \mathbb{R}$ is differentiable at $a \in I$ then f is also continuous at a.*

Proof. We want to show that $\forall \varepsilon > 0 \;\; \exists \delta > 0$ s.t.

$$|x - a| < \delta \implies |f(x) - f(a)| < \varepsilon$$

Since f is differentiable, we know that for any such ε, we can find some $\widetilde{\delta} > 0$ s.t.

$$\frac{|f(x) - f(a)|}{\widetilde{\delta}} < \frac{|f(x) - f(a)|}{|x - a|} = \left| \frac{f(x) - f(a)}{x - a} - L + L \right| \leq \left| \frac{f(x) - f(a)}{x - a} - L \right| + |L| < \varepsilon + |L|$$

for some L. That is, we know the derivative exists and it is equal to some L, so by the definition of a limit, we can find $\widetilde{\delta} > 0$ s.t.

$$|x - a| < \widetilde{\delta} \implies \left| \frac{f(x) - f(a)}{x - a} - L \right| < \varepsilon$$

Hence whenever $|x - a| < \widetilde{\delta}$ we get

$$|f(x) - f(a)| < (\varepsilon + |L|)\, \widetilde{\delta}$$

If we can find $\delta \leq \widetilde{\delta}$ s.t. $(\varepsilon + |L|)\, \delta < \varepsilon$ then we'd be done. If $\widetilde{\delta} < 1$, then take $\delta = \widetilde{\delta} \cdot \dfrac{\varepsilon}{\varepsilon + |L|} \leq \widetilde{\delta}$ (since $|L| \geq 0$, we multiply $\widetilde{\delta}$ by something < 1). Then we have

$$|x - a| < \delta \leq \widetilde{\delta} \implies |f(x) - f(a)| < (\varepsilon + |L|)\, \frac{\varepsilon}{\varepsilon + |L|} \widetilde{\delta} = \varepsilon \widetilde{\delta} < \varepsilon$$

where the last inequality holds if $\widetilde{\delta} < 1$. If $\delta \geq 1$, then set $\delta = \dfrac{\varepsilon}{\varepsilon + |L|} \leq 1 \leq \delta$ and we get

$$|x - a| < \delta \leq \widetilde{\delta} \implies |f(x) - f(a)| < (\varepsilon + |L|)\, \frac{\varepsilon}{\varepsilon + |L|} = \varepsilon$$

$\square$

Theorem 2. *Let $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ be differentiable.*

- $\frac{d}{dx}[cf(x)] = cf'(x)$.
- $\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$.
- ***Product rule***: $\frac{d}{dx}f(x)g(x) = f'(x)g(x) + f(x)g'(x)$.
- ***Power rule***: $\frac{d}{dx}x^k = kx^{k-1}$.
- ***Chain rule***: $(f \circ g)(x) = f(g(x)) = f'(g(x))g'(x)$.
- ***Quotient rule***: $\frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$.

Some useful special results:

$$\begin{aligned} \frac{d}{dx}e^x &= e^x \\ \frac{d}{dx}\log(x) &= \frac{1}{x} \\ \frac{d}{dx}\sin(x) &= \cos(x) \\ \frac{d}{dx}\cos(x) &= -\sin(x) \end{aligned}$$

Definition 2. A function $f : I \to \mathbb{R}$ is ***continuously differentiable*** if f' is continuous.

Example 1. $f(x) = x^2$ is continuously differentiable since $f'(x) = 2x$ is continuous. However,

$$f(x) = \begin{cases} x^2 \sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is *not* continuously differentiable. In particular, for $x \neq 0$,

$$f'(x) = 2x\cos(1/x) - x^2\frac{1}{x^2}\cos(1/x) = 2x\cos(1/x) - \cos(1/x)$$

and for $x = 0$ the derivative is 0:

$$\lim_{x\to 0}\frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0}\frac{x^2\sin(1/x)}{x} = \lim_{x\to 0} x\sin(1/x)$$

With $\sin(1/x)$ bounded, $x \to 0 \implies x\sin(x) \to 0$. However, the derivative itself is not continuous at 0. Note that while $2x\cos(1/x) \xrightarrow{x\to 0} 0$, $-\cos(1/x)$ does not have a limit, so the derivative does not have a limit as $x \to 0$ either, meaning it cannot be continuous. □

Theorem 3 (L'Hôpital's rule.). *Take $f, g : [a, b] \to \mathbb{R}$ be continuous functions on $[a, b]$ and differentiable on (a, b) except for at most some point $c \in (a, b)$. Further, let*

$$\lim_{x\to c}\frac{f'(x)}{g'(x)} = L$$

and $\lim_{x\to c} g(x) = \lim_{x\to c} f(x) = 0$*. Then*

$$\lim_{x\to c} \frac{f(x)}{g(x)} = L$$

If f, g are differentiable at c then the intuition can be made plain, as this implies $f(c) = g(c) = 0$, and

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f(x)-0}{g(x)-0} \cdot \frac{x-c}{x-c} = \lim_{x\to c} \frac{(f(x)-f(c))/(x-c)}{(g(x)-g(c))/(x-c)} = \frac{f'(c)}{g'(c)}$$

Theorem 4 (Taylor's theorem)**.** *Let* $n \in \mathbb{N}$ *and* $f : [a, b] \to \mathbb{R}$ *be continuously differentiable* n *times on* (a, b)*. Then for all* $\widetilde{x}, x \in (a, b)$ *there is some function* h *s.t.* $\lim_{x\to\widetilde{x}} h(x) = 0$ *and*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(\widetilde{x})}{k!}(x-\widetilde{x})^k + h(x)(x-\widetilde{x})^n$$

The first term is called the ***nth order Taylor approximation*** *of* f *around* $\widetilde{x}$ *and the second term is called the error or remainder term of the Taylor expansion. If we further have that* f *is* $n+1$ *times continuously differentiable then* $\exists c \in (\widetilde{x}, x)$ *(or* $c \in (x, \widetilde{x})$*) s.t.*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(\widetilde{x})}{k!}(x-\widetilde{x})^k + \frac{f^{(n+1)}(c)}{(n+1)!}(x-\widetilde{x})^{n+1}$$

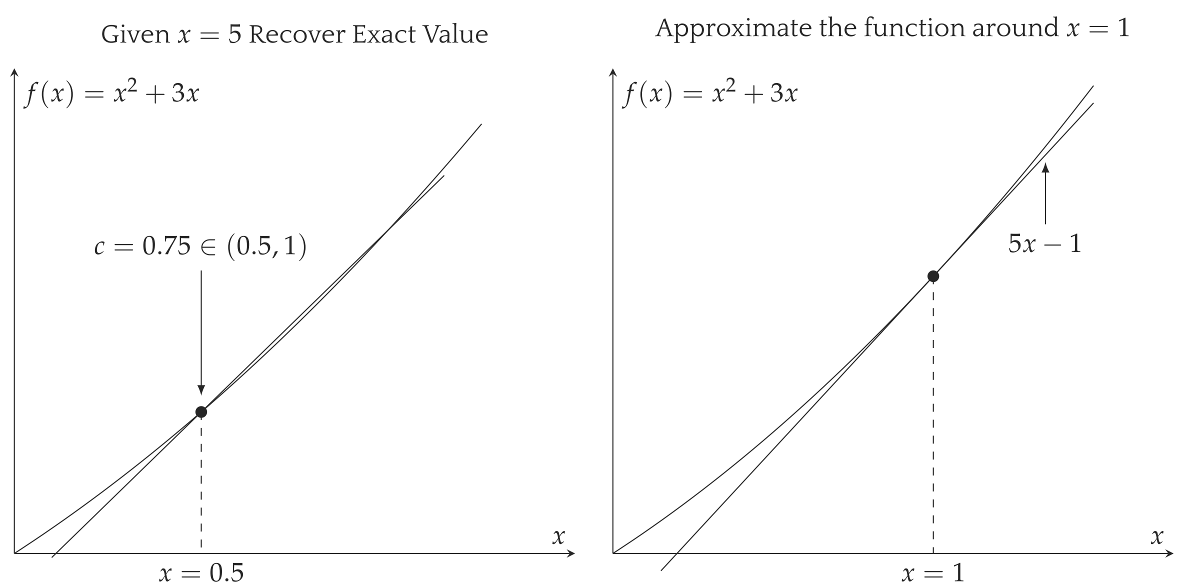
Example 2.

- The Taylor expansion of any polynomial is the polynomial itself. Consider $f(x) = x^2 + 3x$

$$\begin{aligned} f(x) &= \frac{f(1)}{0!}(x-1)^0 + \frac{f'(1)}{1!}(x-1)^1 + \frac{f''(c)}{2!}(x-1)^2 \\ &= 4 + 5(x-1) + \frac{2}{2}(x-1)^2 \\ &= 4 + 5x - 5 + x^2 - 2x + 1 = x^2 + 3x \end{aligned}$$

However, for lower-order Taylor expansions the theorem still applies. Visually:

Figure 2: Visualizing Taylor's Theorem



The left and right figures correspond to:

$$\begin{aligned} f(x) &= \frac{f(1)}{0!}(x-1)^0 + \frac{f'(c)}{1!}(x-1)^1 \\ &= 4 + (2c+3)(x-1) \\ f(x) &\approx \frac{f(1)}{0!}(x-1)^0 + \frac{f'(1)}{1!}(x-1)^1 \\ &= 4 + 5(x-1) \end{aligned}$$

We can see on the left that there is indeed some number c s.t. Taylor's theorem holds. For our sample value of $x = 0.5$ we find $c = 0.75 \in (0.5, 1)$. On the right figure, on the other hand, we plot the *approximation*. In this case, the function and the approximation are exactly equal at 1, since the approximation at that point simplifies to $f(1)$. We can also see that around $x = 1$ the approximation is fairly good! However, farther away the error increases, as we would expect.

- One common Taylor approximation is for the logarithm. In particular, the first order Taylor expansion around 1:

$$\log(x) \approx \log(1) + \frac{1}{1}(x-1) = x - 1$$

Another version of this approximation is for $\log(1+x)$ around 0:

$$\log(1+x) \approx \log(1) + \frac{1}{1+0}(x-0) = x$$

Take the relation:

$$\log Y = \alpha \log K + (1-\alpha)\log L$$

Using the approximation above, we can say that if K increases by 10%, then Y increases by $\alpha \log(1.1) \approx \alpha \cdot 0.1$, that is, $10\alpha\%$.

This second approximation is often used when dealing with percentage changes. You will hear the term ***log-linearization*** thrown around, and this is what that's in reference to: Logarithms can be approximated as a percentage for small values. □

1.1. Partial Derivatives.

Definition 3. Let $f : S \subseteq \mathbb{R}^N \to \mathbb{R}^M$, $e = (e_1, \ldots, e_N)$ a standard basis of $\mathbb{R}^N$, $u = (u_1, \ldots, u_M)$ a standard basis for $\mathbb{R}^M$. $\forall x \in S, f(x) \in \mathbb{R}^M$, $f(x)$ is a linear combination of u and some set of functions $\{f_1, \ldots, f_M\}$ s.t. $f_i : S \to \mathbb{R}$ with

$$f(x) = \sum_{i=1}^{M} f_i(x) u_i$$

Definition 4. The ***partial derivative*** of a function $f : S \subseteq \mathbb{R}^N \to \mathbb{R}^M$ at $x \in S$ is

$$\frac{\partial}{\partial x_j} f_i(x) = \lim_{t \to 0} \frac{f_i(x + t \cdot e_j) - f_i(x)}{t}$$

for $i = 1, \ldots, M$ and $j = 1, \ldots, N$.

Note that partial derivatives needn't imply anything about the behavior of the function overall. Take, for instance, $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

The partial derivatives at 0 are all 0 (crucially, we take the limits one dimension at a time, so it's not that $(x,y) \to (0,0)$, but rather $x \to 0$ and then $y \to 0$. Now take any $y_m \to 0$ and $x_m = y_m^2$ s.t. $y_m \neq 0$ for all n.

$$\lim_{n\to\infty} f(x_m, y_m) = \frac{1}{2} \neq 0$$

which means the function is not even continuous at 0.

Theorem 5 (Schwarz's Theorem)**.** *Let $f : S \subseteq \mathbb{R}^N \to \mathbb{R}$; then*

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

if all the partial derivatives exist and are continuous; that is, mixed partials are symmetric.

Definition 5. The ***gradient*** of $f : S \subseteq \mathbb{R}^N \to \mathbb{R}$ at $x \in S$ is

$$(\nabla f)(x) = \begin{bmatrix} \dfrac{\partial}{\partial x_1} f(x) \\ \vdots \\ \dfrac{\partial}{\partial x_N} f(x) \end{bmatrix}$$

The gradient can also be denoted as $(Df)(x)$ or $D_x f(x)$.

Definition 6. The ***Hessian*** of $f : S \subseteq \mathbb{R}^N \to \mathbb{R}$ at $x \in S$ is

$$(D^2 f)(x) = \begin{bmatrix} \dfrac{\partial^2}{\partial x_1^2} f(x) & \cdots & \dfrac{\partial^2}{\partial x_N \partial x_1} f(x) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2}{\partial x_1 \partial x_N} f(x) & \cdots & \dfrac{\partial^2}{\partial x_N^2} f(x) \end{bmatrix}$$

Where D^2 denotes the application of the D operator twice. The Hessian can also be denoted $D_{x'} D_x f(x)$ (this is analogous to the d/dx^2 operator for univatiate functions, where you can think x^2 for a vector is $x'x$).

NB: I am using x' to denote the transpose; alternatively this can be denoted x^T.

Definition 7. Let $f : S \subseteq \mathbb{R}^N \to \mathbb{R}^M$ and $x \in S$; the ***Jacobian*** of f is the $M \times N$ matrix with i, j entry equal to $\dfrac{\partial}{\partial x_j} f_i(x)$, denoted $D_{x'} f(x)$.

1.2. Mean Value Theorem (MVT).

Theorem 6 (Mean Value Theorem (MVT))**.** *Let $f : [a,b] \to \mathbb{R}$ be a continuous function on the closed interval $[a,b]$ and differentiable on the open interval (a,b). Then $\exists c \in (a,b)$ s.t.*

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

We will prove an equivalent, albeit perhaps conceptually easier, version of this:

Theorem 7 (Rolle's Theorem)**.** *Let $g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) with $g(a) = g(b)$. Then $\exists c \in (a, b)$ s.t.*

$$f'(c) = 0$$

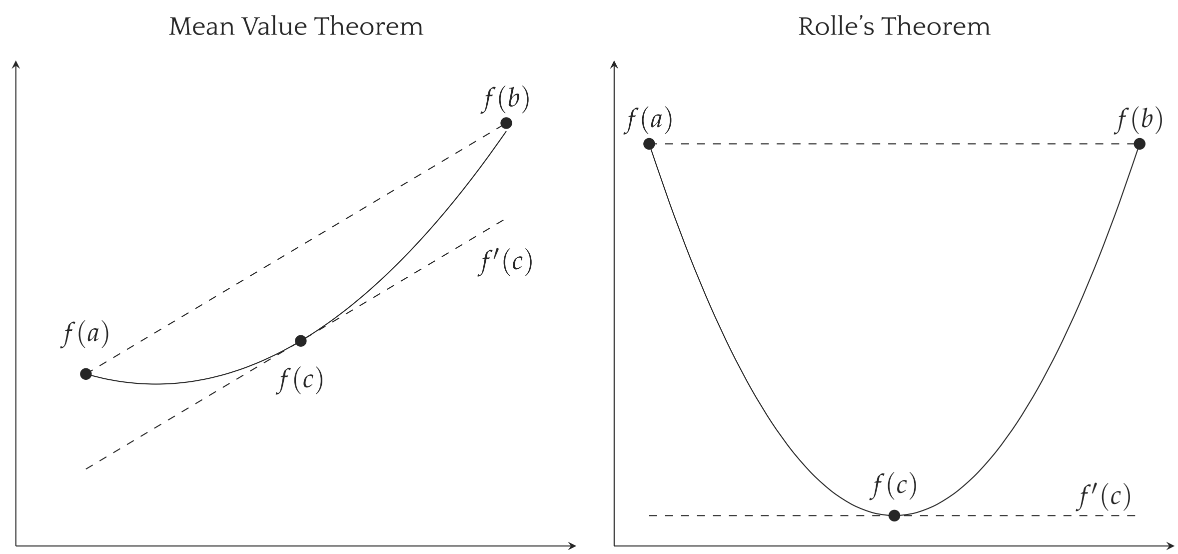


Figure 3: Graphical depiction of MVT

Claim 1. *Rolle's Theorem iff Mean Value Theorem (MVT).*

Proof. The mean value theorem implies Rolle's theorem by definition. Simply note that by the MVT there is some c s.t.

$$f'(c) = \frac{g(b) - g(a)}{b - a} = 0$$

Now to show Rolle's theorem implies MVT, we only need a simple transformation:

$$g(x) = (f(x) - f(a)) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Note that $g(a) = g(b) = 0$. Hence $\exists c$ s.t.

$$g'(c) = 0$$

But we can see that

$$g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0 \implies f'(c) = \frac{f(b) - f(a)}{b - a}$$

□

Now we show Rolle's Theorem.

Proof. For Rolle's theorem we will use the extreme value theorem, and we know that f is bounded and attains its $\sup$ and its $\inf$ in $[a, b]$. If the $\sup$ and the $\inf$ are *both* at $\{a, b\}$, then because by assumption

$f(a) = f(b) = 0$, $f(x) = 0$ everywhere on $[a, b]$. Take any $c \in (a, b)$, and

$$f'(c) = \lim_{y \to c} \frac{f(y) - f(c)}{y - c} = \lim_{y \to c} \frac{0}{y - c} = \lim_{y \to c} 0 = 0$$

Suppose then that either the $\sup$ or the $\inf$ occur at at interior point $c \in (a, b)$. Take the $\sup$ (the case for the $\inf$ will be analogous). By assumption f is differentiable, so $f'(c)$ exists. That is, we know that

$$f'(c) = \lim_{y \to c} \frac{f(y) - f(c)}{y - c} = L$$

Consider $y \to c^-$, that is y approaching c from the left. Since the $\sup$ is attained at c, $f(c) \geq f(y)$ for all $c > y$, which in turn gives

$$\frac{f(y) - f(c)}{y - c} \geq 0 \qquad \forall c > y$$

Now take $y \to c^+$, that is y approaching c from the right. Again, since the $\sup$ is attained at c, $f(c) \geq f(y)$ for all $c < y$, which in turn gives

$$\frac{f(y) - f(c)}{y - c} \leq 0 \qquad \forall c < y$$

Which means that

$$L_- = \lim_{y \to c^-} \frac{f(y) - f(c)}{y - c} \geq 0$$

Suppose not, then $L_- < 0$. But for every $\varepsilon > 0$ there is some δ s.t.

$$c^- - \delta < y < c^- < c^- + \delta \implies L_- - \varepsilon < \frac{f(y) - f(c)}{y - c} < L_- + \varepsilon$$

Take ε s.t. $L_- < -\varepsilon < 0$. Then for some $\delta > 0$, we get

$$c^- - \delta < y < c^- < c^- + \delta \implies L_- - \varepsilon < \frac{f(y) - f(c)}{y - c} < L_- + \varepsilon < 0$$

But we have established that $y < c$ gives $\frac{f(y) - f(c)}{y - c} \geq 0$, a contradiction. By that same logic, we can write

$$L_+ = \lim_{y \to c^+} \frac{f(y) - f(c)}{y - c} \leq 0$$

Since $f'(c)$ exists, it must be the case that

$$f'(c) = \lim_{y \to c^+} \frac{f(y) - f(c)}{y - c} = L_+ = \lim_{y \to c^-} \frac{f(y) - f(c)}{y - c} = L_- = L$$

Now we can see that $L_+ \leq 0$, $L_- \geq 0$, and $L_+ = L_- = L$ gives that $L = 0$. The proof for the case when the $\inf$ is interior is completely analogous. $\square$

Corollary 1. *If f is continuous on $[a, b]$ and differentiable on (a, b) and obtains a local minimum or maximum at c, then $f'(c) = 0$.*

Corollary 2. *If f is continuous on $[a, b]$ and differentiable on (a, b) and $f'(x) > 0$ for every $x \in (a, b)$ then f is increasing on (a, b). Conversely, if $f'(x) < 0$ for every $x \in (a, b)$ then f is decreasing on (a, b).*

Theorem 8 (Generalized MVT). *Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions on $[a, b]$ and differentiable on (a, b). Then $\exists c \in (a, b)$ s.t.*

$$g'(c)\,(f(b) - f(a)) = f'(c)\,(g(b) - g(a))$$

The MVT is a case where $g(x) = x$, the identity function.

2. Implicit Function Theorem (IFT)

2.1. Motivation. Consider a function $f(x, y) = 0$. How do we characterize a function relating x to y? We can write that for $f : \mathbb{R}^2 \to \mathbb{R}$ continuously differentiable on an open set O with $f(x, y) = 0$, there exists some function h s.t.

$$f(x, h(x)) = 0 \text{ and } \frac{dy}{dx} = -\frac{\partial f/\partial x}{\partial f/\partial y}$$

One classic example is how to characterize the slope of a tangent line at a point (x, y) of some circle of radius r centered at $(0, 0)$. That is, $x^2 + y^2 = r^2$. We can write

$$f(x, y) = x^2 + y^2 - r^2 = 0$$

So we know that for some $h(x)$, $f(x, h(x)) = 0$. Furthermore,

$$\frac{\partial f}{\partial x} = 2x \quad \frac{\partial f}{\partial y} = 2y \implies \frac{dy}{dx} = -\frac{x}{y}$$

We will look at the general version of the theorem. We often work with a parameter space and a variable space, and we want to express the variables in terms of the parameters (or with exogenous and endogenous variables, and we want to express the endogenous variables in terms of the exogenous variables). Take

$$(\theta_1, \ldots, \theta_N) = \theta \in \mathbb{R}^N$$

to be the parameters (or exogenous) and

$$(x_1, \ldots, x_M) = x \in \mathbb{R}^M$$

to be the variables (or endogenous). It's typically not common to have an explicit expression for the latter in terms of the former, but often we will encounter an implicit relation of the form

$$f(\theta, x) = 0$$

For example, some system of equations

$$\begin{aligned} f_1(\theta, x) &= 0 \\ &\vdots \\ f_M(\theta, x) &= 0 \end{aligned}$$

The IFT gives a result we can apply to these types of problems.

2.2. The IFT.

Theorem 9 (Implicit Function Theorem (IFT)). *Take a function $f : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}^M$ that is continuously differentiable and fix a point $(\widetilde{\theta}, \widetilde{x}) \in \mathbb{R}^N \times \mathbb{R}^M$ s.t. $f(\widetilde{\theta}, \widetilde{x}) = 0$. If $D_{x'} f(\widetilde{\theta}, \widetilde{x})$ is non-singular (i.e. full-rank,* *or has a non-0 determinant) then for some open sets A, B s.t. $\widetilde{\theta} \in A$, $\widetilde{x} \in B$, there exist a unique function $h : A \to B$ that is continuously differentiable in A s.t.*

$$f(\theta, h(\theta)) = 0$$

for all $\theta \in A$, and $h(\widetilde{\theta}) = \widetilde{x}$. Taking derivatives with respect to θ', we further have

$$D_{\theta'} f(\theta, h(\theta)) + D_{x'} f(\theta, h(\theta)) D_{\theta'} h(\theta) = 0$$
$$D_{\theta'} h(\theta) = -\left[D_{x'} f(\theta, h(\theta))\right]^{-1} D_{\theta'} f(\theta, h(\theta))$$

2.3. Example. Take a simplified version of the IS-LM model

$$\begin{aligned} Y &= C + I + G \\ C &= C(Y - T) \\ I &= I(r) \\ M^S &= M^D(Y, r) \end{aligned}$$

with

$$0 < C'(x) < 1 \qquad I'(r) < 0 \qquad \frac{\partial M^D}{\partial Y} > 0 \qquad \frac{\partial M^D}{\partial r} < 0$$

National income must equal consumption plus investment (savings) plus government spending; consumption is some function of income minus taxes, the level of investment is determined by the interest rate, and money supply must equal money demand. We have that

$$\begin{aligned} Y - C(Y - T) - I(r) - G &= 0 \\ M^S - M^D(Y, r) &= 0 \end{aligned}$$

which is the exact type of problem the IFT can help us solve. We have endogenous variables $x = (Y, r)$, national income and the interest rate, and exogenous variables $\theta = (M^S, G, T)$, money supply, government spending, and taxes. Hence

$$f(\theta, x) = \begin{bmatrix} f_1(\theta, x) \\ f_2(\theta, x) \end{bmatrix} = \begin{bmatrix} Y - C(Y - T) - I(r) - G \\ M^S - M^D(Y, r) \end{bmatrix} = 0 \tag{1}$$

Hence for some h, we can write

$$h(\theta) = \begin{bmatrix} Y(M^S, G, T) \\ r(M^S, G, T) \end{bmatrix}$$
$$D_{\theta'} f(\theta, h(\theta)) + D_{x'} f(\theta, h(\theta)) D_{\theta'} h(\theta) = 0$$

We have that

$$D_{\theta'} f(\theta, h(\theta)) = \begin{bmatrix} 0 & -1 & C'(Y(\cdot) - T) \\ 1 & 0 & 0 \end{bmatrix}$$

$$D_{x'}f(\theta,h(\theta)) = \begin{bmatrix} \frac{\partial f_1}{\partial Y} = 1 - C'(Y(\cdot)-T) & \frac{\partial f_1}{\partial r} = -I'(r(\cdot)) \\ \frac{\partial f_2}{\partial Y} = -\frac{\partial M^D}{\partial Y} & \frac{\partial f_2}{\partial r} = -\frac{\partial M^D}{\partial r} \end{bmatrix}$$

$$[D_{x'}f(\theta,h(\theta))]^{-1} = \frac{1}{D}\begin{bmatrix} -\frac{\partial M^D}{\partial r} & I'(r(\cdot)) \\ \frac{\partial M^D}{\partial Y} & 1 - C'(Y(\cdot)-T) \end{bmatrix}$$

$$D = \underbrace{-\overbrace{\frac{\partial M^D}{\partial r}}^{<0}\underbrace{\Big(1 - C'(Y(\cdot)-T)\Big)}_{>0}}_{>0} \underbrace{-\overbrace{I'(r(\cdot))}^{<0}\underbrace{\frac{\partial M^D}{\partial Y}}_{>0}}_{>0} \implies D > 0$$

A non-zero determinant implies that the inverse exists. Hence we find that

$$D_{\theta'}h(\theta) = \begin{bmatrix} \frac{\partial Y}{\partial M^S} & \frac{\partial Y}{\partial G} & \frac{\partial Y}{\partial T} \\ \frac{\partial r}{\partial M^S} & \frac{\partial r}{\partial G} & \frac{\partial r}{\partial T} \end{bmatrix} = -\frac{1}{D}\begin{bmatrix} I'(r(\cdot)) & \frac{\partial M^D}{\partial r} & -\frac{\partial M^D}{\partial r}C'(Y(\cdot)-T) \\ 1-C'(Y(\cdot)-T) & -\frac{\partial M^D}{\partial Y} & \frac{\partial M^D}{\partial Y}C'(Y(\cdot)-T) \end{bmatrix}$$

$$= -\frac{1}{D}\begin{bmatrix} <0 & <0 & >0 \\ >0 & <0 & >0 \end{bmatrix} = \frac{1}{D}\begin{bmatrix} >0 & >0 & <0 \\ <0 & >0 & <0 \end{bmatrix}$$

Which means that for some $x = (Y,r), \theta = (M^S, G, T)$ that satisfies Equation (1) there is some local neighborhood around θ where we can characterize the behavior of (Y,r) with respect to each of the variables in θ. In particular, income reacts positively to increases money supply or government spending but negatively to taxes, while the interest rate goes down with increases in the money supply or taxes but goes up with increases in government spending.

3. Unconstrained Optimization

Definition 8. Let $f : A \to B$

1. $x \in A$ is a ***local maximum*** of f if $\exists \varepsilon > 0$ s.t.

$$y \in B_\varepsilon(x) \cap A \implies f(x) \geq f(y)$$

   The local maximum is ***strict*** if the inequality is strict. (Note the intersection: For example, $f(x) = x$ has no local maximum on $\mathbb{R}$, but every point is a local maximum if we define the function over $x \in \mathbb{N}$.)

2. The ***local minimum*** definition is analogous.

3. $x \in A$ is a ***global maximum*** of f if $\forall y \in A$, $f(x) \geq f(y)$. It is a strict global maximum if whenever $x \neq y$ we have $f(x) > f(y)$.

4. The ***global minimum*** definition is analogous.

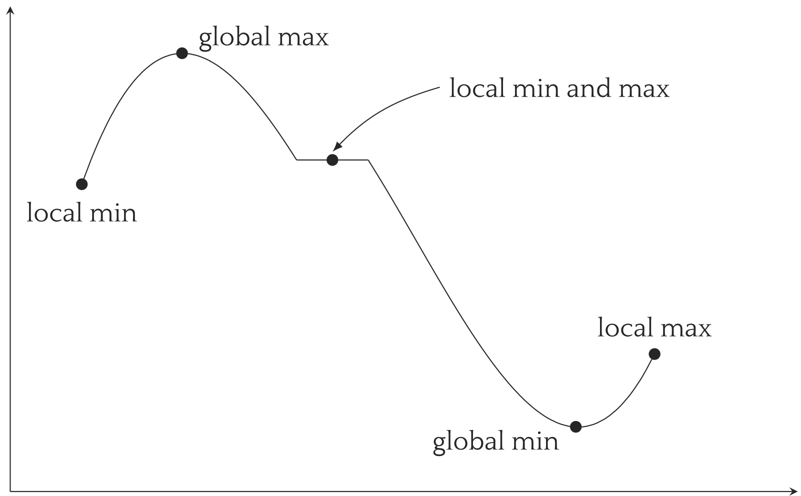


Figure 4: Examples of local and global maxima and minima

Definition 9. The argmax of a function f is the set

$$\underset{x \in A}{\operatorname{argmax}} f(x) = \{x \in A : f(x) \geq f(y) \ \ \forall y \in A\}$$

The argmin is analogously defined.

3.1. Linear Algebra Review.

Definition 10. A square $N \times N$ matrix S with elements in $\mathbb{R}$ is ***positive semidefinite*** (PSD) if $\forall p \in \mathbb{R}^N$

$$p^T S p \geq 0$$

and ***positive definite*** (PD) if the inequality is strict whenever $p \neq 0$.

Definition 11. A square $N \times N$ matrix S with elements in $\mathbb{R}$ is ***negative semidefinite*** (NSD) if $\forall p \in \mathbb{R}^N$

$$p^T S p \leq 0$$

and ***negative definite*** (ND) if the inequality is strict $p \neq 0$.

Definition 12. Let S be a $N \times N$ matrix with elements in $\mathbb{R}$. If $\exists p_1, p_2 \in \mathbb{R}^N$ s.t.

$$p_1^T S p_1 > 0 \quad \text{and} \quad p_2^T S p_2 < 0$$

then we say S is ***indefinite***.

Example 3. Consider the matrix

$$S = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

and take $p_1 = (1, 0, 0)$, $p_2 = (-2, 1, 0)$. Then

$$p_1^T S p_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1 > 0$$

$$p_2^T S p_2 = \begin{bmatrix} -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} = -3 < 0$$

So S is indefinite. In general we only need one counterexample for indefiniteness, but we need to check that *every* vector p gives a positive or negative quadratic form for definiteness. It turns out for symmetric matrices there are rules definite matrices have to follow that will help us determine their definiteness. □

Definition 13. Let S be a $N \times N$ matrix with elements in $\mathbb{R}$. A kth order ***principal submatrix*** is the submatrix of S obtained by removing $N - k$ rows and the corresponding columns of S. A kth order ***principal minor*** is the determinant of a kth order principal submatrix.

It's easiest to talk about the principal minors using examples: Take

$$S = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

The 1st order principal submatrices are

$$\begin{bmatrix} 1 \end{bmatrix} \qquad \begin{bmatrix} 5 \end{bmatrix} \qquad \begin{bmatrix} 9 \end{bmatrix}$$

and the principal minors are the determinants therein. The 2nd order principal submatrices are

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix} \qquad \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}$$

and the $2nd$ order principal minors are their determinants:

$$\det \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} = -3 \qquad \det \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix} = -12 \qquad \det \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} = -3$$

Finally, the 3rd order principal submatrix is just the matrix S itself, and the 3rd order principal minor is the determinant of S (in this case, 0).

Definition 14. Let S be a $N \times N$ matrix with elements in $\mathbb{R}$. The kth ***leading principal minor*** is the principal minors obtained by removing the "last" $N - k$ columns and rows of S.

In our example above, these are the determinants of the matrix S itself (3rd leading principal minor), and

$$\text{1st} \rightarrow \det[1] = 1 \qquad \text{2nd} \rightarrow \det \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} = -3 \qquad \text{3rd} \rightarrow \det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = 0$$

Definition 15. A matrix S is ***symmetric*** if $S = S^T$; that is, so S with entries $s(i,j)$ we have

$$s(i,j) = s(j,i)$$

Theorem 10. *Let S be a $N \times N$ symmetric matrix.*

1. *If all the leading principal minors are strictly positive, then S is positive definite.*
2. *If for every $k \leq N$ the kth order leading principal minor has sign $(-1)^k$ (that is, positive for k even and negative for k odd), then S is negative definite.*
3. *If all principal minors of S are weakly positive (≥ 0) then S is positive semidefinite.*
4. *If for every $k \leq N$ the kth order principal minors are ≤ 0 when k is odd and ≥ 0 when k is even, then S is negative semidefinite.*
5. *If the principal minors do not fit any of the above patterns then S is indefinite.*

Example 4. Consider the matrices

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

A, the identity, is positive definite. Consider any non-zero p:

$$p^T A p = p^T p = \sum_i p_i^2 > 0$$

if $p_i \neq 0$ for some i. Note all the leading principal minors are 1. For B, we similarly have

$$p^T B p = -p^T p = p^T p = \sum_i p_i^2 < 0$$

if $p_i \neq 0$ for some i. Note the leading principal minors are $-1, 1, 1$. For C, take $p_1 = (0,1,0)$, $p_2 = (0,0,-1)$:

$$p_1^T C p_1 = 1 > 0 \qquad p_2^T C p_2 = -1 < 0$$

so C is indefinite. Note the leading principal minors are all 0, but the non-leading principal minors do not obey the pattern that gives semi-definiteness. In this case, the 1st-order principal minors are 0 (leading), 1, and -1, which is already an issue since the sign flips within a given set of principal minors. ☐

3.2. First Order Conditions (FOC).

Theorem 11. *Let $f : A \rightarrow \mathbb{R}$ be a continuously differentiable function on an open set $A \subseteq \mathbb{R}^N$. If $x^* \in A$ is a local minimum or maximum, then*

$$Df(x^*) = 0$$

that is, the first-order partials evaluated at x^* equal 0.

Remark 1. In general the converse need not be true. For instance $f(x) = x^3$. We have $f'(x) = 3x^2 = 0$ if $x = 0$. However the function does not have a local minimum or maximum at 0. Hence $Df(x) = 0$ is a necessary but not sufficient condition. □

Some examples

1. Take $f(x) = 2x^3 - 3x^2$. We have
$$Df(x) = 6x^2 - 6x$$
$Df(x) = 0$ at $x = 0, 1$, so if f has local maxima or minima they must occur at those points, but we don't yet know how to check whether they are local maxima or minima.

2. $f(x, y) = x^3 - y^3 + 9xy$, so $f : \mathbb{R}^2 \to \mathbb{R}$. We have
$$Df(x,y) = \begin{bmatrix} 3x^2 + 9y \\ -3y^2 + 9x \end{bmatrix}$$
$Df(x, y) = 0$ at $(0, 0)$ and $(3, -3)$. Note
$$\begin{aligned} 0 &= 3x^2 + 9y \\ 0 &= -3y^2 + 9x \\ 3y^2 &= 9x \\ x^2 &= -3y \\ x^4 &= 9y^2 = 3(3y^2) = 3(9x) \end{aligned}$$
So
$$\begin{aligned} x^4 = 27x &\implies x = 3 \\ 27 + 9y = 0 &\implies y = -3 \end{aligned}$$

3.3. Second Order Conditions (SOC).

Theorem 12. *Let $f : A \to \mathbb{R}$ be a twice continuously differentiable function on an open set $A \subseteq \mathbb{R}^N$ with*
$$Df(x^*) = 0$$
for some $x^* \in A$. If $D^2 f(x^*)$ the Hessian at x^* is negative definite then x^* is a local maximum, and if it is positive definite then it is a local minimum.

Remark 2. The converse need not hold. For instance, take $f(x) = x^4, Df(x) = 4x^3, D^2 f(x) = 12x^2$. At $x^* = 0$, we have $Df(0) = 0$, but $D^2 f(0) = 0$ is neither positive nor negative definite. Hence a strictly definite Hessian is a sufficient condition for a local maximum or minimum, but it is not necessary. □

Theorem 13. *Let $f : A \to \mathbb{R}$ be a twice continuously differentiable function on an open set $A \subseteq \mathbb{R}^N$. If $x^* \in A$ is a local maximum, then*
$$Df(x^*) = 0$$

and $D^2 f(x^*)$ is negative semidefinite. If $x^* \in A$ is a local minimum, then $Df(x^*) = 0$ and $D^2 f(x^*)$ is positive semidefinite.

Remark 3. Again, the converse need not be true. In our previous example, $x^* = 0$ is actually a local minimum, and we can check that $Df(x^*) = 4(0^3) = 0$ and $D^2 f(x^*) = 12(0^2) = 0 \geq 0$ (positive semidefinite). However, $D^2 f(x^*) \leq 0$ means that it is negative semidefinite as well, but that does not imply a local maximum. Hence the condition is necessary but not sufficient. □

Let us take $f(x) = 2x^3 - 3x^2$ again. We saw that $Df(x) = 6x^2 - 6x = 0$ at $x = 0, 1$. Now we have

$$D^2 f(x) = 12x - 6$$
$$D^2 f(0) = -6 < 0 \implies \text{local max}$$
$$D^2 f(1) = 6 > 0 \implies \text{local min}$$

What about $f(x, y) = x^3 - y^3 + 9xy$? Recall

$$Df(x,y) = \begin{bmatrix} 3x^2 + 9y \\ -3y^2 + 9x \end{bmatrix} = 0 \impliedby (x, y) = (0, 0) \text{ or } (3, -3)$$

We find that

$$D^2 f(x,y) = \begin{bmatrix} 6x & 9 \\ 9 & -6y \end{bmatrix}$$

The second order principal minor is the determinant of the Hessian itself. The 1st order principal minor is $6x$. Note the determinant of the Hessian is

$$\left|D^2 f(x,y)\right| = -36xy - 81$$

1. For $(0, 0)$, we have $6(0) = 0$ and $-36(0)(0) - 81 = -81 < 0$. The 2nd leading principal minor is negative, so the Hessian at that point cannot be positive definite. Further, $(-1)^2$ is positive, so it cannot be negative definite either. Hence the Hessian at $(0, 0)$ is indefinite.
2. For $(3, -3)$, we have $6(3) = 18 > 0$ and $-36(3)(-3) - 81 = 324 - 81 = 243 > 0$. The 1st and 2nd leading principal minors are both positive, which means the Hessian at that point is positive definite and $(3, -3)$ is a local minimum.

NB: This is exactly the second-derivative test you may have learned in early calculus. For a function $f : \mathbb{R}^2 \to \mathbb{R}$, the Hessian is given by

$$D^2 f(x,y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix}$$

with f_{xy} denoting the partials with respect to x and then y.

1. The second-oder leading principal minor has to be positive for the Hessian to be definite, positive or negative. This is the determinant of the Hessian itself, or $f_{xx} f_{yy} - 2f_{xy} > 0$.

2. The first-oder leading principal minor has to be positive for a min, or $f_{xx} > 0$, which gives PD.

3. The first-oder leading principal minor has to be negative for a max, or $f_{xx} < 0$, which gives ND.

Remark 4. One way to think about why it is that definiteness of the Hessian gives local extrema is to use a "multivariate" version of Taylor's theorem. Let $f : A \to \mathbb{R}$ be a twice continuously differentiable function on an open set with $Df(x^*) = 0$ for some x^*. Note any x can be written as $x^* + \alpha z$ for some unit vector z and some α. Note all such x are at most α away from x^* (i.e. $x \in B_\alpha(x^*)$). Let $g(\alpha) = f(x^* + \alpha z)$ and consider the second-order Taylor expansion of g around $\alpha = 0$:

$$g(\alpha) = g(0) + g'(0)\alpha + \frac{1}{2}g''(0)\alpha^2 + h(\alpha)\alpha^2$$

$$f(x^* + \alpha z) = f(x^*) + \alpha z^T Df(x^*) + \frac{\alpha^2}{2} z^T D^2 f(x^*) z + h(\alpha)\alpha^2$$

with $h(\alpha) \to 0$ as $\alpha \to 0$. We know $Df(x^*) = 0$ by premise. If $D^2 f(x^*)$ is positive definite, it turns out there exists some λ s.t. $z^T D^2 f(x^*) z \geq \lambda z^T z = \lambda > 0$ for any unit vector z.[1] Hence

$$f(x^* + \alpha z) = f(x^*) + \frac{\alpha^2}{2} z^T D^2 f(x^*) z + h(\alpha)\alpha^2 > f(x^*) + \left(\frac{\lambda}{2} + h(\alpha)\right)\alpha^2$$

for *any* z s.t. $x^* + \alpha z \in B_\delta(x^*)$. Since $h(\alpha) \to 0$, there should be an α small enough to make the last term above positive; thus for some $\alpha > 0$ we have

$$f(x^* + \alpha z) > f(x^*)$$

for any such z; the last step is to remark once again $x = x^* + \alpha z \in B_\alpha(x^*)$. Hence if $D^2 f(x^*)$ is PD we have a local min for the α-neighborhood. The steps for a local max are analogous (note λ will be negative). □

3.4. Concavity and Convexity.

Definition 16. A function $f : A \to \mathbb{R}$ is ***concave*** if for any $\alpha \in [0, 1]$ and $x, y \in A$

$$\alpha f(x) + (1-\alpha) f(y) \leq f(\alpha x + (1-\alpha) y)$$

It is ***strictly concave*** if the above holds strictly for $\alpha \in (0, 1)$.

Definition 17. A function $f : A \to \mathbb{R}$ is ***convex*** if for any $\alpha \in [0, 1]$ and $x, y \in A$

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y)$$

It is ***strictly convex*** if the above holds strictly for $\alpha \in (0, 1)$.

Theorem 14. *Let $f : A \to \mathbb{R}$ be a twice continuously differentiable function on an open set $A \subseteq \mathbb{R}^N$.*

1. *$f(x)$ is concave $\iff$ $D^2 f(x)$ is negative semidefinite. (Negative definite $\implies$ strictly concave.)*
2. *$f(x)$ is convex $\iff$ $D^2 f(x)$ is positive semidefinite. (Positive definite $\implies$ strictly convex.)*

[1] We will discuss this in the last lecture, but a symmetric matrix is PD iff all its eigenvalues are strictly positive. Further, a symmetric matrix S can be decomposed into $C\Lambda C^T$ with Λ a diagonal matrix of eigenvalues and C an orthonormal matrix of eigenvectors. Hence $z^T S z = z^T C \Lambda C^T z = p^T \Lambda p = \sum_i \lambda_i p_i^2$ for $p = C^T z$. Let $\lambda = \min_i \lambda_i$ and we have the result, since $z^T S z \geq \lambda p^T p = z^T C C^T z = z^T z$ since C is orthonormal. Last, a symmetric matrix is ND iff all its eigenvalues are strictly negative, and the analogous steps give the result with the inequality flipped.

Remark 5. As was pointed out during the lecture, $f(x) = x^4$ is convex but the Hessian is not positive definite everywhere. In particular, $D^2 f(0) = 0$. Hence while a positive definite Hessian implies concavity, the converse only gives weak concavity. ☐

Remark 6. The univariate intuition for concave and convex functions generalizes. The first derivative of a concave function is decreasing (the function is either increasing at a decreasing rate or decreasing at an increasing rate), so the second derivative must be negative; the converse for a convex function. To see why this intuition is sufficient for multivariate functions, consider

$$g(\alpha) = f(\alpha x + (1-\alpha)y)$$

and note

$$g''(\alpha) = (x-y)^T D^2 f(\alpha x + (1-\alpha)y)(x-y)$$

which is the square form for the Hessian (and exactly the form we want to check for definiteness). We will show f is concave iff g is concave. If f is concave,

$$\begin{aligned}
g(\alpha a + (1-a)\beta) &= f((\alpha a + (1-a)\beta)x + (1 - (\alpha a + (1-a)\beta))y) \\
&= f\left(\alpha a x + (1-a)\beta x + ay + (1-a)y - \alpha a y - (1-a)\beta y\right) \\
&= f\left(a(\alpha x + (1-\alpha)y) + (1-a)(\beta x + (1-\beta)y)\right) \\
&\geq a f(\alpha x + (1-\alpha)y) + (1-a) f(\beta x + (1-\beta)y) \\
&= a g(\alpha) + (1-a) g(\beta)
\end{aligned}$$

If g is concave,

$$f(\alpha x + (1-\alpha)y) = g(\alpha) \geq \alpha g(1) + (1-\alpha)g(0) = \alpha f(x) + (1-\alpha)f(y)$$

Noting $\alpha = 1 \cdot \alpha + (1-\alpha) \cdot 0$. Hence it is sufficient to show g has a negative second derivative, which follows from the univariate version of the theorem. ☐

Theorem 15. *Let $f : A \to \mathbb{R}$ be a twice continuously differentiable function on an open set $A \subseteq \mathbb{R}^N$.*

1. *If f is concave and x^* is s.t. $Df(x^*) = 0$ then x^* is a global maximum. (If f is strictly concave the global maximum is unique.)*

2. *If f is convex and x^* is s.t. $Df(x^*) = 0$ then x^* is a global minimum. (If f is strictly convex the global minimum is unique.)*

Some examples:

1. Take the function $f(x, y) = x^2 + y^2$. We have

$$Df(x,y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix} \qquad D^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

   Note that $Df(0,0) = (0,0)$. Further, the 1st leading principal minor of the Hessian is $2 > 0$; for the 2nd leading principal minor we have $2(2) - 0(0) = 4 > 0$. Thus $D^2 f(x)$ is positive definite; this means f is convex and $(0,0)$ is a global minimum.

2. What about $f(x) = x^4$? In this case

$$Df(x) = 4x^3 \qquad D^2 f(x,y) = 12x^2$$

$Df(0) = 0$ and $D^2 f(0) = 0$. However, $12x^2 \geq 0$ for all x, so the function is convex, which means that 0 is a global minimum.

3. $f(x,y) = x^2 y^2$

$$Df(x,y) = \begin{bmatrix} 2xy^2 \\ 2x^2 y \end{bmatrix} \qquad D^2 f(x,y) = \begin{bmatrix} 2y^2 & 4xy \\ 4xy & 2x^2 \end{bmatrix}$$

Note $Df(x,0) = Df(0,y) = (0,0)$, but the kth order principal minors are all 0 at $(x,0)$ or $(0,y)$. More genearlly we have that while the 1st principal minors are $2y^2 \geq 0$ and $2x^2 \geq 0$, the determinant of the 2nd principal minor is

$$4x^2 y^2 - 16x^2 y^2 \leq 0$$

Hence we cannot even say whether it is positive or negative semidefinite.