Math Camp 2012:
Calculus

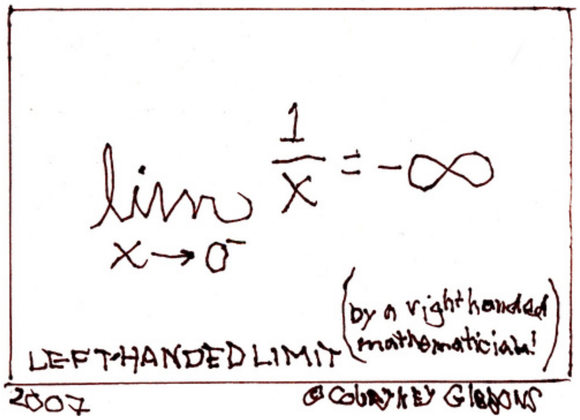
August 2012

Topics*:

- Limits
  - The limits of a function
  - Continuity
  - The limit of a series
- Change we can believe in
  - Derivatives
  - Higher-Order Derivatives
  - Maxima and Minima
  - Composite Functions
  - The Chain Rule, Derivatives of Exp and Ln
  - L'Hospital's Rule
- The area under a curve
  - The Indefinite Integral: The Antiderivative
  - The Definite Integral: The Area under the Curve
  - Integration by Substitutions
  - Integration by Parts
- Calculus in higher dimensions
  - Partial Derivatives
  - Partial Derivatives of Higher Order
  - Multidimensional Integrals
  - Calculus in vector and matrix form
  - First order conditions
  - Second order conditions and the Hessian
  - Global or local?

*These notes were prepared by Jacob Montgomery. Much of the material and examples for this lecture are taken from Harvard "Math (P)refresher" class notes whose authors are listed here and *A Mathematical Primer for Social Statistics* by John Fox.

1 Limits

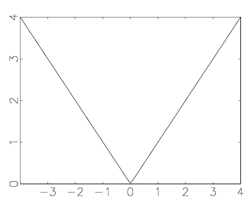


- We're often interested in determining if a function f approaches some number L as its independent variable x moves to some number c (usually 0 or $\pm\infty$). If it does, we say that $f(x)$ approaches L as x approaches c, or $\lim_{x\to c} f(x) = L$.

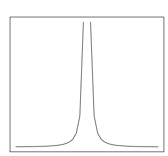
- **Limit of a function**. Let f be defined at each point in some open interval containing the point c, although possibly not defined at c itself. Then $\lim\limits_{x\to c} f(x) = L$ if for any (small positive) number ϵ, there exists a corresponding number $\delta > 0$ such that if $0 < |x - c| < \delta$, then $|f(x) - L| < \epsilon$.

- Examples:

  1. $\lim\limits_{x\to c} k = k$
  2. $\lim\limits_{x\to c} x = c$
  3. $\lim\limits_{x\to 0} |x| = 0$

  4. $\lim\limits_{x\to 0} \left(1 + \frac{1}{x^2}\right) = \infty$

- Uniqueness: $\lim\limits_{x\to c} f(x) = L$ and $\lim\limits_{x\to c} f(x) = M \Longrightarrow L = M$

- Properties: Let f and g be functions with $\lim\limits_{x\to c} f(x) = A$ and $\lim\limits_{x\to c} g(x) = B$.

  1. $\lim\limits_{x\to c}[f(x) + g(x)] = \lim\limits_{x\to c} f(x) + \lim\limits_{x\to c} g(x) = A + B$
  2. $\lim\limits_{x\to c} \alpha f(x) = \alpha \lim\limits_{x\to c} f(x) = \alpha A$
  3. $\lim\limits_{x\to c} f(x)g(x) = [\lim\limits_{x\to c} f(x)][\lim\limits_{x\to c} g(x)] = AB$

4. $\lim\limits_{x \to c} \frac{f(x)}{g(x)} = \frac{\lim\limits_{x \to c} f(x)}{\lim\limits_{x \to c} g(x)} = \frac{A}{B}$, provided $B \neq 0$

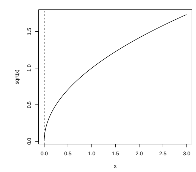
- Examples:

  1. $\lim\limits_{x \to 2}(2x - 3) = 2 \lim\limits_{x \to 2} x - 3 \lim\limits_{x \to 2} 1 = 2 \times 2 - 3 \times 1 = 1$
  2. $\lim\limits_{x \to c} x^n = [\lim\limits_{x \to c} x] \cdots [\lim\limits_{x \to c} x] = c \cdots c = c^n$

- Other types of limits:

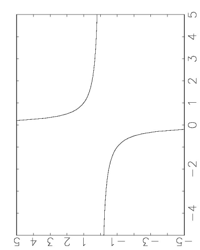
  1. **Right-hand limit:** $\lim\limits_{x \to c^+} f(x) = L$, if $c < x < c + \delta \Longrightarrow |f(x) - L| < \epsilon$

     Example: $\lim\limits_{x \to 0^+} \sqrt{x} = 0$

  2. **Left-hand limit:** $\lim\limits_{x \to c^-} f(x) = L$, if $c - \delta < x < c \Longrightarrow |f(x) - L| < \epsilon$
  3. **Infinity:** $\lim\limits_{x \to \infty} f(x) = L$, if $x > N \Longrightarrow |f(x) - L| < \epsilon$
  4. **$-$Infinity:** $\lim\limits_{x \to -\infty} f(x) = L$, if $x < -N \Longrightarrow |f(x) - L| < \epsilon$

     Example: $\lim\limits_{x \to \infty} 1/x = \lim\limits_{x \to -\infty} 1/x = 0$

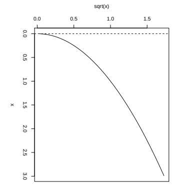
- **Caution**: In some situations, you will not be able to calculate a limit. For instance, $\lim_{x \to \infty} \frac{x}{-x}$. The numerator is headed towards ∞ while the denominator is headed towards $-\infty$. In this case the limit does not exist. In other circumstances, the limit may exist but additional steps need to be taken.

2 Continuity

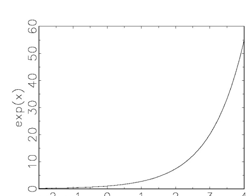
- **Continuity**: Suppose that the domain of the function f includes an open interval containing the point c. Then f is continuous at c if $\lim\limits_{x \to c} f(x)$ exists and if $\lim\limits_{x \to c} f(x) = f(c)$. Further, f is continuous on an open interval (a, b) if it is continuous at each point in the interval.

- Examples: Continuous functions.

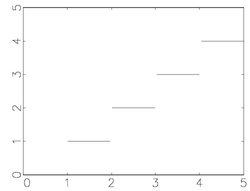
  $f(x) = \sqrt{x}$

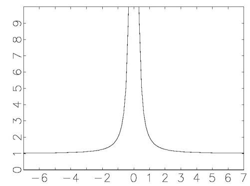
  

  $f(x) = e^x$

- Examples: Discontinuous functions.

$f(x) = \text{floor}(x)$ 

$f(x) = 1 + \frac{1}{x^2}$ 

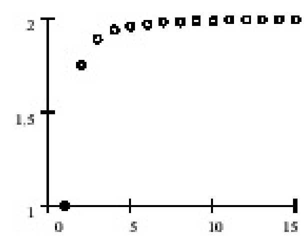
- Properties:

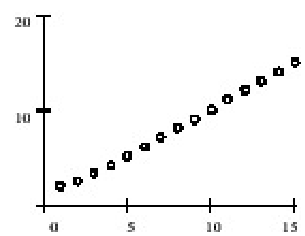
1. If f and g are continuous at point c, then $f+g$, $f-g$, fg, $|f|$, and αf are continuous. f/g is continuous, provided $g(c) \neq 0$.
2. **Boundedness**: If f is continuous on the closed bounded interval $[a,b]$, then there is a number K such that $|f(x)| \leq K$ for each x in $[a,b]$.
3. **Max/Min**: If f is continuous on the closed bounded interval $[a,b]$, then f has a maximum and a minimum on $[a,b]$, possibly at the end points.
4. The range of a closed bounded interval $[a,b]$ under a continuous function f is also a closed bounded interval $[m,M]$.

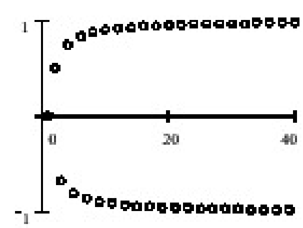
2.1 Sequences

- A **sequence** $\{y_n\} = \{y_1, y_2, y_3, \ldots, y_n\}$ is an ordered set of real numbers, where y_1 is the first term in the sequence and y_n is the nth term. Generally, a sequence extends to $n = \infty$. We can also write the sequence as $\{y_n\}_{n=1}^{\infty}$.

- Examples:

1. $\{y_n\} = \left\{2 - \frac{1}{n^2}\right\} = \left\{1, \frac{7}{4}, \frac{17}{9}, \frac{31}{16}, \ldots\right\}$ 

2. $\{y_n\} = \left\{\frac{n^2+1}{n}\right\} = \left\{2, \frac{5}{2}, \frac{10}{3}, \ldots\right\}$ 

3. $\{y_n\} = \left\{(-1)^n\left(1 - \frac{1}{n}\right)\right\} = \{0, \frac{1}{2}, -\frac{2}{3}, \frac{3}{4}, \ldots\}$ 

- Think of sequences like functions. Before, we had $y = f(x)$ with x specified over some domain. Now we have $\{y_n\} = \{f(n)\}$ with $n = 1, 2, 3, \ldots$.

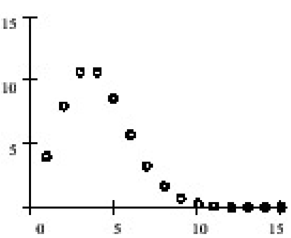
- Three kinds of sequences:

1. Sequences like 1 above that converge to a limit.
2. Sequences like 2 above that increase without bound.
3. Sequences like 3 above that neither converge nor increase without bound — alternating over the number line.

- Boundedness and monotonicity:
  1. **Bounded**: if $|y_n| \leq K$ for all n
  2. **Monotone Increasing**: $y_{n+1} > y_n$ for all n
  3. **Monotone Decreasing**: $y_{n+1} < y_n$ for all n
- **Subsequence**: choose an infinite collection of entries from $\{y_n\}$, retaining their order.

2.2 The Limit of a Sequence

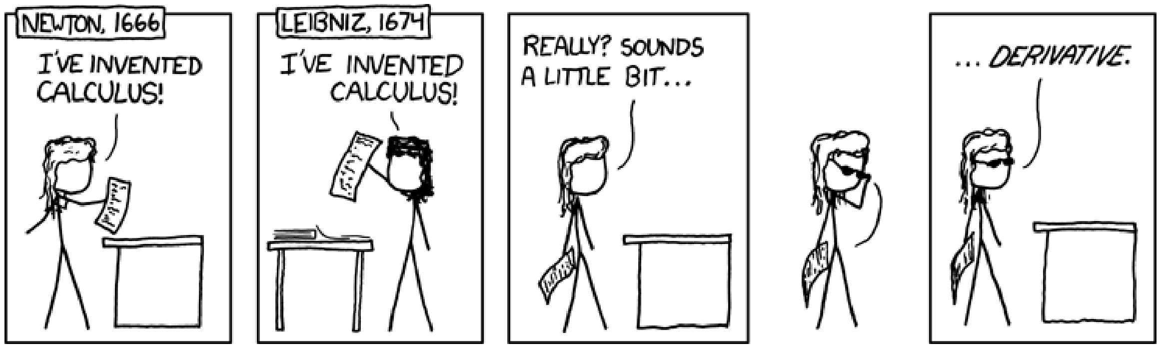
- We're often interested in whether a sequence **converges** to a **limit**. Limits of sequences are conceptually similar to the limits of functions addressed in the previous lecture.
- **Definition: (Limit of a sequence)**. The sequence $\{y_n\}$ has the limit L, that is $\lim_{n\to\infty} y_n = L$, if for any $\epsilon > 0$ there is an integer N (which depends on ϵ) with the property that $|y_n - L| < \epsilon$ for each $n > N$. $\{y_n\}$ is said to converge to L. If the above does not hold, then $\{y_n\}$ diverges.
- Examples:
  1. $\lim_{n\to\infty} \left\{2 - \frac{1}{n^2}\right\} = 2$
  2. $\lim_{n\to\infty} \left\{\frac{4^n}{n!}\right\} = 0$



- Uniqueness: If $\{y_n\}$ converges, then the limit L is unique.
- Properties: Let $\lim_{n\to\infty} y_n = A$ and $\lim_{n\to\infty} z_n = B$. Then
  1. $\lim_{n\to\infty} [\alpha y_n + \beta z_n] = \alpha A + \beta B$
  2. $\lim_{n\to\infty} y_n z_n = AB$
  3. $\lim_{n\to\infty} \frac{y_n}{z_n} = \frac{A}{B}$, provided $B \neq 0$
- Finding the limit of a sequence in $\mathbf{R}^n$ is similar to that in $\mathbf{R}^1$.
- **Limit of a sequence of vectors.** The sequence of vectors $\{\mathbf{y_n}\}$ has the limit $\mathbf{L}$, that is $\lim_{n\to\infty} \mathbf{y_n} = \mathbf{L}$, if for any ϵ there is an integer N where $||\mathbf{y_n} - \mathbf{L}|| < \epsilon$ for each $n > N$. The sequence of vectors $\{\mathbf{y_n}\}$ is said to converge to the vector $\mathbf{L}$ — and the distances between $\mathbf{y_n}$ and $\mathbf{L}$ converge to zero.
- Think of each coordinate of the vector $\mathbf{y_n}$ as being part of its own sequence over n. Then a sequence of vectors in $\mathbf{R}^n$ converges if and only if all n sequences of its components converge. Examples:

1. The sequence $\{y_n\}$ where $y_n = \left(\frac{1}{n}, 2 - \frac{1}{n^2}\right)$ converges to $(0, 2)$.
2. The sequence $\{y_n\}$ where $y_n = \left(\frac{1}{n}, (-1)^n\right)$ does not converge, since $\{(-1)^n\}$ does not converge.

3 Change we can believe in



Although you will be using matrix notation and linear algebra on a pretty regular basis (mostly in your statistics classes), encounters with basic concepts from calculus will be a daily event in your methods training. In the case of derivatives, you will primarily be using them to:

- Find the maximum and minimum values of specific functions including utility and likelihood functions;
- Derive a probability density function given its cumulative density function

This probably doesn't mean much to you at the moment. But what they have in common is that we are trying to understand a certain property of a function – the rate of change. Usually we want to find when the rate of change is zero, as that indicates either a maximum or minimum. These concepts were developed in the 17th century to understand physical motion. If a function describes the location of a baseball flying through the air, the first derivative describes the velocity of the ball, and the second derivative represents the acceleration.

3.1 Derivative

- The derivative of f at x is its rate of change at x — i.e., how much $f(x)$ changes with a change in x.
  - For a line, the derivative is the slope.
  - For a curve, the derivative is the tangent line at x.
- **Derivative**: Let f be a function whose domain includes an open interval containing the point x. The derivative of f at x is given by

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{(x+h) - x} \\ &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \end{aligned}$$

- Don't succumb to the temptation to take this equation as a given without any understanding. This idea is over 400 years old now, and is not beyond your abilities to understand.

  Imagine that you want to calculate how a function changes between time period 1 and period 2. With a little thought we would calculate this as:

  $$m = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

  This is same "rise over run" calculation we used to calculate the slope of a line in the first lecture. If we set $h = x_2 - x_1$ and $x = \frac{x_1 + x_2}{2}$ then we can re-parametrize this as:

  $$\begin{aligned} m &= \frac{f(x + h/2) - f(x - h/2)}{h} \\ &= \frac{f(x+h) - f(x)}{h} \end{aligned}$$

  What we want to calculate is the "instantaneous" rate of change. That is, we want to know the rate of change as x_1 and x_2 get closer and closer together such that time period between when we measure them moves towards zero. In other words, what happens as h gets very close to to zero. This is the conceptual leap that it takes a genius like Newton to discover, but which we can accept with a bit of thought.

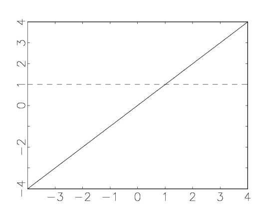
- If $f'(x)$ exists at a point x, then f is said to be **differentiable** at x. Similarly, if $f'(x)$ exists for every point along an interval, then f is differentiable along that interval. For f to be differentiable at x, f must be both continuous and "smooth" at x. The process of calculating $f'(x)$ is called **differentiation**.

- For historical reasons, there are multiple notations for derivatives.

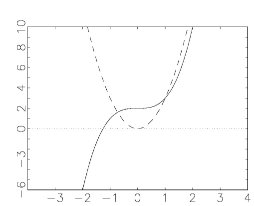
  1. y', $f'(x)$ (Prime or Lagrange Notation)
  2. $\frac{dy}{dx}$, $\frac{df}{dx}(x)$ (Leibniz's Notation)

- Examples:

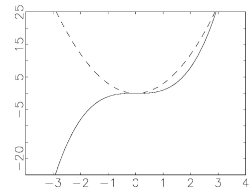
  1. $f(x) = c$
     $f'(x) = \lim\limits_{h \to 0} \frac{f(x+h)-f(x)}{h} = \lim\limits_{h \to 0} \frac{c-c}{h} = \lim\limits_{h \to 0} 0 = 0$



  2. $f(x) = x$
     $f'(x) = \lim\limits_{h \to 0} \frac{(x+h)-x}{h} = \lim\limits_{h \to 0} 1 = 1$



  3. $f(x) = x^2$
     $f'(x) = \lim\limits_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim\limits_{h \to 0} \frac{2xh + h^2}{h} = \lim\limits_{h \to 0} (2x + h) = 2x$



4. $f(x) = x^3$
   $f'(x) = \lim_{h \to 0} \frac{(x+h)^3 - x^3}{h} = \lim_{h \to 0} \frac{3x^2h + 3xh^2 + h^3}{h}$
   $= \lim_{h \to 0} \left(3x^2 + 3xh + h^2\right) = 3x^2$

- **Aside:** A function is *monotonically increasing* in a domain if it has a positive derivative over that domain. Likewise, it is *monotonically decreasing* if it has a negative derivative.

- **Existence**: $f'(x)$ at some particular value of x exits iff $f(x)$ is continuous at that value of x. $f'(x)$ exists for all values of x iff $f(x)$ is continuous at all values of x in the domain of $f(x)$. In other words, there must be no point where the right-hand derivative and the left-hand derivative are different. The basic test is "can you draw the function without lifting your pencil." The classic case where this does not work are:

  1. $$f(x) = |x|$$
  2. $$f(x) = \frac{1}{x-1}$$
  3. $$f(x) = 1, x \in [0, 1]$$
     $$f(x) = 0, x > 1 \text{ or } x < 0$$

- Doing this calculation for every function is cumbersome, and sometimes difficult. Fortunately, hundreds of mathematical graduate students and professors have developed several "rules" that help us calculate derivatives quickly. All of these rules come with proofs, but for our purposes we usually just memorize them and use them. The first (and easiest) is that the derivative of a constant is zero.

  Properties of derivatives: Suppose that f and g are differentiable at x and that α is a constant. Then the functions $f \pm g$, αf, fg, and f/g (provided $g(x) \neq 0$) are also differentiable at x. Additionally,

Power rule:	$[x^k]' = kx^{k-1}$
Sum rule:	$[f(x) \pm g(x)]' = f'(x) \pm g'(x)$
Constant rule:	$[\alpha f(x)]' = \alpha f'(x)$
Product rule:	$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$
Quotient rule:	$[f(x)/g(x)]' = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$, $g(x) \neq 0$

- Examples:

  1. $f(x) = 3x^2 + 2x^{1/3}$
     $f'(x) = (3x^2)' + (2x^{1/3})' = 3(2x) + 2(\frac{1}{3}x^{-2/3}) = 6x + \frac{2}{3}x^{-2/3}$

  2. $f(x) = (x^3)(2x^4)$
     $f'(x) = (x^3)'(2x^4) + (x^3)(2x^4)' = (3x^2)(2x^4) + (x^3)(8x^3) = 6x^6 + 8x^6 = 14x^6$
     or
     $f'(x) = (2x^7)' = 14x^6$

3. $f(x) = \frac{x^2+1}{x^2-1}$
   $f'(x) = \frac{(x^2+1)'(x^2-1)-(x^2+1)(x^2-1)'}{(x^2-1)^2} = \frac{2x(x^2-1)-(x^2+1)2x}{(x^2-1)^2} = \frac{-4x}{(x^2-1)^2}$

3.2 Higher-order derivatives

- One conceptual leap you need to make is that $f'(x)$ is just another function of x. An input value for x gives an output value determined by $f'(x)$. For some people it might help to change the name so that $h(x) = f'(x)$. The important thing to understand is that just as $f(x)$ can have a derivative, $f'(x)$ can also have have a derivative.
- We can keep applying the differentiation process to functions that are themselves derivatives. The derivative of $f'(x)$ with respect to x, would then be

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}$$

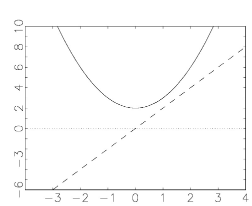
  and so on. Similarly, the derivative of $f''(x)$ would be denoted $f'''(x)$.

- **First derivative**: $f'(x)$, y', $\frac{df(x)}{dx}$, $\frac{dy}{dx}$
  Second derivative: $f''(x)$, y'', $\frac{d^2f(x)}{dx^2}$, $\frac{d^2y}{dx^2}$
  nth derivative: $\frac{d^nf(x)}{dx^n}$, $\frac{d^ny}{dx^n}$
- Example: $f(x) = x^3$, $f'(x) = 3x^2$, $f''(x) = 6x$, $f'''(x) = 6$, $f''''(x) = 0$
- Example:

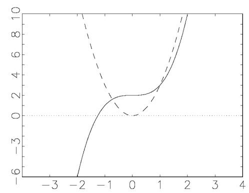
$$f(x) = 4x^4 + 12x^2 + x - 2$$
$$f'(x) = 16x^3 - 24x + 1$$
$$f''(x) = 48x^2 - 24$$
$$f'''(x) = 96x$$
$$f''''(x) = 96$$
$$f'''''(x) = 0$$

3.3 Maxima and minima

- The first derivative $f'(x)$ identifies whether the function $f(x)$ at the point x is
  1. **Increasing:** $f'(x) > 0$
  2. **Decreasing:** $f'(x) < 0$
  3. **Extremum/Saddle**: $f'(x) = 0$
- Examples:
  1. $f(x) = x^2 + 2$, $f'(x) = 2x$



2. $f(x) = x^3 + 2$, $f'(x) = 3x^2$



- The second derivative $f''(x)$ identifies whether the function $f(x)$ at the point x is
  1. **Concave down:** $f''(x) < 0$
  2. **Concave up:** $f''(x) > 0$

- **Maximum (Minimum)**: x_0 is a **local** maximum (minimum) if $f(x_0) > f(x)$ ($f(x_0) < f(x)$) for all x within some open interval containing x_0. x_0 is a **global** maximum (minimum) if $f(x_0) > f(x)$ ($f(x_0) < f(x)$) for all x in the domain of f.

- **Critical points**: Given the function f defined over domain D, all of the following are critical points:
  1. Any interior point of D where $f'(x) = 0$.
  2. Any interior point of D where $f'(x)$ does not exist.
  3. Any endpoint that is in D.

  The maxima and minima will be a subset of the critical points.

- Combined, the first and second derivatives can tell us whether a point is a maximum or minimum of $f(x)$.

  Local Maximum: $f'(x) = 0$ and $f''(x) < 0$
  Local Minimum: $f'(x) = 0$ and $f''(x) > 0$
  Need more info: $f'(x) = 0$ and $f''(x) = 0$

- **Global Maxima and Minima**. Sometimes no global max or min exists — e.g., $f(x)$ not bounded above or below. However, three situations where we can fairly easily identify global max or min.
  1. **Functions with only one critical point.** If x_0 is a local maximum of f and it is the only critical point, then it is a global maximum.
  2. **Globally concave up or concave down functions.** If f'' is never zero, then there is at most one critical point, which is a global maximum if $f'' < 0$ and a global minimum if $f'' > 0$.
  3. **Functions over closed and bounded intervals** must have both a global maximum and a global minimum.

- Examples:
  1. $f(x) = x^2 + 2$, $f'(x) = 2x$, $f''(x) = 2$

     $f'(x) = 2x = 0$ at $x = 0$ and it is the only critical point. Checking $f''(x)$, we see that $f''(x) > 0$ for all x. Therefore, $x = 0$ is a global minimum. We could also have noted that $f''(x)$ is globally concave up and concluded that $x = 0$ is a global minimum.

2. $f(x) = x^3 + 2$ $\quad f'(x) = 3x^2$ $\quad f''(x) = 6x$

   $f'(x) = 0$ at $x = 0$ and it is the only critical point. However, $f''(0) = 0$, so we need more information to determine if $x = 0$ is a maximum, minimum, or saddle. If we examined either the graph or values of $f''(x)$ around $x = 0$, we would find that $x = 0$ is in fact a saddle point. Since $f(x)$ is unbounded above or below, there are no maxima or minima.

3.4 Chain rule

- **Composite functions** are formed by substituting one function into another and are denoted by
$$(f \circ g)(x) = f[g(x)]$$
To form $f[g(x)]$, the range of g must be contained (at least in part) within the domain of f. The domain of $f \circ g$ consists of all the points in the domain of g for which $g(x)$ is in the domain of f.

- Examples:

  1. $f(x) = \ln x,$ $\quad 0 < x < \infty$
     $g(x) = x^2$ $\quad -\infty < x < \infty$
     $(f \circ g)(x) = \ln x^2,$ $\quad -\infty < x < \infty - \{0\}$
     $(g \circ f)(x) = [\ln x]^2,$ $\quad 0 < x < \infty$
     Notice that $f \circ g$ and $g \circ f$ are not the same functions.
  2. $f(x) = 4 + \sin x,$ $\quad -\infty < x < \infty$
     $g(x) = \sqrt{1 - x^2},$ $\quad -1 \leq x \leq 1$
     $(f \circ g)(x) = 4 + \sin\sqrt{1 - x^2},$ $\quad -1 \leq x \leq 1$
     $(g \circ f)(x)$ does not exist, since the range of f, $[3, 5]$, has no points in common with the domain of g.

- **Chain Rule**: Let $y = f(z)$ and $z = g(x)$. Then, $y = (f \circ g)(x) = f[g(x)]$ and the derivative of y with respect to x is
$$\frac{d}{dx}\{f[g(x)]\} = f'[g(x)]g'(x)$$
which can also be written as
$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$$
(Note: the above does not imply that the dz's cancel out, as in fractions. They are part of the derivative notation and have no separate existence.) The chain rule can be thought of as the derivative of the "outside" times the derivative of the "inside," remembering that the derivative of the outside function is evaluated at the value of the inside function.

- **Generalized Power Rule**: If $y = [g(x)]^k$, then $dy/dx = k[g(x)]^{k-1}g'(x)$.

- Examples:

1. Find dy/dx for $y = (3x^2 + 5x - 7)^6$. Let $f(z) = z^6$ and $z = g(x) = 3x^2 + 5x - 7$. Then, $y = f[g(x)]$ and

$$\begin{aligned}\frac{dy}{dx} &= f'(z)g'(x) \\ &= \left(6z^5\right)(6x+5) \\ &= 6\left(3x^2+5x-7\right)^5(6x+5)\end{aligned}$$

2. Find dy/dx for $y = \sin(x^3 + 4x)$. (Note: the derivative of $\sin x$ is $\cos x$.) Let $f(z) = \sin z$ and $z = g(x) = x^3 + 4x$. Then, $y = f[g(x)]$ and

$$\begin{aligned}\frac{dy}{dx} &= f'(z)g'(x) \\ &= (\cos z)\left(3x^2+4\right) \\ &= \cos\left(x^3+4x\right)\left(3x^2+4\right)\end{aligned}$$

3.5 Derivatives of Exp and Ln

- **Derivatives of Exp**:
  1. $\frac{d}{dx}\alpha e^x = \alpha e^x$
  2. $\frac{d^n}{dx^n}\alpha e^x = \alpha e^x$
  3. $\frac{d}{dx}e^{u(x)} = e^{u(x)}u'(x)$
- Examples: Find dy/dx for
  1. $y = e^{-3x}$ Let $u(x) = -3x$. Then $u'(x) = -3$ and $dy/dx = -3e^{-3x}$.
  2. $y = e^{x^2}$ Let $u(x) = x^2$. Then $u'(x) = 2x$ and $dy/dx = 2xe^{x^2}$.
  3. $y = e^{\sin 2x}$ Let $u(x) = \sin 2x$. Then $u'(x) = 2\cos 2x$ and $dy/dx = (2\cos 2x)e^{\sin 2x}$.
- **Derivatives of Ln**:
  1. $\frac{d}{dx}\ln x = \frac{1}{x}$
  2. $\frac{d}{dx}\ln x^k = \frac{d}{dx}k\ln x = \frac{k}{x}$
  3. $\frac{d}{dx}\ln u(x) = \frac{u'(x)}{u(x)}$ (by the chain rule)
- Examples: Find dy/dx for
  1. $y = \ln(x^2 + 9)$ Let $u(x) = x^2 + 9$. Then $u'(x) = 2x$ and $dy/dx = u'(x)/u(x) = 2x/(x^2 + 9)$.
  2. $y = \ln(\ln x)$ Let $u(x) = \ln x$. Then $u'(x) = 1/x$ and $dy/dx = 1/(x\ln x)$.
  3. $y = (\ln x)^2$ Use the generalized power rule. $dy/dx = (2\ln x)/x$.
  4. $y = \ln e^x$ (We know that $\ln e^x = x$ and that $dx/dx = 1$, but let's double check.) Let $u(x) = e^x$. Then $u'(x) = e^x$ and $dy/dx = u'(x)/u(x) = e^x/e^x = 1$.

3.6 L'Hospital's Rule

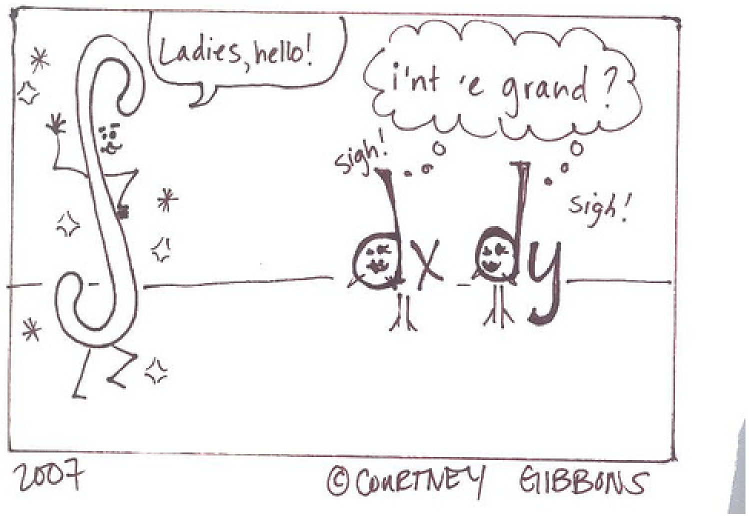
- In studying limits, we saw that $\lim_{x\to c} f(x)/g(x) = \left(\lim_{x\to c} f(x)\right) / \left(\lim_{x\to c} g(x)\right)$, provided that $\lim_{x\to c} g(x) \neq 0$, which will cause the limit to be unbounded.
- If both $\lim_{x\to c} f(x) = 0$ and $\lim_{x\to c} g(x) = 0$, then we get an **indeterminate form** of the type 0/0 as $x \to c$. However, we can still analyze such limits using L'Hospital's rule.
- **L'Hospital's Rule**: Suppose f and g are differentiable on $a < x < b$ and that either
  1. $\lim_{x\to a^+} f(x) = 0$ and $\lim_{x\to a^+} g(x) = 0$, or
  2. $\lim_{x\to a^+} f(x) = \pm\infty$ and $\lim_{x\to a^+} g(x) = \pm\infty$

  Suppose further that $g'(x)$ is never zero on $a < x < b$ and that
  $$\lim_{x\to a^+} \frac{f'(x)}{g'(x)} = L$$
  then
  $$\lim_{x\to a^+} \frac{f(x)}{g(x)} = L$$
- Examples: Use L'Hospital's rule to find the following limits:
  1. $\lim_{x\to 0^+} \frac{\ln(1+x^2)}{x^3}$ Let $f(x) = \ln(1+x^2)$ and $g(x) = x^3$. Then $f'(x) = 2x/(1+x^2)$ and $g'(x) = 3x^2$. Using L'Hospital's rule, $\lim_{x\to 0^+} \frac{2x/(1+x^2)}{3x^2} = \lim_{x\to 0^+} \frac{2}{3x(1+x^2)} = \infty$.
  2. $\lim_{x\to 0^+} \frac{e^{1/x}}{1/x}$ Let $f(x) = e^{1/x}$ and $g(x) = 1/x$. Then $f'(x) = -\frac{1}{x^2}e^{1/x}$ and $g'(x) = -1/x^2$. Using L'Hospital's rule, $\lim_{x\to 0^+} \frac{-\frac{1}{x^2}e^{1/x}}{-1/x^2} = \lim_{x\to 0^+} e^{1/x} = \infty$.
  3. $\lim_{x\to 2} \frac{x-2}{(x+6)^{1/3}-2}$ Let $f(x) = x-2$ and $g(x) = (x+6)^{1/3} - 2$. Then $f'(x) = 1$ and $g'(x) = \frac{1}{3}(x+6)^{-2/3}$. Using L'Hospital's rule, $\lim_{x\to 2} \frac{1}{\frac{1}{3}(x+6)^{-2/3}} = 3(8)^{2/3} = 12$.

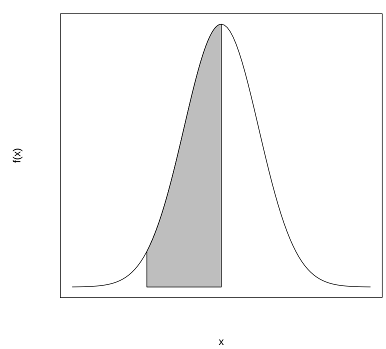
3.7 Newton-Raphson and Taylor Series

- For various reasons, derivatives can be difficult to write down. This is particularly true in the kinds of higher-dimensional settings we discuss below. The *Newton-Raphson* and *Taylor Series expansion* approaches are sometimes used in these situations to find the value when $f'(x) = 0$ or at some value x_1 respectively.
- For reasons of time, we are not going to cover this in great detail, but more advanced students may want to carefully look at section 6.4.2 in the book and/or consult more advanced calculus texts. In particular, you are very likely to see a Taylor series expansion in the next couple of years.

4 The area under a curve



If you were trying to find the area of a rectangle or in a trapezoid, you would have no problem. But what if you were trying to find the area under a curve like this?



This is an important question, because the above function is a drawing of the normal distribution – the most commonly used probability function in all of statistics. The above picture is essentially asking the question: what is the probability (i.e., $f(x)$) of observing a value of x between -2 and 0?

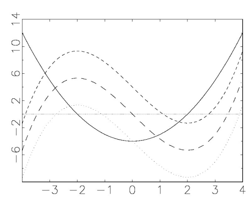
The answer, as it turns out, is sometimes very difficult to get. Nonetheless it is very important. You will be doing *a lot* of integration in statistics (especially if you venture into Bayesian statistics). And there are many, many applications of this technique in game theory. But in general, you will *always* be using these techniques for one of these two goals:

- Finding the area under a curve
- Finding a function given its derivative.

Be aware that integration is sometimes (usually?) *hard*. Sometimes it is impossible. There are many important functions (e.g., the normal probability density function) whose indefinite integral has never been derived. Bayesian statistics was held back for hundreds of years by the difficulties of integrating until computational methods such as MCMC were developed and refined in the 1990s that made it possible to get answers without actually solving. Don't expect solutions to integrals to jump off of the page at you. Focus on understanding the basic concept, and starting develop a library of "tricks" that mathematicians frequently use to solve these kinds of problems.

4.1 The Indefinite Integral: The Antiderivative

- Sometimes we're interested in exactly the reverse: finding the function f for which g is its derivative. We refer to f as the **antiderivative** of g.
- Let DF be the derivative of F. And let $DF(x)$ be the derivative of F evaluated at x. Then the antiderivative is denoted by D^{-1} (i.e., the inverse derivative). If $DF = f$, then $F = D^{-1}f$.
- **Indefinite Integral**: Equivalently, if F is the antiderivative of f, then F is also called the indefinite integral of f and written $F(x) = \int f(x)dx$.
- Examples:
  1. $\int \frac{1}{x^2}dx = -\frac{1}{x} + c$
  2. $\int 3e^{3x}dx = e^{3x} + c$
  3. $\int (x^2 - 4)dx = \frac{1}{3}x^3 - 4x + c$



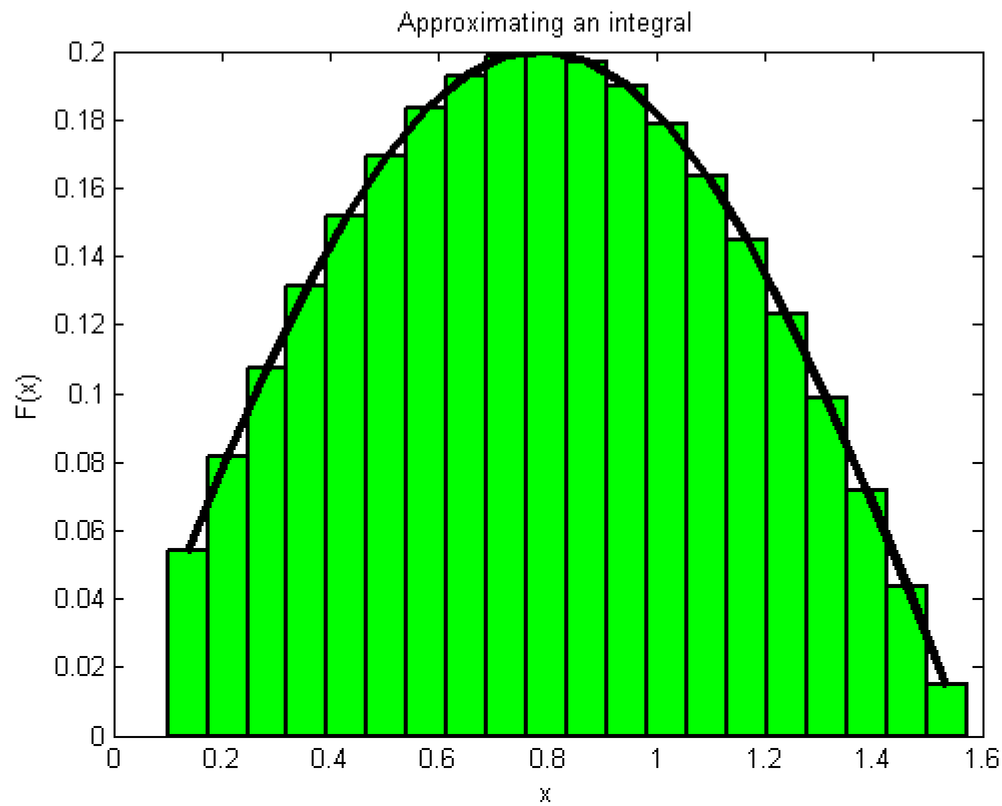
- Notice from these examples that while there is only a single derivative for any function, there are multiple antiderivatives: one for any arbitrary constant c. c just shifts the curve up or down on the y-axis. If more info is present about the antiderivative — e.g., that it passes through a particular point — then we can solve for a specific value of c.
- Common rules of integration:
  1. $\int af(x)dx = a\int f(x)dx$
  2. $\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$
  3. $\int x^n dx = \frac{1}{n+1}x^{n+1} + c$
  4. $\int e^x dx = e^x + c$
  5. $\int \frac{1}{x}dx = \ln x + c$
  6. $\int e^{f(x)}f'(x)dx = e^{f(x)} + c$
  7. $\int [f(x)]^n f'(x)dx = \frac{1}{n+1}[f(x)]^{n+1} + c$
  8. $\int \frac{f'(x)}{f(x)}dx = \ln f(x) + c$
- Examples:
  1. $\int 3x^2dx = 3\int x^2dx = 3\left(\frac{1}{3}x^3\right) + c = x^3 + c$
  2. $\int (2x + 1)dx = \int 2xdx + \int 1dx = x^2 + x + c$
  3. $\int e^x e^{e^x}dx = e^{e^x} + c$

4.2 The Definite Integral: The Area under the Curve

- **Riemann Sum**: Suppose we want to determine the area $A(R)$ of a region R defined by a curve $f(x)$ and some interval $a \leq x \leq b$. One way to calculate the area would be to divide the interval $a \leq x \leq b$ into n subintervals of length Δx and then approximate the region with a series of rectangles, where the base of each rectangle is Δx and the height is $f(x)$ at the midpoint of that interval. $A(R)$ would then be approximated by the area of the union of the rectangles, which is given by

$$S(f, \Delta x) = \sum_{i=1}^{n} f(x_i)\Delta x$$

and is called a Riemann sum.



- As we decrease the size of the subintervals Δx, making the rectangles "thinner," we would expect our approximation of the area of the region to become closer to the true area. This gives the limiting process

$$A(R) = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i)\Delta x$$

- **Riemann Integral**: If for a given function f the Riemann sum approaches a limit as $\Delta x \to 0$, then that limit is called the Riemann integral of f from a to b. Formally,

$$\int_a^b f(x)dx = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i)\Delta x$$

- **Definite Integral**: We use the notation $\int_a^b f(x)dx$ to denote the definite integral of f from a to b. In words, the definite integral $\int_a^b f(x)dx$ is the area under the "curve" f(x) from $x = a$ to $x = b$.

- **First Fundamental Theorem of Calculus**: Let the function f be bounded on $[a, b]$ and continuous on (a, b). Then the function

$$F(x) = \int_a^x f(s)ds, \quad a \leq x \leq b$$

has a derivative at each point in (a, b) and

$$F'(x) = f(x), \quad a < x < b$$

This last point shows that differentiation is the inverse of integration.

- **Second Fundamental Theorem of Calculus**: Let the function f be bounded on $[a, b]$ and continuous on (a, b). Let F be any function that is continuous on $[a, b]$ such that $F'(x) = f(x)$ on (a, b). Then

$$\int_a^b f(x)dx = F(b) - F(a)$$

- Procedure to calculate a "simple" definite integral $\int_a^b f(x)dx$:

  1. Find the indefinite integral $F(x)$.
  2. Evaluate $F(b) - F(a)$.

- Examples:

  1. $\int_1^3 3x^2 dx = 3\left(\frac{1}{3}x^3\right)\Big|_1^3 = (3)^3 - (1)^3 = 26$
  2. $\int_{-2}^2 e^x e^{e^x} dx = e^{e^x}\Big|_{-2}^2 = e^{e^2} - e^{e^{-2}} = 1617.033$

- Properties of Definite Integrals:

  1. $\int_a^a f(x)dx = 0$ — There is no area below a point.
  2. $\int_a^b f(x)dx = -\int_b^a f(x)dx$ — Reversing the limits changes the sign of the integral.
  3. $\int_a^b [\alpha f(x) + \beta g(x)]dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx$
  4. $\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$

- Examples:

1. $\int_1^1 3x^2dx = x^3\big|_1^1 = (1)^3 - (1)^3 = 0$
2. $\int_0^4 (2x+1)dx = 2\int_0^4 xdx + \int_0^4 1dx = x^2\big|_0^4 + x|_0^4 = (16-0)+(4-0) = 20$
3. $\int_{-2}^0 e^x e^{e^x} dx + \int_0^2 e^x e^{e^x} dx = e^{e^x}\big|_{-2}^0 + e^{e^x}\big|_0^2 = e^{e^0} - e^{e^{-2}} + e^{e^2} - e^{e^0} = e^{e^2} - e^{e^{-2}} = 1617.033$

4.3 Integration by Substitutions

- Sometimes the integrand doesn't appear integrable using common rules and antiderivatives. A method one might try is **integration by substitutions**, which is related to the Chain Rule.
- Suppose we want to find the indefinite integral $\int g(x)dx$ and assume we can identify a function $u(x)$ such that $g(x) = f[u(x)]u'(x)$. Let's refer to the antiderivative of f as F. Then the chain rule tells us that $\frac{d}{dx}F[u(x)] = f[u(x)]u'(x)$. So, $F[u(x)]$ is the antiderivative of g. We can then write

$$\int g(x)dx = \int f[u(x)]u'(x)dx = \int \frac{d}{dx}F[u(x)]dx = F[u(x)] + c$$

- Procedure to determine the indefinite integral $\int g(x)dx$ by the method of substitutions:
  1. Identify some part of $g(x)$ that might be simplified by substituting in a single variable u (which will then be a function of x).
  2. Determine if $g(x)dx$ can be reformulated in terms of u and du.
  3. Solve the indefinite integral.
  4. Substitute back in for x
- Substitution can also be used to calculate a definite integral. Using the same procedure as above,

$$\int_a^b g(x)dx = \int_c^d f(u)du = F(d) - F(c)$$

  where $c = u(a)$ and $d = u(b)$.

- Examples:
  1. $\int x^2\sqrt{x+1}dx$

     The problem here is the $\sqrt{x+1}$ term. However, if the integrand had $\sqrt{x}$ times some polynomial, then we'd be in business. Let's try $u = x+1$. Then $x = u-1$ and $dx = du$. Substituting these into the above equation, we get

$$\begin{aligned}\int x^2\sqrt{x+1}dx &= \int (u-1)^2\sqrt{u}du \\ &= \int (u^2-2u+1)u^{1/2}du \\ &= \int (u^{5/2} - 2u^{3/2} + u^{1/2})du\end{aligned}$$

We can easily integrate this, since it's just a polynomial. Doing so and substituting $u = x + 1$ back in, we get

$$\int x^2\sqrt{x+1}dx = 2(x+1)^{3/2}\left[\frac{1}{7}(x+1)^2 - \frac{2}{5}(x+1) + \frac{1}{3}\right] + c$$

2. For the above problem, we could have also used the substitution $u = \sqrt{x+1}$. Then $x = u^2 - 1$ and $dx = 2udu$. Substituting these in, we get

$$\int x^2\sqrt{x+1}dx = \int (u^2-1)^2 u 2udu$$

which when expanded is again a polynomial and gives the same result as above.

3. $\int\limits_0^1 \frac{5e^{2x}}{(1+e^{2x})^{1/3}}dx$

When an expression is raised to a power, it's often helpful to use this expression as the basis for a substitution. So, let $u = 1 + e^{2x}$. Then $du = 2e^{2x}dx$ and we can set $5e^{2x}dx = 5du/2$. Additionally, $u = 2$ when $x = 0$ and $u = 1 + e^2$ when $x = 1$. Substituting all of this in, we get

$$\begin{aligned}\int\limits_0^1 \frac{5e^{2x}}{(1+e^{2x})^{1/3}}dx &= \frac{5}{2}\int\limits_2^{1+e^2} \frac{du}{u^{1/3}} \\ &= \frac{5}{2}\int\limits_2^{1+e^2} u^{-1/3}du \\ &= \frac{15}{4}u^{2/3}\Big|_2^{1+e^2} \\ &= 9.53\end{aligned}$$

4.4 Integration by Parts

- Another useful integration technique is **integration by parts**, which is related to the Product Rule of differentiation. The product rule states that

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

Integrating this and rearranging, we get

$$\int u\frac{dv}{dx}dx = uv - \int v\frac{du}{dx}dx$$

or

$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx$$

More frequently remembered as

$$\int udv = uv - \int vdu$$

where $du = u'(x)dx$ and $dv = v'(x)dx$.

- For definite integrals: $\int_a^b u\frac{dv}{dx}dx = uv|_a^b - \int_a^b v\frac{du}{dx}dx$
- Our goal here is to find expressions for u and dv that, when substituted into the above equation, yield an expression that's more easily evaluated.
- Examples:

1. $\int xe^{ax}dx$

   Let $u = x$ and $dv = e^{ax}dx$. Then $du = dx$ and $v = (1/a)e^{ax}$. Substituting this into the integration by parts formula, we obtain

$$\begin{aligned}\int xe^{ax}dx &= uv - \int vdu \\ &= x\left(\frac{1}{a}e^{ax}\right) - \int \frac{1}{a}e^{ax}dx \\ &= \frac{1}{a}xe^{ax} - \frac{1}{a^2}e^{ax} + c\end{aligned}$$

2. $\int x^n e^{ax}dx$

   As in the first problem, let's let $u = x^n$ and $dv = e^{ax}dx$. Then $du = nx^{n-1}dx$ and $v = (1/a)e^{ax}$. Substituting these into the integration by parts formula gives

$$\begin{aligned}\int x^n e^{ax}dx &= uv - \int vdu \\ &= x^n\left(\frac{1}{a}e^{ax}\right) - \int \frac{1}{a}e^{ax}nx^{n-1}dx \\ &= \frac{1}{a}x^n e^{ax} - \frac{n}{a}\int x^{n-1}e^{ax}dx\end{aligned}$$

   Notice that we now have an integral similar to the previous one, but with x^{n-1} instead of x^n. For a given n, we would repeat the integration by parts procedure until the integrand was directly integrable — e.g., when the integral became $\int e^{ax}dx$.

3. $\int x^3 e^{-x^2}dx$

   We could, as before, choose $u = x^3$ and $dv = e^{-x^2}dx$. But we can't then find v — i.e., integrating $e^{-x^2}dx$ isn't possible. Instead, notice that $d(e^{-x^2})/dx = -2xe^{-x^2}$, which can be factored out of the original integrand

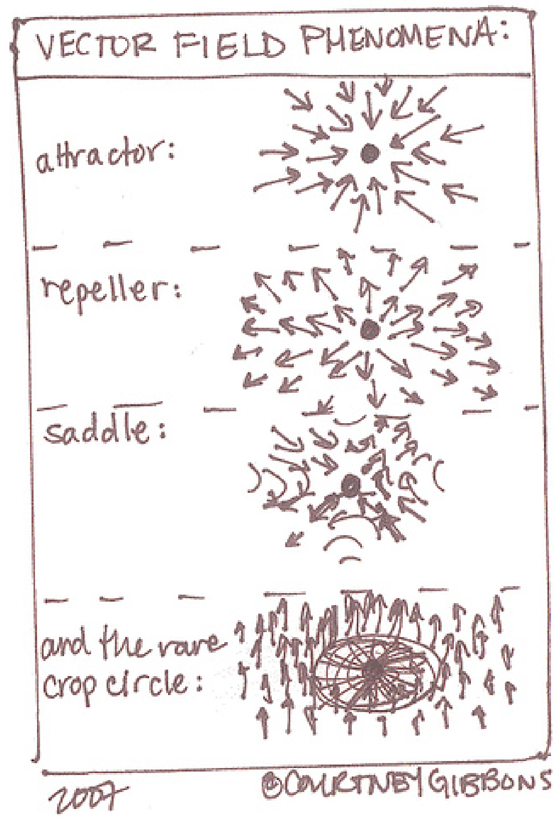
$$\int x^3 e^{-x^2}dx = \int x^2(xe^{-x^2})dx$$

   We can then let $u = x^2$ and $dv = xe^{-x^2}dx$. Then $du = 2xdx$ and $v = -\frac{1}{2}e^{-x^2}$. Substituting these in, we have

$$\begin{aligned}\int x^3 e^{-x^2}dx &= uv - \int vdu \\ &= x^2\left(-\frac{1}{2}e^{-x^2}\right) - \int\left(-\frac{1}{2}e^{-x^2}\right)2xdx \\ &= -\frac{1}{2}x^2e^{-x^2} + \int xe^{-x^2}dx \\ &= -\frac{1}{2}x^2e^{-x^2} - \frac{1}{2}e^{-x^2} + c\end{aligned}$$

and the rest is algebra.

5 Calculus with a matrix



This is where things get cool (but also a bit tricky). Just like we often have too many equations and variables to deal with efficiently in pieces for basic algebraic operations, we will often want a better way to do calculus. In these situations, very smart people have developed matrix methods for handling differentiation equivalent to first (gradients) and second (Hessians) derivatives. This is pretty useful for finding global maxima and minimum in multi-dimensional spaces.

Unfortunately, there aren't a set of equally clean methods for dealing with multidimensional integration. That's good news for you today (since you don't have to learn about it), but bad news for the rest of your life since you will spend a lot of your time working with imperfect numerical approximations of high-dimensional integrals.

5.1 Differentiation in several variables

- Suppose we have a function f now of two (or more) variables and we want to determine the rate of change relative to one of the variables. To do so, we would find its partial derivative, which is defined similar to the derivative of a function of one variable.

- **Partial Derivative**: Let f be a function of the variables $(x_1, \ldots, x_n)$. The partial derivative of f with respect to x_i is

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}$$

  Only the ith variable changes — the others are treated as constants.

- We can take higher-order partial derivatives, like we did with functions of a single variable, except now we the higher-order partials can be with respect to multiple variables.

- Examples:

  1. $f(x, y) = x^2 + y^2$
     $\frac{\partial f}{\partial x}(x, y) = 2x$
     $\frac{\partial f}{\partial y}(x, y) = 2y$
     $\frac{\partial^2 f}{\partial x^2}(x, y) = 2$
     $\frac{\partial^2 f}{\partial x \partial y}(x, y) = 0$
  2. $f(x, y) = x^3 y^4 + e^x - \ln y$
     $\frac{\partial f}{\partial x}(x, y) = 3x^2 y^4 + e^x$
     $\frac{\partial f}{\partial y}(x, y) = 4x^3 y^3 - 1/y$
     $\frac{\partial^2 f}{\partial x^2}(x, y) = 6xy^4 + e^x$
     $\frac{\partial^2 f}{\partial x \partial y}(x, y) = 12x^2 y^3$

5.2 Integration with several variables

- Now suppose we want to reverse the process. Say we have a function f now of two (or more) variables and we want to determine the area under the surface (or hypersurface).

  Take the total function and choose one of the variables (although you will want to be strategic about this choice). Perform the integration, while *treating the other variable as a constant.* Make sure you keep track of the ∂x_i symbols.

$$\int \int (2x + 2y) \partial x \partial y = \int x^2 + 2xy \partial y + c$$
$$= yx^2 + xy^2 + ?$$

- This is not as straight-forward as it seems because indefinite integrals are only correct up to a constant. For instance:

$$\int x^3 = \frac{1}{4} x^4 + 10$$

  or

$$= \frac{1}{4} x^4 - 10$$

  or (if we are treating y as a constant)

$$= \frac{1}{4} x^4 - e^y$$

- Likewise:

$$\int\int 12x^2y^3\partial x\partial y$$

$$= x^3y^4 + e^x - ln(y)$$

or

$$= x^3y^4 + 24x - y^{monkey}$$

- Usually we get around this by making use of the fact that we know that function we are working with integrates to a known constant or when the integrated function passes through a specific point. For instance, we know that any probability function must integrate to 1. But often, even this doesn't help that much when we are integrating many times across many dimensions.

- None of this is particularly important for anything you will be doing soon. Just something to keep in mind as you work towards more advanced methods.

5.3 A quick look at Constraint Optimization with Lagrange Multipliers

- When solving formal models or problems in microeconomics we often want to minimize of maximize functions, for example your utility function! Yet, often in life we can't get everything we want. And so we need to constrain our optimization.

- Take the function $f(x) = x^2 + y^2$. We want to find the maximum but to the constraint that $h(x) \equiv x + 2x = 10$

- Write out the Lagrangian $L(x, y, \lambda) = x^2 + y^2 - \lambda(x + 2y - 1)$

- Now we take the derivatives with respect to x, y, λ and set each of the partials equal to zero

  1. $\frac{\partial L}{\partial x} = 2x - \lambda = 0$
  2. $\frac{\partial L}{\partial y} = 2y - 2\lambda = 0$
  3. $\frac{\partial L}{\partial \lambda} = -(x + 2y - 10) = 0$

- From the first equation: $x = \frac{\lambda}{2}$

- From the second equation: $y = \lambda$, thus $y = 2x$

- Thus: $-(x + 2(2x) - 10) = 0$

- now we can solve for y, x, and λ!

- $x = 2$, $y = 4$, $\lambda = 4$, so max $f(x, y) = 18$ s.t. $x + 2y = 10$

5.4 Vector representation of calculus

- The function $y = f(x_1, x_2, \ldots, x_n)$ of the independent variable $x_1, x_2, \ldots, x_n$ can be written as the function $y = f(\mathbf{x})$ where $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$.

- The **gradient** is the vector of partial derivatives and is denoted:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

- The **Hessian** $\mathbf{H}(\mathbf{x})$ is an $n \times n$ matrix, where the (i, j)th element is the second order partial derivative of $f(\mathbf{x})$ with respect to x_i and x_j:

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

5.5 Maxima and Minima in $\mathbf{R}^n$

- **Conditions for Extrema**: The conditions for extrema are similar to those for functions on $\mathbf{R}^1$. Let $f(\mathbf{x})$ be a function of n variables. Let $B(\mathbf{x}, \epsilon)$ be the ϵ-ball about the point $\mathbf{x}$. Then

  1. $f(\mathbf{x}^*) > f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon) \implies$ Strict Local Max
  2. $f(\mathbf{x}^*) \geq f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon) \implies$ Local Max
  3. $f(\mathbf{x}^*) < f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon) \implies$ Strict Local Min
  4. $f(\mathbf{x}^*) \leq f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon) \implies$ Local Min

5.6 First Order Conditions

- When we examined functions of one variable x, we found critical points by taking the first derivative, setting it to zero, and solving for x. For functions of n variables, the critical points are found in much the same way, except now we set the partial derivatives equal to zero.[†]
- $\mathbf{x}^*$ is a critical point iff $\nabla f(\mathbf{x}^*) = 0$.
- Example: Find the critical points of $f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + 1$

  1. The partial derivatives of $f(\mathbf{x})$ are

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial x_1} &= 2(x_1 - 1) \\ \frac{\partial f(\mathbf{x})}{\partial x_2} &= 2x_2 \end{aligned}$$

  2. Setting each partial equal to zero and solving for x_1 and x_2, we find that there's a critical point at $\mathbf{x}^* = (1, 0)$.

[†]We will only consider critical points on the interior of a function's domain.

5.7 Second Order Conditions

- When we found a critical point for a function of one variable, we used the second derivative as an indicator of the curvature at the point in order to determine whether the point was a min, max, or saddle. For functions of n variables, we use second order partial derivatives as an indicator of curvature.

- **Curvature and The Taylor Polynomial as a Quadratic Form**: The Hessian is used in a Taylor polynomial approximation to $f(\mathbf{x})$ and provides information about the curvature of $f(\mathbf{x})$ at $\mathbf{x}$ — e.g., which tells us whether a critical point $\mathbf{x}^*$ is a min, max, or saddle point.

  1. The second order Taylor polynomial about the critical point $\mathbf{x}^*$ is

  $$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T \mathbf{H}(\mathbf{x}^*)\mathbf{h} + R(\mathbf{h})$$

  2. Since we're looking at a critical point, $\nabla f(\mathbf{x}^*) = 0$; and for small $\mathbf{h}$, $R(\mathbf{h})$ is negligible. Rearranging, we get

  $$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^T \mathbf{H}(\mathbf{x}^*)\mathbf{h}$$

  3. The RHS is a quadratic form and we can determine the definiteness of $\mathbf{H}(\mathbf{x}^*)$.

     (a) If $\mathbf{H}(\mathbf{x}^*)$ is positive definite, then the RHS is positive for all small $\mathbf{h}$:

     $$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) > 0 \quad \Longrightarrow \quad f(\mathbf{x}^* + \mathbf{h}) > f(\mathbf{x}^*)$$

     i.e., $f(\mathbf{x}^*) < f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$, so $\mathbf{x}^*$ is a strict local min.

     (b) Conversely, if $\mathbf{H}(\mathbf{x}^*)$ is negative definite, then the RHS is negative for all small $\mathbf{h}$:

     $$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) < 0 \quad \Longrightarrow \quad f(\mathbf{x}^* + \mathbf{h}) < f(\mathbf{x}^*)$$

     i.e., $f(\mathbf{x}^*) > f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$, so $\mathbf{x}^*$ is a strict local max.

- **Summary of Second Order Conditions**:

  Given a function $f(\mathbf{x})$ and a point $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = 0$,

  1. $\mathbf{H}(\mathbf{x}^*)$ Positive Definite $\Longrightarrow$ Strict Local Min
  2. $\mathbf{H}(\mathbf{x})$ Positive Semidefinite $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$ $\Longrightarrow$ Local Min
  3. $\mathbf{H}(\mathbf{x}^*)$ Negative Definite $\Longrightarrow$ Strict Local Max
  4. $\mathbf{H}(\mathbf{x})$ Negative Semidefinite $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$ $\Longrightarrow$ Local Max
  5. $\mathbf{H}(\mathbf{x}^*)$ Indefinite $\Longrightarrow$ Saddle Point

- Example: We found that the only critical point of $f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + 1$ is at $\mathbf{x}^* = (1, 0)$. Is it a min, max, or saddle point?

1. Recall that the gradient of $f(\mathbf{x})$ is

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2(x_1 - 1) \\ 2x_2 \end{pmatrix}$$

Then the Hessian is

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

2. To check the definiteness of $\mathbf{H}(\mathbf{x}^*)$, we could use either of two methods:
   (a) Determine whether $\mathbf{x^T H}(\mathbf{x}^*)\mathbf{x}$ is greater or less than zero for all $\mathbf{x} \neq \mathbf{0}$:

$$\mathbf{x^T H}(\mathbf{x}^*)\mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1^2 + 2x_2^2$$

   For any $\mathbf{x} \neq \mathbf{0}$, $2(x_1^2 + x_2^2) > 0$, so the Hessian is positive definite and $\mathbf{x}^*$ is a strict local minimum.

   (b) Using the method of leading principal minors, we see that $M_1 = 2$ and $M_2 = 4$. Since both are positive, the Hessian is positive definite and $\mathbf{x}^*$ is a strict local minimum.

- Does this seem confusing? What it mean for a matrix to be "definite"?

5.8 Redux: Definiteness

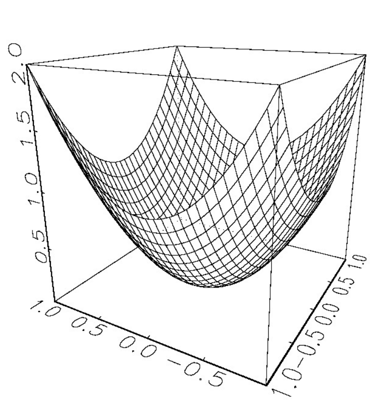
- **A positive definite matrix is always nonsingular**. When you come across this term, it is usually there to specify that the matrix can be inverted, and a solution to some system of equations is possible.
- When some $n \times n$ matrix $\mathbf{A}$ is pre- and post-multiplied by a conformable non-zero matrix $\mathbf{x}$, we get the equation:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = c$$

- Some properties of the quadratic. For all nonzero vectors $\mathbf{x}$:
  - $\mathbf{A}$ is said to be **positive definite** if $c > 0$.
  - $\mathbf{A}$ is said to be **positive semidefinite** if $c \geq 0$.
  - $\mathbf{A}$ is said to be **negative definite** if $c < 0$.
  - $\mathbf{A}$ is said to be **negative semidefinite** if $c \leq 0$.
- Examples:

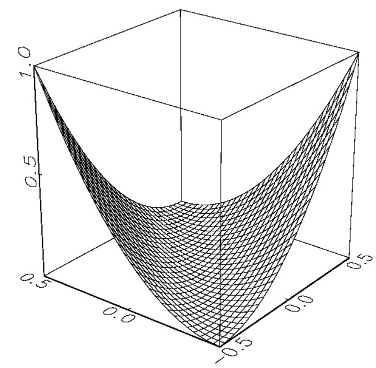
1. Positive Definite:

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x} \\ &= x_1^2 + x_2^2 \end{aligned}$$



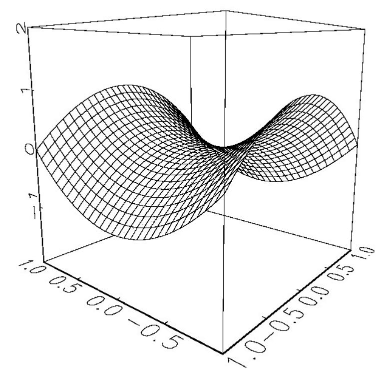
2. Positive Semidefinite:

$$
\begin{aligned}
Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \mathbf{x} \\
&= (x_1 - x_2)^2
\end{aligned}
$$



3. Indefinite:

$$
\begin{aligned}
Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{x} \\
&= x_1^2 - x_2^2
\end{aligned}
$$



5.9 Redux: Test for definiteness

- Given an $n \times n$ matrix $\mathbf{A}$, kth order **principal minors** are the determinants of the $k \times k$ submatrices along the diagonal obtained by deleting $n-k$ columns and the same $n-k$ rows from $\mathbf{A}$.
- Example: For a 3×3 matrix $\mathbf{A}$,
  1. First order principal minors:
  $$|a_{11}|, \quad |a_{22}|, \quad |a_{33}|$$
  2. Second order principal minors:
  $$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$
  3. Third order principal minor: $|\mathbf{A}|$
- Define the kth **leading principal minor** M_k as the determinant of the $k \times k$ submatrix obtained by deleting the last $n-k$ rows and columns from $\mathbf{A}$.
- Example: For a 3×3 matrix $\mathbf{A}$, the three leading principal minors are
$$M_1 = |a_{11}|, \quad M_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad M_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$
- Algorithm: If $\mathbf{A}$ is an $n \times n$ symmetric matrix, then
  1. $M_k > 0, \ k = 1, \ldots, n \qquad \Longrightarrow \qquad$ Positive Definite

2. $M_k < 0$, for odd k and $M_k > 0$, for even k $\Longrightarrow$ Negative Definite
3. $M_k \neq 0$, $k = 1, \ldots, n$, but does not fit the pattern of 1 or 2. $\Longrightarrow$ Indefinite.

- If some leading principal minor is zero, but all others fit the pattern of the preceding conditions 1 or 2, then

  1. Every principal minor ≥ 0 $\Longrightarrow$ Positive Semidefinite
  2. Every principal minor of odd order ≤ 0 and every principal minor of even order ≥ 0 $\Longrightarrow$ Negative Semidefinite

5.10 Global maxima and minima

- To determine whether a critical point is a global min or max, we can check the concavity of the function over its entire domain. Here again we use the definiteness of the Hessian to determine whether a function is globally concave or convex:

  1. $\mathbf{H}(\mathbf{x})$ Positive Semidefinite $\forall \mathbf{x}$ $\Longrightarrow$ Globally Convex
  2. $\mathbf{H}(\mathbf{x})$ Negative Semidefinite $\forall \mathbf{x}$ $\Longrightarrow$ Globally Concave

  Notice that the definiteness conditions must be satisfied over the entire domain.

- Given a function $f(\mathbf{x})$ and a point $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = 0$,

  1. $f(\mathbf{x})$ Globally Convex $\Longrightarrow$ Global Min
  2. $f(\mathbf{x})$ Globally Concave $\Longrightarrow$ Global Max

- Note that showing that $\mathbf{H}(\mathbf{x}^*)$ is negative semidefinite is not enough to guarantee $\mathbf{x}^*$ is a local max. However, showing that $\mathbf{H}(\mathbf{x})$ is negative semidefinite for all $\mathbf{x}$ guarantees that x^* is a global max. (The same goes for positive semidefinite and minima.)

- Example: Take $f_1(x) = x^4$ and $f_2(x) = -x^4$. Both have $x = 0$ as a critical point. Unfortunately, $f_1''(0) = 0$ and $f_2''(0) = 0$, so we can't tell whether $x = 0$ is a min or max for either. However, $f_1''(x) = 12x^2$ and $f_2''(x) = -12x^2$. For all x, $f_1''(x) \geq 0$ and $f_2''(x) \leq 0$ — i.e., $f_1(x)$ is globally convex and $f_2(x)$ is globally concave. So $x = 0$ is a global min of $f_1(x)$ and a global max of $f_2(x)$.

5.11 Example (thanks Harvard!)

- Given $f(\mathbf{x}) = x_1^3 - x_2^3 + 9x_1x_2$, find any maxima or minima.

  1. First-order conditions. Set the gradient equal to zero and solve for x_1 and x_2.

  $$\begin{aligned} \frac{\partial f}{\partial x_1} &= 3x_1^2 + 9x_2 = 0 \\ \frac{\partial f}{\partial x_2} &= -3x_2^2 + 9x_1 = 0 \end{aligned}$$

  We have two equations in two unknowns. Solving for x_1 and x_2, we get two critical points: $\mathbf{x}_1^* = (0, 0)$ and $\mathbf{x}_1^* = (3, -3)$.

2. Second order conditions. Determine whether the Hessian is positive or negative definite. The Hessian is

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 6x_1 & 9 \\ 9 & -6x_2 \end{pmatrix}$$

   Evaluated at $\mathbf{x}_1^*$,

$$\mathbf{H}(\mathbf{x_1^*}) = \begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix}$$

   The two leading principal minors are $M_1 = 0$ and $M_2 = -81$, so $\mathbf{H}(\mathbf{x_1^*})$ is indefinite and $\mathbf{x}_1^* = (0, 0)$ is a saddle point.

   Evaluated at $\mathbf{x}_2^*$,

$$\mathbf{H}(\mathbf{x_2^*}) = \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}$$

   The two leading principal minors are $M_1 = 18$ and $M_2 = 243$. Since both are positive, $\mathbf{H}(\mathbf{x_2^*})$ is positive definite and $\mathbf{x}_2^* = (3, -3)$ is a strict local min.

3. Global concavity/convexity. In evaluating the Hessians for $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ we saw that the Hessian is not everywhere positive semidefinite. Hence, we can't infer that $\mathbf{x}_2^* = (3, -3)$ is a global minimum. In fact, if we set $x_1 = 0$, the $f(\mathbf{x}) = -x_2^3$, which will go to $-\infty$ as $x_2 \to \infty$.