

ECPR Winter School in Methods and Techniques, 11-18 February 2012

Additional Topics in Scalar and Vector Calculus

JEFF GILL, Professor

Department of Political Science, College of Arts & Sciences

Division of Biostatistics, School of Medicine

Department of Surgery (Division of Public Health Sciences), School of Medicine

Washington University, St. Louis

Objectives

- ▶ The material here extends the basic calculus we have already done.
- ▶ These topics include more advanced uses of derivatives like partial derivatives and higher order partial derivatives; root finding (locating an important point along some function of interest); analyzing function minima, maxima, and inflection points (points where derivatives change); integrals on functions of multiple variables; and, finally, the idea of an abstract series.
- ▶ Importantly, we also turn to functions of multiple variables are handled with different operations.

Partial Derivatives

- ▶ A *partial derivative* is a regular derivative, except that the operation is performed on a function of multiple variables where the derivative is taken only with respect to one of them and the others are treated as constants.

- ▶ We start with the generic function

$$u = f(x, y, z)$$

specified for the three variables, x , y , and z .

- ▶ Define the three partial derivatives

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial x} f(x, y, z) = g_1(x, y, z)$$

$$\frac{\partial u}{\partial y} = \frac{\partial}{\partial y} f(x, y, z) = g_2(x, y, z)$$

$$\frac{\partial u}{\partial z} = \frac{\partial}{\partial z} f(x, y, z) = g_3(x, y, z),$$

- ▶ The resulting functions are labeled as $g_i()$, $i = 1, 2, 3$, to indicate that not only are they different from the original function, they are also distinct from each other because different variables are manipulated with the derivative processes.

Partial Derivatives

- ▶ As an example, consider the function

$$f(x, y, z) = x \exp(y) \log(z),$$

- ▶ This has the three partial derivatives

$$\frac{\partial}{\partial x} f(x, y, z) = \exp(y) \log(z) \frac{\partial}{\partial x} x = \exp(y) \log(z)$$

$$\frac{\partial}{\partial y} f(x, y, z) = x \log(z) \frac{\partial}{\partial y} \exp(y) = x \exp(y) \log(z)$$

$$\frac{\partial}{\partial z} f(x, y, z) = x \exp(y) \frac{\partial}{\partial z} \log(z) = x \exp(y) \frac{1}{z}.$$

- ▶ Notice that the variables that are not part of the derivation process are simply treated as constants in these operations.

Partial Derivatives

- We can also evaluate a more complex function with additional variables:

$$f(u_1, u_2, u_3, u_4, u_5) = u_1^{u_2 u_3} \sin\left(\frac{u_1}{2}\right) \log\left(\frac{u_4}{u_5}\right) u_3^2,$$

- This produces

$$\frac{\partial}{\partial u_1} f(u_1, u_2, u_3, u_4, u_5) = \left[u_1^{u_2 u_3} \log\left(\frac{u_4}{u_5}\right) u_3^2 \right] \left[\frac{u_2 u_3}{u_1} \sin\left(\frac{u_1}{2}\right) + \frac{1}{2} \cos\left(\frac{u_1}{2}\right) \right]$$

$$\frac{\partial}{\partial u_2} f(u_1, u_2, u_3, u_4, u_5) = \log(u_1) u_1^{u_2 u_3} \sin\left(\frac{u_1}{2}\right) \log\left(\frac{u_4}{u_5}\right) u_3^3$$

$$\frac{\partial}{\partial u_3} f(u_1, u_2, u_3, u_4, u_5) = u_1^{u_2 u_3} \sin\left(\frac{u_1}{2}\right) \log\left(\frac{u_4}{u_5}\right) u_3 [u_2 \log(u_1) u_3 + 2]$$

$$\frac{\partial}{\partial u_4} f(u_1, u_2, u_3, u_4, u_5) = u_1^{u_2 u_3} \sin\left(\frac{u_1}{2}\right) \frac{u_3^2}{u_4}$$

$$\frac{\partial}{\partial u_5} f(u_1, u_2, u_3, u_4, u_5) = -u_1^{u_2 u_3} \sin\left(\frac{u_1}{2}\right) \frac{u_3^2}{u_5}.$$

- These results make extensive use of the handy reexpression: $a^x = (e^{\log(a)})^x = e^{x \log(a)}$.

Example: The Relationship Between Age and Earnings

- ▶ Rees and Schultz (1970) analyzed the relationship between workers' ages and their earnings with the idea that earnings increase early in a worker's career but tend to decrease toward retirement.
- ▶ For maintenance electricians, they posited an additive relationship for income (in units of \$1,000):

$$\begin{aligned} \text{Income} = & \beta_0 + \beta_1(\text{Seniority}) + \beta_2(\text{School.Years}) \\ & + \beta_3(\text{Experience}) + \beta_4(\text{Training}) \\ & + \beta_5(\text{Commute.Distance}) + \beta_6(\text{Age}) + \beta_7(\text{Age}^2), \end{aligned}$$

where the β values are scalar values for how much each factor individually affects *Income*.

- ▶ Since *Age* and *Age*² are both included in the analysis, the effect of the workers' age is parabolic, which is exactly as the authors intended: rising early, cresting, and then falling back.
- ▶ Taking the first partial derivative with respect to age, we get

$$\frac{\partial}{\partial \text{Age}} \text{Income} = \beta_6 + 2\beta_7 \text{Age} = (0.031) + 2(-0.00032)\text{Age}.$$

- ▶ Thus at age 25 the incremental effect of one additional year is $0.031 + 2(-0.00032)(25) = 0.015$, at age 50 it is $0.031 + 2(-0.00032)(50) = -0.001$, and at 75 it is $0.031 + 2(-0.00032)(75) = -0.017$.

Indexing Socio-Economic Status (SES)

- ▶ An early measure of socio-economic status by Gordon (1969) reevaluated conventional compilation of various social indicators as a means of measuring “the position of the individual in some status ordering as determined by the individual’s characteristics—his education, income, position, in the community, the market place, etc.”
- ▶ The idea is to combine separate collected variables into a single measure because using any one single measure does not fully provide an overall assessment of status.
- ▶ The means by which these indicators are combined can vary considerably: additive or multiplicative, scaled or unscaled, weighted according to some criteria, and so on.
- ▶ Gordon proposed an alternative multiplicative causal expression of the form for an individual’s socio-economic status:

$$SES = AE^b I^c M^d,$$

where

A = all terms not explicitly included in the model

E = years of education

I = amount of income

M = percent of time employed annually

and b, c, d are the associated “elasticities” providing a measure of strength for each causal term.

Indexing Socio-Economic Status (SES)

- ▶ The marginal (i.e., incremental) impacts on SES for each of the three terms of interest are given by partial derivatives with respect to the term of interest:

$$\frac{\partial SES}{\partial E} = AbE^{b-1}I^cM^d$$

$$\frac{\partial SES}{\partial I} = AcE^bI^{c-1}M^d$$

$$\frac{\partial SES}{\partial M} = AdE^bI^cM^{d-1}.$$

- ▶ Individual effects can be pulled from the general multiplicative specification with the partial derivatives.
- ▶ Note that due to the multiplicative nature of the proposed model, the marginal impacts are still dependent on levels of the other variables in a way that a strictly additive model would not produce: $SES = A + E^b + I^c + M^d$.

Derivatives and Partial Derivatives of Higher Order

- ▶ Derivatives of higher order than one iterated applications of the derivative process.
- ▶ The *second derivative* of a function with respect to x is

$$\frac{\partial^2 f(x)}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} f(x) \right),$$

also denoted $f''(x)$ for functions of only one variable since $f''(x, y)$ would be ambiguous.

- ▶ It follows then how to denote, calculate, and notate higher order derivatives:

$$\frac{\partial^3 f(x)}{\partial x^3}, \quad \frac{\partial^4 f(x)}{\partial x^4}, \quad \frac{\partial^5 f(x)}{\partial x^5},$$

(i.e., $f'''(x)$, $f''''(x)$, and $f'''''(x)$, although this notation gets cumbersome).

- ▶ Note the convention with respect to the order designation here, and that it differs by placement in the numerator (∂^5) and denominator (∂x^5).

Example: Derivatives of Higher Order

- For instance, given the function $f(x) = 3x^3$,

$$\frac{\partial f(x)}{\partial x} = 9x^2$$

$$\frac{\partial^4 f(x)}{\partial x^4} = 0$$

$$\frac{\partial^2 f(x)}{\partial x^2} = 18x$$

$$\frac{\partial^5 f(x)}{\partial x^5} = 0$$

$$\frac{\partial^3 f(x)}{\partial x^3} = 18$$

⋮

- Notice that we “run out” of derivatives eventually, and all derivatives of order four or higher remain zero.

Interpretation of Derivatives of Higher Order

- ▶ Because the first derivative is the rate of change of the original function, the second derivative is the rate of change for the rate of that change, and so on.
- ▶ Consider the simple example of the velocity of a car.
- ▶ The first derivative describes the rate of change of this velocity: very high when first starting out from a traffic light, and very low when cruising on the highway.
- ▶ The second derivative describes how fast this rate of change is changing.
- ▶ The second derivative is very high early in the car's path as it increases speed rapidly but is low under normal cruising conditions.
- ▶ Third-order and higher derivatives work in exactly this same way on respective lower orders, but the interpretation is less intuitive.

Higher Order Partial Derivatives

- ▶ Given the function $f(x, y) = 3x^3y^2$, we can calculate

$$\frac{\partial^2}{\partial x \partial y} f(x, y) = \frac{\partial}{\partial y} (9x^2y^2) = 18x^2y$$

and

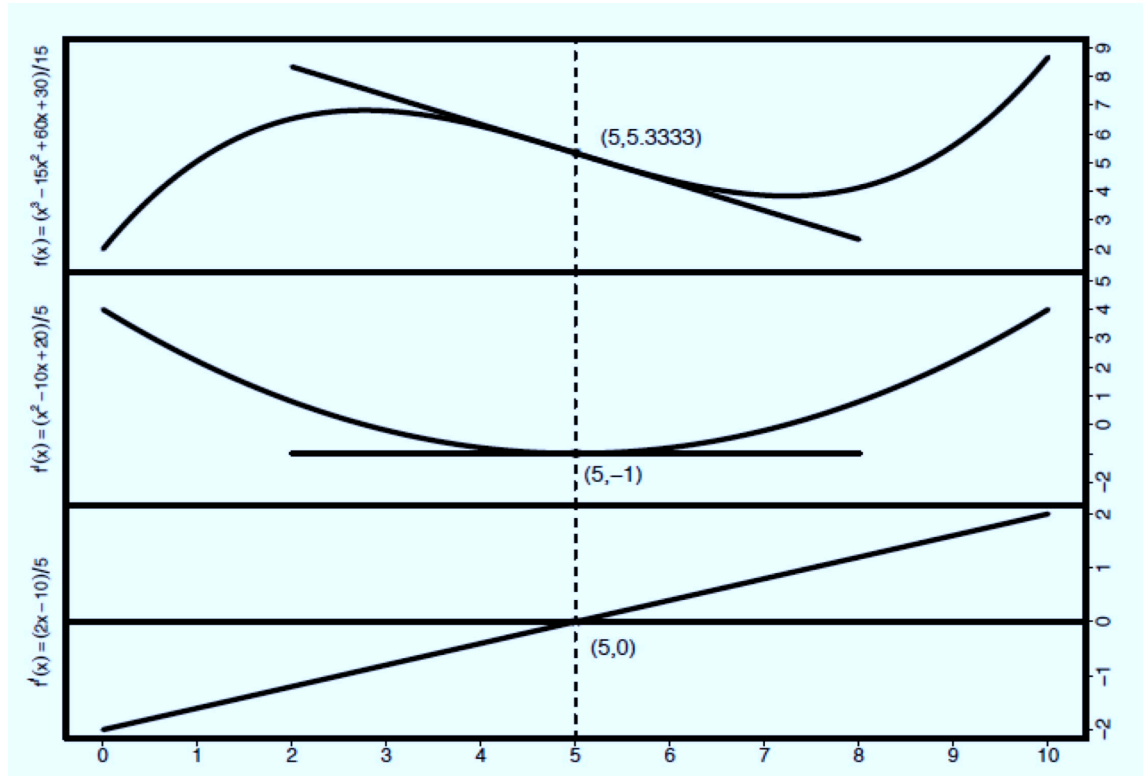
$$\frac{\partial^3}{\partial x^2 \partial y} f(x, y) = \frac{\partial^2}{\partial x \partial y} (9x^2y^2) = \frac{\partial}{\partial y} (18xy^2) = 36xy.$$

- ▶ Thus the order hierarchy in the denominator gives the “recipe” for how many derivatives of each variable to perform.
- ▶ The sequence of operations does not change the answer and therefore should be done in an order that makes the problem as easy as possible.
- ▶ There are other higher order partial derivatives that could be calculated for this function.
- ▶ In fact, if a function has k variables each of degree n , the number of derivatives of order n is given by $\binom{n+k-1}{n}$.

Maxima, Minima, and Root Finding

- ▶ Derivatives can be used to find points of interest along a given function.
- ▶ One point of interest is the point where the “curvature” of the function changes, given by the following definition:
- ▶ **[Inflection Point.]** For a given function, $y = f(x)$, a point (x^*, y^*) is called an *inflection point* if the second derivative immediately on one side of the point is signed oppositely to the second derivative immediately on the other side.
- ▶ So if x^* is indeed an inflection point, then for some small interval around x^* , $[x^* - \delta, x^* + \delta]$, $f''(x)$ is positive on one side and negative on the other.
- ▶ If $f''(x)$ is continuous at the inflection point $f(x^*)$, then $f''(x^*) = 0$.
- ▶ This makes intuitive sense since on one side $f'(x)$ is increasing so that $f''(x)$ must be positive, and on the other side $f'(x)$ is decreasing so that $f''(x)$ must be negative.
- ▶ Making δ arbitrarily small and decreasing it toward zero, x^* is the point where δ vanishes and the second derivative is neither positive nor negative and therefore must be zero.

Illustrating the Inflection Point



Illustrating the Inflection Point

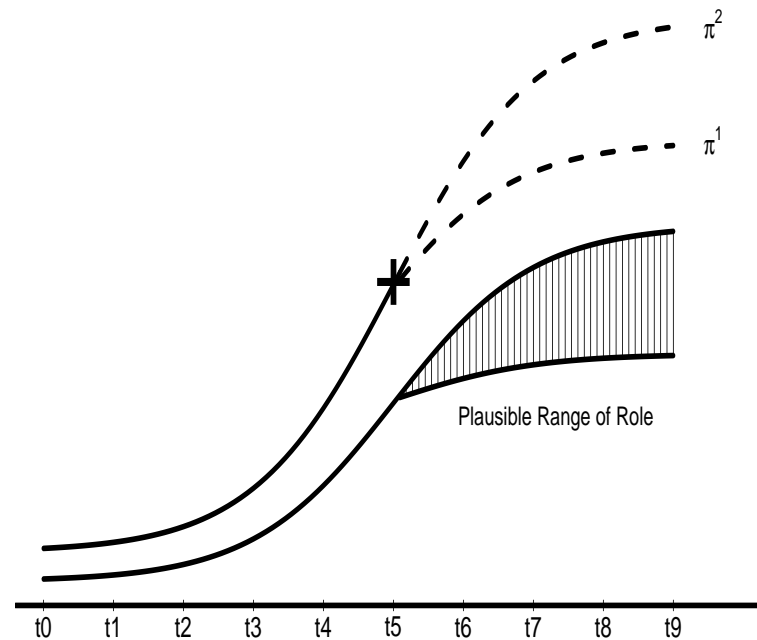
- ▶ The tangent line at the inflection point crosses the function such that it is on one side before the point and on the other side afterward.
- ▶ This is a consequence of the change of sign of the second derivative of the function and is illustrated in the figure with the function $f(x) = (x^3 - 15x^2 + 60x + 30)/15$.
- ▶ The function (characteristically) curves away from the tangent line on one side and curves away in the opposite direction on the other side.
- ▶ In the first panel of this figure the function itself is plotted with the tangent line shown.
- ▶ In the second panel the first derivative function, $f'(x) = (x^2 - 10x + 20)/5$, is plotted with the tangent line shown at the corresponding minima.
- ▶ Finally the third panel shows the second derivative function, $f''(x) = (2x - 10)/5$, with a horizontal line where the function crosses zero on the y -axis.
- ▶ Having a zero second derivative is a necessary but not sufficient condition for being an inflection point since the function $f'(5)$ has a zero second derivative, but $f'''(5)$ does not change sign just around 5.

Power Cycle Theory

- ▶ Sometimes inflection points can be substantively helpful in analyzing politics.
- ▶ Doran (1989) looked at critical changes in relative power between nations, evaluating *power cycle theory*, which asserts that war is caused by changes in the gap between state power and state interest.
- ▶ The key is that when dramatic differences emerge between relative power (capability and prestige) and systematic foreign policy role (current interests developed or allowed by the international system), existing balances of power are disturbed.
- ▶ Doran particularly highlighted the case where a nation ascendent in both power and interest, with power in excess of interest, undergoes a “sudden violation of the prior trend” in the form of an inflection point in their power curve.

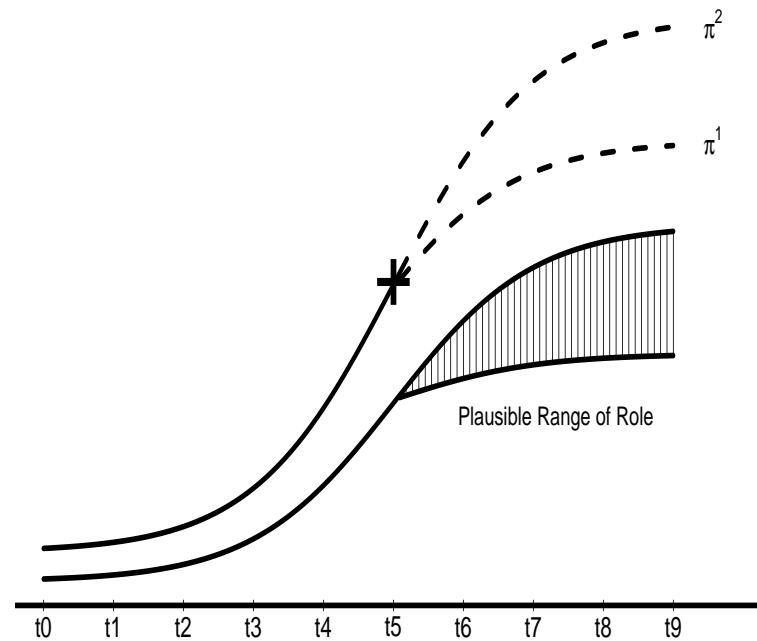
Power/Role Gap Changes Induced by the Inflection Point

This is illustrated here where the inflection point at time t_5 in the power curve introduces uncertainty in the anticipated decline in increasing power (i.e., the difference in π^1 and π^2 in the figure) and the subsequent potential differences between power and the range of plausible roles.



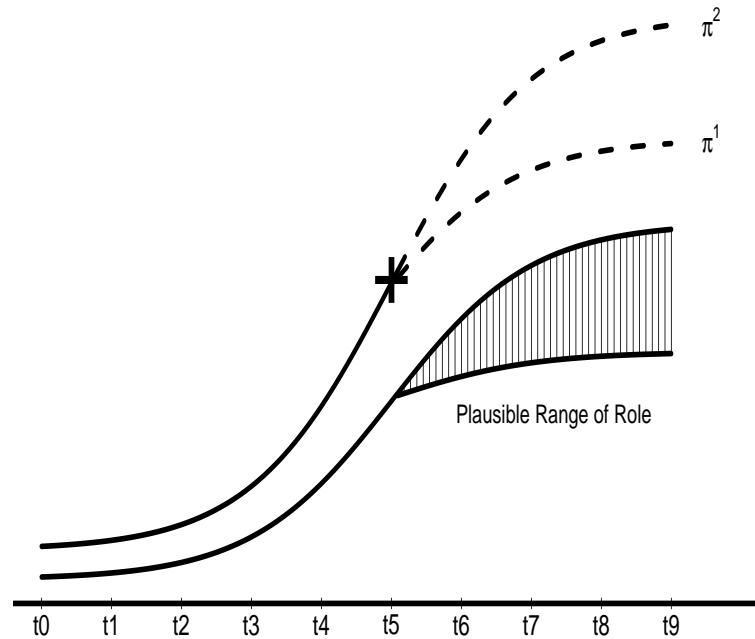
Power/Role Gap Changes Induced by the Inflection Point

Thus the problem introduced by the inflection point is that it creates changes in the slope difference and thus changes in the gap between role and power that the state and international system must account for.



Power/Role Gap Changes Induced by the Inflection Point

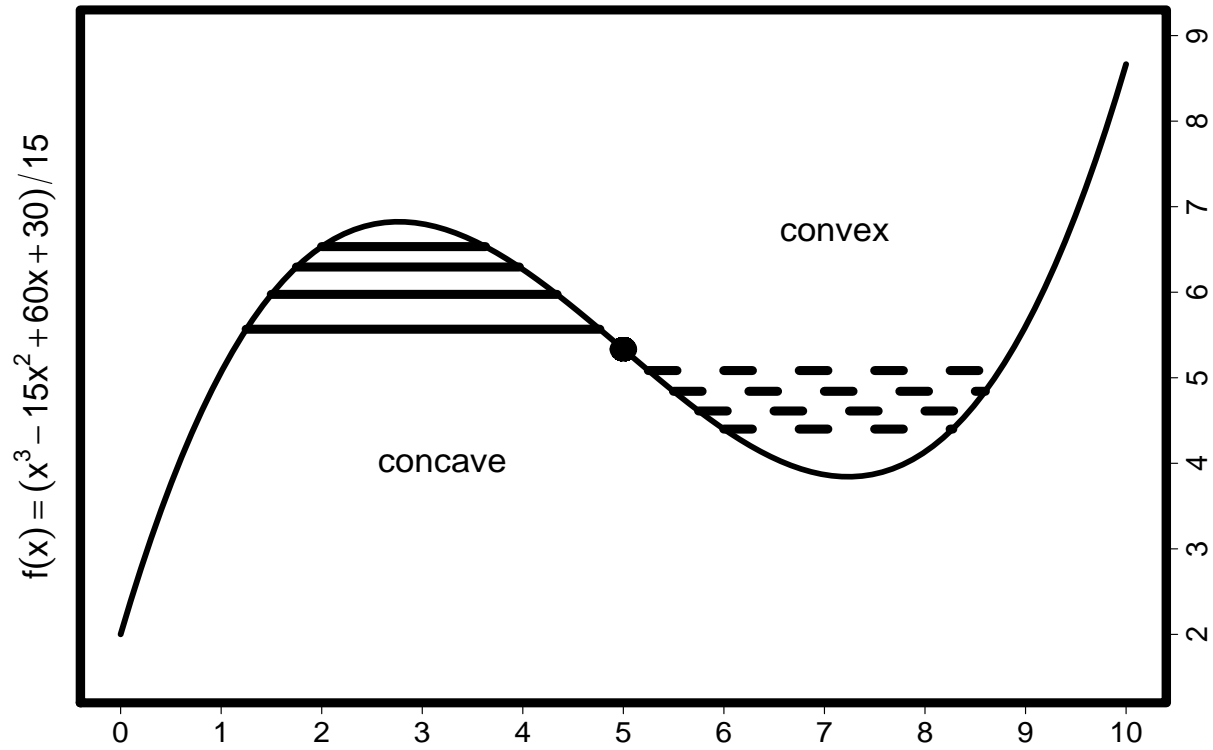
The function $f(x) = (x^3 - 15x^2 + 60x + 30)/15$ has a single inflection point in the illustrated range because there is one point where the concave portion of the function meets the convex portion.



Concave and Convex

- ▶ If a function (or portion of a function) is *convex* (also called *concave upward*), then every possible chord is above the function, except for their endpoints.
- ▶ A *chord* is just a line segment connecting two points along a curve.
- ▶ If a function (or portion of a function) is *concave* (or *concave downward*), then every possible chord is below the function, except for their endpoints.

Concave and Convex Portions of a Function



Formal Definition of Concave and Convex

- ▶ If a function $f(x)$ is twice differentiable and concave over some open interval, then $f''(x) \leq 0$, and if a function $f(x)$ is twice differentiable and convex over some open interval then $f''(x) \geq 0$.
- ▶ The reverse statement is also true: A twice differentiable function that has a nonpositive second derivative over some interval is concave over that interval, and a twice differentiable function that has a nonnegative derivative over some interval is convex over that interval.
- ▶ So the sign of the second derivative gives a handy test.

Evaluating Zero-Derivative Points

- ▶ Repeated application of the derivative also gives us another general test: the *second derivative test*.
- ▶ Points where the first derivatives are equal to zero are either a maxima or a minima of the function (modes or anti-modes), but without graphing how would we know which?
- ▶ Visually it is clear that the rate of change of the first derivative declines, moving away from a relative maximum point in both directions, and increases, moving away from a relative minimum point.
- ▶ The term “relative” here reinforces that these functions may have multiple maxima and minima, and we mean the behavior in a small neighborhood around the point of interest.
- ▶ This rate of change observation of the first derivative means that we can test with the second derivative: If the second derivative is negative at a point where the first derivative is zero, then the point is a relative maximum, and if the second derivative is positive at such a point, then the point is a relative minimum.

Multiple Modes

- ▶ Higher order polynomials often have more than one mode.
- ▶ For example, we can evaluate the function

$$f(x) = \frac{1}{4}x^4 - 2x^3 + \frac{11}{2}x^2 - 6x + \frac{11}{4}.$$

- ▶ The first derivative is

$$f'(x) = x^3 - 6x^2 + 11x - 6,$$

which can be factored as

$$\begin{aligned} &= (x^2 - 3x + 2)(x - 3) \\ &= (x - 1)(x - 2)(x - 3). \end{aligned}$$

- ▶ Since the critical values are obtained by setting this first derivative function equal to zero and solving, it is apparent that they are simply 1, 2, and 3.
- ▶ Substituting these three x values into the original function shows that at the three points $(1, 0.5)$, $(2, 0.75)$, $(3, 0.5)$ there is a function maximum or minimum.

Multiple Modes

- ▶ To determine whether each of these is a maximum or a minimum, we must first obtain the second derivative form,

$$f''(x) = 3x^2 - 12x + 11,$$

and then plug in the three critical values:

$$f''(1) = 3(1)^2 - 12(1) + 11 = 2$$

$$f''(2) = 3(2)^2 - 12(2) + 11 = -1$$

$$f''(3) = 3(3)^2 - 12(3) + 11 = 2.$$

- ▶ This means that $(1, 0.5)$ and $(3, 0.5)$ are minima and $(2, 0.75)$ is a maximum.

Multiple Modes

- ▶ We can think of this in terms of Rolle's Theorem:
 - ▷ If we modified the function slightly by subtracting $\frac{1}{2}$ (i.e., add $\frac{9}{4}$ instead of $\frac{11}{4}$), then two minima would occur on the x -axis.
 - ▷ This change does not alter the fundamental form of the function or its maxima and minima; it just shifts the function up or down on the y -axis.
 - ▷ By Rolle's Theorem there is guaranteed to be another point in between where the first derivative is equal to zero; the point at $x = 2$ here is the only one from the factorization.
 - ▷ Since the points at $x = 1$ and $x = 3$ are minima, then the function increases away from them, which means that $x = 2$ has to be a maximum.

Root Finding with Newton-Raphson

- ▶ A *root* of a function is the point where the function crosses the x -axis: $f(x) = 0$.
- ▶ This value is a “root” of the function $f()$ in that it provides a solution to the polynomial expressed by the function.
- ▶ It is also the point where the function crosses the x -axis in a graph of x versus $f(x)$.
- ▶ Roots are typically substantively important points along the function, and it is therefore useful to be able to find them.

Root Finding with Newton-Raphson

- ▶ Previously, we were able to easily factor such functions as $f(x) = x^2 - 1$ to find the roots.
- ▶ However, this is not always realistically the case, so it is important to have a more general procedure.
- ▶ One such procedure is *Newton's method* (also called *Newton-Raphson*).
- ▶ The general form of Newton's method also (to be derived) is a series of steps according to

$$x_1 \cong x_0 - \frac{f(x_0)}{f'(x_0)},$$

where we move from a starting point x_0 to x_1 , which is closer to the root, using characteristics of the function itself.

- ▶ Newton's method converges quadratically in time (number of steps) to a solution provided that the selected starting point is reasonably close to the solution, although the results can be very bad if this condition is not met.
- ▶ The key problem with distant starting points is that if $f''(x)$ changes sign between this starting point and the objective, then the algorithm may even move *away* from the root (diverge).

Root Finding with Newton-Raphson, Taylor Series Expansion

- ▶ Newton's method exploits the *Taylor series expansion*, which gives the relationship between the value of a mathematical function at the point x_0 and the function value at another point, x_1 , given (with continuous derivatives over the relevant support) as

$$f(x_1) = f(x_0) + (x_1 - x_0)f'(x_0) + \frac{1}{2!}(x_1 - x_0)^2 f''(x_0) + \frac{1}{3!}(x_1 - x_0)^3 f'''(x_0) + \dots,$$

where f' is the first derivative with respect to x , f'' is the second derivative with respect to x , and so on.

- ▶ Infinite precision between the values $f(x_1)$ and $f(x_0)$ is achieved with the infinite extending of the series into higher order derivatives and higher order polynomials:

$$f(x_1) = f(x_0) + (x_1 - x_0)f'(x_0) + \frac{1}{2!}(x_1 - x_0)^2 f''(x_0) + \dots + \dots \frac{1}{\infty!}(x_1 - x_0)^\infty f^{(\infty)}(x_0)$$

Root Finding with Newton-Raphson, Taylor Series Expansion

- ▶ We are actually interested in finding a root of the function, which we will suppose exists at the undetermined point x_1 .
- ▶ What we know so far is that for any point x_0 that we would pick, it is possible to relate $f(x_1)$ and $f(x_0)$ with the Taylor series expansion.
- ▶ This is simplified in two ways.
 - ▷ First, note that if x_1 is a root, then $f(x_1) = 0$, meaning that the left-hand side of the two above equations is really zero.
 - ▷ Second, while we cannot perfectly relate the $f(x_0)$ to the function evaluated at the root because the expansion can never be *fully* evaluated, it should be obvious that using some of the first terms at least gets us closer to the desired point.

Root Finding with Newton-Raphson, Taylor Series Expansion

- ▶ Since the factorial function is increasing rapidly, let us use these last two facts and assert that

$$0 \cong f(x_0) + (x_1 - x_0)f'(x_0),$$

where the quality of the approximation can presumably be improved with better guesses of x_0 .

- ▶ Rearrange the equation so that the quantity of interest is on the left-hand side:

$$x_1 \cong x_0 - \frac{f(x_0)}{f'(x_0)},$$

to make this useful for candidate values of x_0 .

- ▶ If this becomes “algorithmic” because x_0 is chosen arbitrarily, then repeating the steps with successive approximations gives a process defined for the $(j + 1)$ th step:

$$x_{j+1} \cong x_j - \frac{f(x_j)}{f'(x_j)},$$

so that progressively improved estimates are produced until $f(x_{j+1})$ is sufficiently close to zero.

Root Finding with Newton-Raphson, Example

- ▶ As a simple example of Newton's method, suppose that we wanted a numerical routine for finding the square root of a number, μ .
- ▶ This is equivalent to finding the root of the simple equation $f(x) = x^2 - \mu = 0$.
- ▶ The first derivative is just $\frac{\partial}{\partial x}f(x) = 2x$.
- ▶ If we insert these functions into the $(j + 1)$ th step:

$$x_{j+1} = x_j - \frac{f(x_j)}{f'(x_j)},$$

we get

$$x_{j+1} = x_j - \frac{(x_j)^2 - \mu}{2x_j} = \frac{1}{2}(x_j + \mu x_j^{-1}).$$

- ▶ Recall that μ is a constant here defined by the problem and x_j is an arbitrary value at the j th step.

Root Finding with Newton-Raphson, Example

- ▶ A very basic algorithm for implementing this in software or on a hand calculator is

```
delta = 0.0000001
x = starting.value
D0:
    x.new = 0.5*(x + mu/x)
    x = x.new
UNTIL:
    abs(x^2 - mu) < delta
```

- ▶ Here **abs()** is the absolute value function and **delta** is the accuracy threshold that we are willing to accept.

Root Finding with Newton-Raphson, Example

- If we are interested in getting the square root of 99 and we run this algorithm starting at the obviously wrong point of $x = 2$, we get:

Iteration	x	x^2
0	2.0	4.0
1	25.75	663.0625
2	14.7973300970874	218.960978002168
3	10.7438631130069	115.430594591030
4	9.97921289277379	99.5846899593026
5	9.94991749815775	99.0008582201457
6	9.94987437115967	99.00000000186

Multidimensional Integrals

- ▶ In multiple dimensions, integrals measure volume and hyper-volume.
- ▶ For two variables, the *iterated integral* (also called the *repeated integral*), is given in definite form here by

$$V = \int_a^b \int_c^d f(x, y) dy dx,$$

where x is integrated between constants a and b , and y is integrated between constants c and d .

- ▶ The best way to think of this is by its inherent “nesting” of operations:

$$V = \int_a^b \left[\int_c^d f(x, y) dy \right] dx,$$

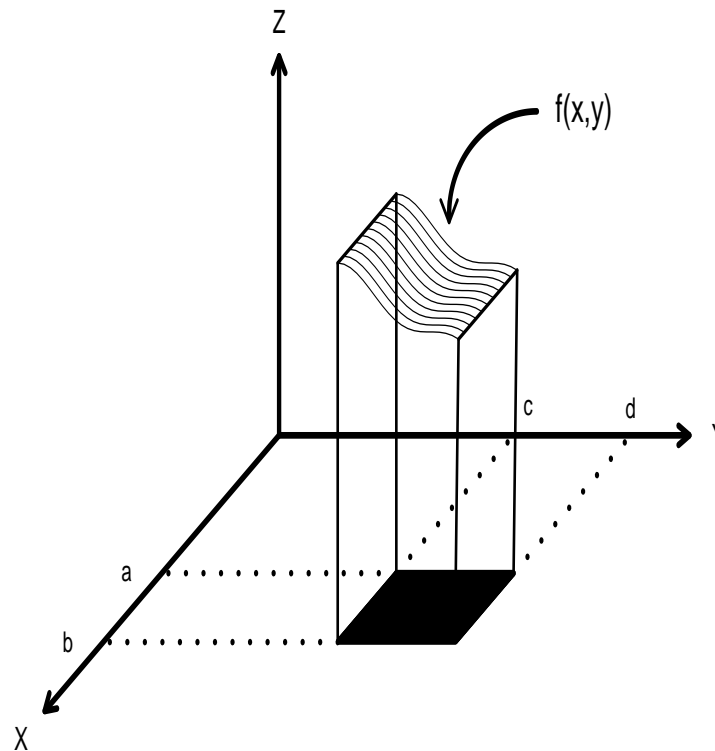
so that after the integration with respect to y there is a single integral left with respect to x of some *new* function that results from the first integration:

$$V = \int_a^b g(x) dx,$$

where $g(x) = \int_c^d f(x, y) dy$.

Multidimensional Integrals

- ▶ In the first (inner) step x is treated as a constant, and once the integration with respect to y is done, y is treated as a constant in the second (outer) step.
- ▶ In the figure, the black rectangle shows the region and the striped region shows $f(x, y)$.
- ▶ The integration provides the volume between this plan and this surface.



Example: Double Integral with Constant Limits

- In the following integral we first perform the operation for the variable y :

$$\begin{aligned}\int_2^3 \int_0^1 x^2 y^3 dy dx &= \int_2^3 \left[\frac{1}{4} x^2 y^4 \Big|_{y=0}^{y=1} \right] dx \\ &= \int_2^3 \left[\frac{1}{4} x^2 (1)^4 - \frac{1}{4} x^2 (0)^4 \right] dx\end{aligned}$$

- Then integrate x :

$$\int_2^3 \int_0^1 x^2 y^3 dy dx = \int_2^3 \frac{1}{4} x^2 dx. \qquad = \frac{1}{12} x^3 \Big|_{x=2}^{x=3} = \frac{19}{12}.$$

Example: Double Integral with Constant Limits

► It turns out that it is not important which variable we integrate first.

► Therefore, in our case,

$$V = \int_a^b \int_c^d f(x, y) dy dx = \int_a^b \left[\int_c^d f(x, y) dy \right] dx = \int_c^d \left[\int_a^b f(x, y) dx \right] dy.$$

► Sometimes though it pays to be strategic as to which variable integration to perform first, because the problem can be made more or less complicated by this decision.

► **Revisiting Double Integrals with Constant Limits.**

Now integrate x first with the same function:

$$\int_0^1 \int_2^3 x^2 y^3 dx dy = \int_0^1 \left[\frac{1}{3} x^3 y^3 \Big|_{x=2}^{x=3} \right] dy = \int_0^1 \left[\frac{1}{3} (3)^3 y^3 - \frac{1}{3} (2)^3 y^3 \right] dy = \int_0^1 \frac{19}{3} y^3 dy = \frac{19}{12} y^4 \Big|_{y=0}^{y=1} = \frac{19}{12}.$$

Limits as Functions

- ▶ So far all the limits of the integration have been constant.
- ▶ This is not realistic because it assumes that we are integrating functions over rectangles only to measure the volume under some surface described by $f(x, y)$.
- ▶ More generally, the region of integration is a function of x and y .
- ▶ Consider the unit circle centered at the origin defined by the equation $1 = x^2 + y^2$.
- ▶ The limits of the integral in the x and the y dimensions now depend on the other variable, and an iterated integral over this region measures the cylindrical volume under a surface defined by $f(x, y)$.
- ▶ Thus, we need a means of performing the integration in two steps as before but accounting for this dependency.

Limits as Functions

- ▶ To generalize the iterated integration process above, we first express the limits of the integral in terms of a single variable.

- ▶ For the circle example we could use either of

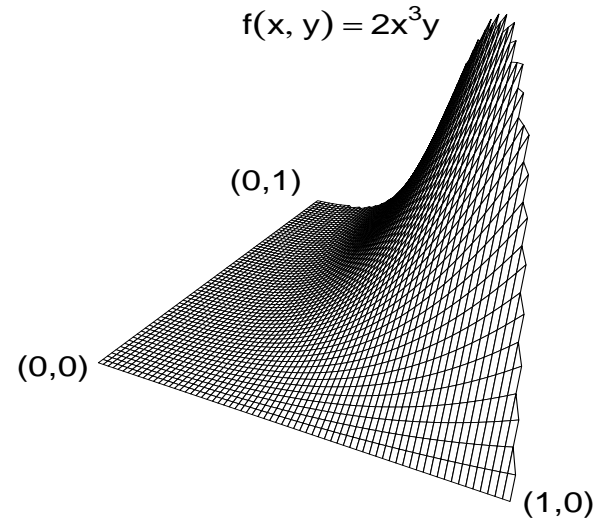
$$y = g_y(x) = \sqrt{1 - x^2} \qquad x = g_x(y) = \sqrt{1 - y^2},$$

depending on our order preference.

- ▶ If we pick the first form, then the integral limits in the inner operation are the expression for y in terms of x , so we label this as the function $g_y(x)$.

Integrating $f(x, y) = 2x^3y$ Over the First Quadrant of the Unit Circle

$$\begin{aligned} V &= \int_0^1 \int_0^{\sqrt{1-x^2}} 2x^3y \, dy \, dx \\ &= \int_0^1 x^3 y^2 \Big|_{y=0}^{y=\sqrt{1-x^2}} dx \\ &= \int_0^1 x^3(1-x^2) dx = \int_0^1 x^3 - x^5 dx \\ &= \frac{1}{4}x^4 - \frac{1}{6}x^6 \Big|_{x=0}^{x=1} = \frac{1}{12}. \end{aligned}$$



Iterated Integral Theorem

:

- ▶ A two-dimensional area of interest, denoted \mathfrak{A} , is characterized by either

$$a \leq x \leq b, \quad g_{y1}(x) \leq y \leq g_{y2}(x)$$

or

$$c \leq y \leq d, \quad g_{x1}(y) \leq x \leq g_{x2}(y).$$

- ▶ The function to be integrated, $f(x, y)$, is continuous for all of \mathfrak{A} .
- ▶ Then the double integral over \mathfrak{A} is equivalent to either of the iterated integrals:

$$\iint_{\mathfrak{A}} f(x, y) d\mathfrak{A} = \int_a^b \int_{g_{y1}(x)}^{g_{y2}(x)} f(x, y) dy dx = \int_c^d \int_{g_{x1}(y)}^{g_{x2}(y)} f(x, y) dx dy.$$

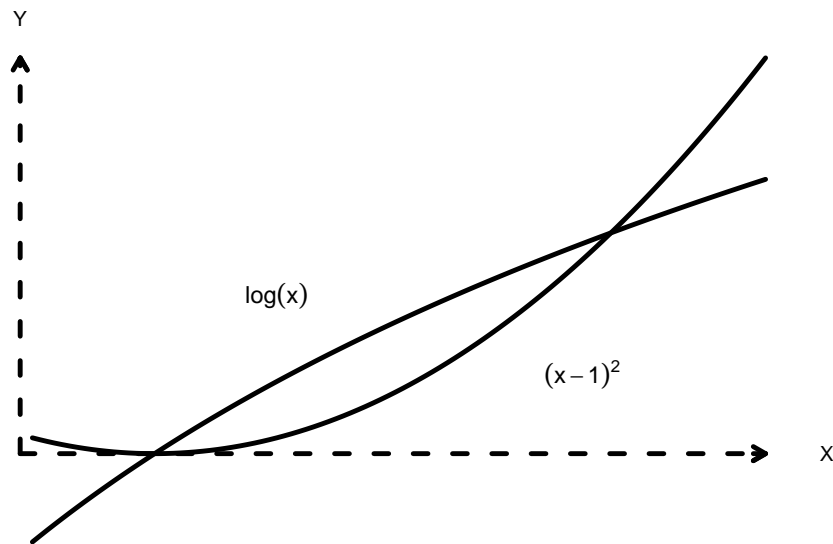
- ▶ This theorem states that, like the case with constant limits, we can switch the order of integration if we like, and that in both cases the result is equivalent to the motivating double integral.

Irregular Region Double Integral

- Consider the problem of integrating the function $f(x, y) = 3 + x^2$ over the region \mathfrak{A} , determined by the intersection of the function $f_1(x) = (x - 1)^2$ and the function $f_2(x) = \log(x)$.

$$V = \int_a^b \int_{(x-1)^2}^{\log(x)} 3 + x^2 dy dx,$$

where a is obviously from the point $(1, 0)$, but it is not clear how to determine b .

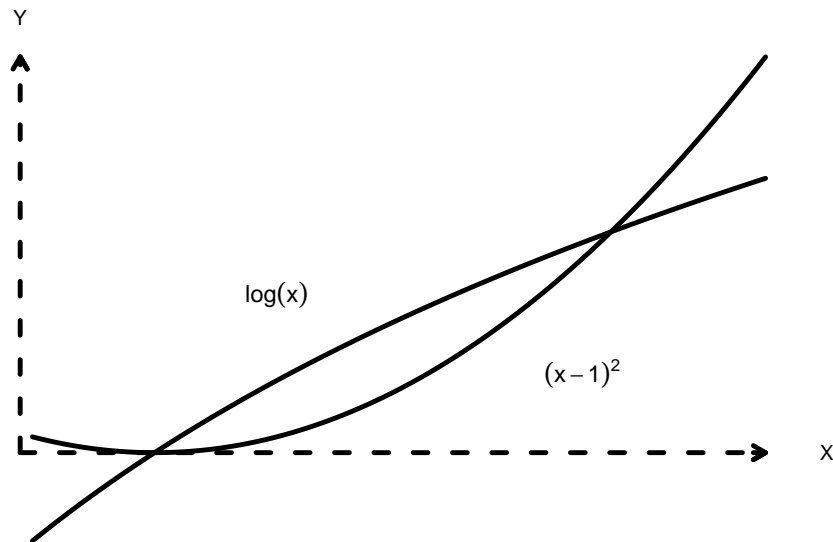


Irregular Region Double Integral

- One way to obtain this point is to define the difference equation and its first derivative:

$$f_d(x) = x^2 - 2x + 1 - \log(x), \quad f'_d(x) = 2x - 2 - \frac{1}{x},$$

which comes from expanding the square and subtracting $\log(x)$.



Irregular Region Double Integral

- ▶ Now we can apply the Newton-Raphson algorithm quite easily, and the result is the point (1.74688, 0.55783).
- ▶ The integration procedure becomes

$$\begin{aligned}
 \int_1^{1.74\dots} \int_{(x-1)^2}^{\log(x)} 3 + x^2 dy dx &= \int_1^{1.74\dots} 3y + x^2 y \Big|_{y=(x-1)^2}^{y=\log(x)} dx \\
 &= \int_1^{1.74\dots} (3 \log(x) - 3(x-1)^2 + x^2 \log(x) - x^2(x-1)^2) dx \\
 &= \int_1^{1.74\dots} (x^2 \log(x) + 3 \log(x) - x^4 + 2x^3 - 4x^2 + 6x - 3) dx \\
 &= \frac{1}{9} x^3 (3 \log(x) - 1) + 3(x \log(x) - x) - \frac{1}{5} x^5 + \frac{1}{2} x^4 - \frac{4}{3} x^3 + 3x^2 - 3x \Big|_1^{1.7468817} \\
 &= 0.02463552 - (-0.1444444) = 0.1690799,
 \end{aligned}$$

where the antiderivative of $x^2 \log(x)$ comes from the rule

$$\int u^n \log(u) du = \frac{u^{n+1}}{(n+1)^2} [(n+1) \log(u) - 1] + k$$

Finite and Infinite Series

- ▶ We use the idea of a mathematical finite or infinite series to model social phenomena.
- ▶ The first distinction that we care about is whether a series converges or diverges.
- ▶ We will also be centrally concerned here with the idea of a limit.

Measuring Small Group Standing

- ▶ Some sociologists care about the relative standing of individuals in small, bounded groups.
- ▶ This can be thought of as popularity, standing, or esteem, broadly defined.
- ▶ Suppose there are N members of the group and \mathbf{A} is the $N \times N$ matrix where a 1 for a_{ij} indicates that individual i (the row) chooses to associate with individual j (the column), and 0 indicates the opposite choice.
- ▶ For convenience, the diagonal values are left to be zero (one cannot choose or not choose in this sense).
- ▶ The early literature (Moreno 1934) posited a ranking of status that simply added up choice reception by individual, which can be done by multiplying the \mathbf{A} matrix by an appropriate unit vector to sum by columns, for instance,

$$\text{status}_1 = \mathbf{A}'u = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}' \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 2 \\ 0 \end{bmatrix}.$$

Measuring Small Group Standing

- ▶ Later it was proposed (Katz 1953) that indirect associations also led to increased standing, so a means of incorporating indirect paths throughout this matrix needed to be included. Of
- ▶ We do not want to require that indirect paths be equal in weight to direct paths, so we include a weighting factor, $0 < \alpha < 1$, that discounts distance.
- ▶ If we include every possible path, then the matrix that keeps track of these is given by the *finite matrix series*:

$$B = \alpha \mathbf{A} + \alpha^2 \mathbf{A}' \mathbf{A} + \alpha^3 \mathbf{A} \mathbf{A}' \mathbf{A} + \alpha^4 \mathbf{A}' \mathbf{A} \mathbf{A}' \mathbf{A} + \dots + \alpha^N \mathbf{A}' \mathbf{A} \cdots \mathbf{A}' \mathbf{A}$$

(assuming N is even).

Measuring Small Group Standing

- So that the Katz measure of standing for the example above is now $status_2 = B'u =$

$$\begin{aligned}
 &= \left(0.5 \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} + 0.5^2 \begin{bmatrix} 3 & 2 & 1 & 0 \\ 2 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + 0.5^3 \begin{bmatrix} 1 & 1 & 2 & 0 \\ 3 & 2 & 1 & 0 \\ 5 & 4 & 2 & 0 \\ 6 & 5 & 4 & 0 \end{bmatrix} + 0.5^4 \begin{bmatrix} 14 & 11 & 7 & 0 \\ 11 & 9 & 6 & 0 \\ 7 & 6 & 6 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right)' \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\
 &= \begin{bmatrix} 1.7500 & 1.3125 & 1.4375 & 0.0000 \\ 2.0625 & 1.3125 & 0.7500 & 0.0000 \\ 1.8125 & 1.6250 & 1.1250 & 0.0000 \\ 1.2500 & 1.1250 & 1.0000 & 0.0000 \end{bmatrix}' \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6.8750 \\ 5.3750 \\ 4.3125 \\ 0.0000 \end{bmatrix}.
 \end{aligned}$$

These calculations obviously used $\alpha = 0.5$.

- The final column vector shows relative standing that includes these other paths.
- This new measure apparently improves the standing of the second person.

Notation for Series

- ▶ A series is the consecutively produced values are generated by some sort of a rule or relationship.
- ▶ Notationally, start with an *infinite series*:

$$S_{\infty} = \sum_{i=1}^{\infty} x_i = x_1 + x_2 + x_3 + \cdots ,$$

which is just a set of enumerated values stretching out to infinity.

- ▶ This is not to say that any one of these values is itself equal to infinity or even that their sum is necessarily infinite, but rather that the *quantity* of them is infinite.
- ▶ We can also define a *finite series* of length n :

$$S_n = \sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \cdots + x_n,$$

which is a series that terminates with the n th value: x_n .

- ▶ This may also simply be the first n values of an infinite series and in this context is called an **nth partial sum** of the larger infinite sequence.

Convergence

- ▶ A series is *convergent* if the limit as n goes to infinity is bounded (noninfinite itself):

$$\lim_{n \rightarrow \infty} S_n = \mathfrak{A}, \text{ where } \mathfrak{A} \text{ is bounded.}$$

- ▶ A series is *divergent* if it is not convergent, that is, if \mathfrak{A} above is positive or negative infinity.
- ▶ Another test is stated in the following theorem.

Integral Test for Series Convergence:

- ▷ If S_n converges, then $\lim_{n \rightarrow \infty} x_n = 0$, and if $\lim_{n \rightarrow \infty} x_n \neq 0$, then S_n diverges.
- ▷ If S_n is a series with all positive terms and $f(x)$ is the corresponding function that must be decreasing and everywhere continuous for values of $x \geq 1$, then the series S_n and the integral $\int_1^{\infty} f(x)dx$ both converge or diverge.

Convergence

- ▶ It is *not* true that a zero limit of x_n shows convergence (the logic only goes in one direction).
- ▶ For instance, a harmonic series has the property that x_n goes to zero in the limit, but it is a well-known divergent series:

$$S_\infty = \sum_{i=1}^{\infty} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots .$$

- ▶ Convergence is a handy result because it means that the infinite series can be approximated by a reasonable length finite series.
- ▶ So additional values become unimportant at some point).

Convergence

- ▶ Evaluate the limiting term of the series:

$$\sum_{i=1}^n \frac{i-1}{i+1} = 0 + \frac{1}{3} + \frac{2}{4} + \frac{3}{5} + \dots .$$

We could note that the values get larger, which is clearly an indication of a diverging series.

- ▶ The integral test also shows this, because

$$\lim_{n \rightarrow \infty} \frac{n-1}{n+1} = \lim_{n \rightarrow \infty} \frac{1 - \frac{1}{n}}{1 + \frac{1}{n}} = 1,$$

which is not zero, indicating divergence of the series.

Convergence

- ▶ The integral part of the statement above relates the characteristic of a series with an integral, so that if we can obtain convergence of one, we can establish convergence of the other.
- ▶ Consider the simple series and associated integral

$$S_{\infty} = \sum_{i=1}^{\infty} \frac{1}{i^3}, \quad I_{\infty} = \int_1^{\infty} \frac{1}{x^3} dx.$$

The integral quantity is $\frac{1}{2}$, so we know that the series converges.

Examples of Series

► **Consecutive Integers.** Given $x_k = x_{k-1} + 1$, then $S_\infty = \sum_{i=1}^{\infty} i = 1 + 2 + 3 + \dots$. This is a divergent series.

► **Telescoping Sum.** Assume $S_n = \sum_{i=1}^n \left(\frac{1}{i} - \frac{1}{i+1}\right)$. On quick inspection, there seems to be no obvious reason why this sum should converge in the limit, but note that it can be reexpressed as

$$\begin{aligned} S_n &= \sum_{i=1}^n \left(\frac{1}{i} - \frac{1}{i+1}\right) = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \dots + \left(\frac{1}{i} - \frac{1}{i+1}\right) \\ &= 1 + \left(-\frac{1}{2} + \frac{1}{2}\right) + \left(-\frac{1}{3} + \frac{1}{3}\right) + \dots + \left(-\frac{1}{i} + \frac{1}{i}\right) + \frac{1}{i+1} \\ &= 1 + \frac{1}{i+1}. \end{aligned}$$

► Take n to infinity:

$$S_\infty = \lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n+1}\right) = 1$$

since no matter how big n becomes, adjacent terms cancel except for the first and last.

► Notice also that the limit of the n th term is zero, as per the integral test.

Examples of Series

► **Geometric Series.** Suppose for some positive integer k and real number r , we have the series

$$S_n = \sum_{i=0}^n kr^i = k + kr + kr^2 + kr^3 + \cdots + kr^n.$$

Specific cases include

$$\begin{array}{ll} k = 1, r = -1 : & 1 - 1 + 1 - 1 \cdots \\ k = 1, r = 2 : & 1 + 2 + 4 + 8 + 16 + \cdots \\ k = 2, r = \frac{1}{2} : & 2 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots \end{array}$$

It turns out that the geometric series converges to $S_n = \frac{k}{1-r}$ if $|r| < 1$ but diverges if $|r| > 1$. The series also diverges for $r = 1$ since it is then simply the sum $k + k + k + k + \cdots$.

► **Repeating Values as a Geometric Series.** Consider the repeating number

$$0.123123123 \dots = \frac{123}{1000^1} + \frac{123}{1000^2} + \frac{123}{1000^3} + \cdots$$

which is expressed in the second form as a geometric series with $k = 123$ and $r = 0.001$. Clearly this sequence converges because r is (much) less than one.

More Tests

► For the infinite series $\sum_{i=1}^{\infty} a_i$:

▷ **Ratio Test.** If every $a_i > 0$ and $\lim_{i \rightarrow \infty} \frac{a_{i+1}}{a_i} = A$, then the series converges for $A < 1$, diverges for $A > 1$, and may converge or diverge for $A = 1$.

▷ **Root Test.** If every $a_i > 0$ and $\lim_{i \rightarrow \infty} (a_i)^{\frac{1}{i}} = A$, then the series converges for $A < 1$, diverges for $A > 1$, and may converge or diverge for $A = 1$.

▷ **Comparison Test.** If there is a convergent series $\sum_{i=1}^{\infty} b_i$ and a positive (finite) integer value J such that $a_i \leq b_i \forall i \geq J$, then $\sum_{i=1}^{\infty} a_i$ converges.

Properties of Convergent Series

- Limiting Values $\lim_{n \rightarrow \infty} a_n = 0$
(if $\lim_{n \rightarrow \infty} a_n \neq 0$, then $\sum_{i=1}^{\infty} a_i$ diverges)
- Summation $\sum_{i=1}^{\infty} a_i + \sum_{i=1}^{\infty} b_i = \sum_{i=1}^{\infty} (a_i + b_i)$
- Scalar Multiplication $\sum_{i=1}^{\infty} k a_i = k \sum_{i=1}^{\infty} a_i$

An Equilibrium Point in Simple Games

- ▶ Consider the basic prisoner's dilemma game, which has many variants, but here two parties obtain \$10 each for both cooperating, \$15 dollars for acting opportunistically when the other acts cooperatively, and only \$5 each for both acting opportunistically.
- ▶ What is the value of this game to a player who intends to act opportunistically at all iterations and expects the other player to do so as well?
- ▶ Assume that each player discounts the future value of payoffs by 0.9 per period.

- ▶ Then this player expects a minimum payout of

$$\$5(0.9^0 + 0.9^1 + 0.9^2 + 0.9^3 + \dots + 0.9^\infty).$$

- ▶ The component in parentheses is a geometric series where $r = 0.9 < 1$, so it converges giving $\$5 \frac{1}{1-0.9} = \50 .
- ▶ The game might be worth slightly more to our player if the opponent was unaware of this strategy on the first or second iteration (presumably it would be quite clear after that).

Other Types of Infinite Series

- ▶ Occasionally there are special characteristics of a given series that allow us to assert convergence.
- ▶ A series where adjacent terms are alternating in sign for the whole series is called an *alternating series*.
- ▶ An alternating series converges if the same series with absolute value terms also converges. So if $\sum_{i=1}^{\infty} a_i$ is an alternating series, then it converges if $\sum_{i=1}^{\infty} |a_i|$ converges.
- ▶ For instance, the alternating series given by

$$\sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i^2}$$

converges if $\sum_{i=1}^{\infty} \frac{1}{i^2}$ converges since the latter is always greater for some given i value.

- ▶ This series converges if the integral is finite

$$\int_1^{\infty} \frac{1}{x^2} dx = -x^{-1} \Big|_1^{\infty} = -\frac{1}{\infty} - \left(-\frac{1}{1}\right) = 1,$$

so the second series converges and thus the original series converges.

Power Series

- ▶ The *power series*, which is a series defined for x of the form

$$\sum_{i=1}^{\infty} a_i x^i,$$

for scalar values $a_1, a_2, \dots, a_{\infty}$.

- ▶ A special case is defined for the difference operator $(x - x_0)$:

$$\sum_{i=1}^{\infty} a_i (x - x_0)^i.$$

This type of power series has the characteristic that if it converges for the given value of $x_0 \neq 0$, then it converges for $|x| < |x_0|$.

- ▶ Conversely, if the power series diverges at x_0 , then it also diverges for $|x| > |x_0|$.

Power Series

- ▶ There are three power series that converge in important ways:

$$\sum_{i=1}^{\infty} \frac{x^i}{i!} = e^x$$

$$\sum_{i=1}^{\infty} \frac{(-1)^i x^{2i+1}}{(2i+1)!} = \sin(x)$$

$$\sum_{i=1}^{\infty} \frac{(-1)^i x^{2i}}{(2i)!} = \cos(x).$$

- ▶ The idea here is bigger than just these special cases.
- ▶ If a function can be expressed in the form $f(x) = \sum_{i=1}^{\infty} a_i (x - x_0)^i$, then it has derivatives of all orders and a_i can be expressed as the i th derivative divided by the i th factorial.

Power Series

- ▶ The converse is not necessarily true in terms of guaranteeing the existence of a series expansion.
- ▶ The function can be expressed as

$$f(x) = \frac{1}{0!}(x - x_0)^0 f(x_0) + \frac{1}{1!}(x - x_0)^1 f'(x_0) + \frac{1}{2!}(x - x_0)^2 f''(x_0) \\ + \frac{1}{3!}(x - x_0)^3 f'''(x_0) + \dots ,$$

which is just the *Taylor series* discussed previously.

Maclaurin Series

- ▶ A special case of the Taylor series is the *Maclaurin series*, which is given when $x_0 = 0$.
- ▶ Many well-known functions can be rewritten as a Maclaurin series.
- ▶ For instance, now express $f(x) = \log(x)$ as a Maclaurin series and compare at $x = 2$ to $x = 1$ where $f(x) = 0$.
- ▶ We first note that

$$\begin{array}{ll}
 f'(x) = \frac{1}{x} & f''(x) = \frac{-1}{x^2} \\
 f'''(x) = \frac{2}{x^3} & f^{(4)}(x) = \frac{-6}{x^4} \\
 \vdots &
 \end{array}$$

- ▶ This leads to the general order form for the derivative

$$f^{(i)}(x) = \frac{(-1)^{i+1}(i-1)!}{x^i}.$$

Maclaurin Series

- So the function of interest can be expressed as follows by plugging in the derivative term and simplifying:

$$\begin{aligned}
 \log(x) &= \sum_{i=1}^{\infty} \frac{1}{i!} (x - x_0)^i f^{(i)}(x_0) \\
 &= \sum_{i=1}^{\infty} \frac{1}{i!} (x - x_0)^i \left(\frac{(-1)^{i+1} (i-1)!}{x_0^i} \right) \\
 &= \sum_{i=1}^{\infty} (-1)^{i+1} \frac{1}{i} \frac{(x - x_0)^i}{x_0^i}.
 \end{aligned}$$

- Now set $x_0 = 1$ and $x = 2$:

$$\log(2) = \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \cdots,$$

which converges to the (correct) value of 0.6931472.

Vector Function Notation

- ▶ Using standard (Hamiltonian) notation, we start with two orthogonal unit vectors \mathbf{i} and \mathbf{j} starting at the origin and following along the x -axis and y -axis correspondingly.
- ▶ Any vector in two-space (\mathfrak{R}^2) can be expressed as a scalar-weighted sum of these two *basis vectors* giving the horizontal and vertical progress:

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j}.$$

- ▶ For example, to characterize the vector from the point $(3, 1)$ to the point $(5, 5)$ we use $\mathbf{v} = (5 - 3)\mathbf{i} + (5 - 1)\mathbf{j} = 2\mathbf{i} + 4\mathbf{j}$.
- ▶ Now instead of the scalars a and b , substitute the real-valued functions $f_1(t)$ and $f_2(t)$ for $t \in \mathfrak{R}$.

Vector Function Notation

- ▶ We can define the *vector function*:

$$\mathbf{f}(t) = f_1(t)\mathbf{i} + f_2(t)\mathbf{j},$$

which gives the x and y vector values $\mathbf{f}(t) = (x, y)$.

- ▶ The parametric representation of the line passing through $(3, 1)$ and $(5, 5)$ is found according to

$$x = x_0 + t(x_1 - x_0) \qquad y = y_0 + t(y_1 - y_0)$$

$$x = 3 + 2t \qquad y = 1 + 4t,$$

meaning that for some value of t we have a point on the line.

- ▶ To get the expression for this line in standard slope-intercept form, we first find the slope by getting the ratio of differences $(5 - 1)/(5 - 3) = 2$ in the standard fashion and subtracting from one of the two points to get the y value where x is zero: $(0, -5)$.
- ▶ Setting $x = t$, we get $y = -5 + 2x$.

Nonlinear Forms

- ▶ Now suppose that we have some curvilinear form in \mathfrak{R} given the functions $f_1(t)$ and $f_2(t)$, and we would like to get the slope of the tangent line at the point $t_0 = (x_0, y_0)$.
- ▶ This is found by evaluating the ratio of first derivatives of the functions

$$\mathbf{R}'(t_0) = \frac{f_2'(t_0)}{f_1'(t_0)},$$

where we have to worry about the restriction that $f_1'(t_0) \neq 0$ for obvious reasons.

- ▶ Consider what we are doing here; the derivatives are producing incremental changes in x and y separately by the construction with \mathbf{i} and \mathbf{j} above.
- ▶ Because of the limits, this ratio is the instantaneous change in y for a change in x , that is, the slope.
- ▶ Consider this logic in the notation:

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta x} = \frac{\frac{\Delta y}{\lim_{\Delta t \rightarrow 0} \Delta t}}{\frac{\Delta x}{\lim_{\Delta t \rightarrow 0} \Delta t}} = \frac{\frac{\partial y}{\partial t}}{\frac{\partial x}{\partial t}} = \frac{\partial y}{\partial t} \frac{\partial t}{\partial x} = \frac{\partial y}{\partial x}.$$

Nonlinear Forms

- ▶ For example, we can find the slope of the tangent line to the curve $x = 3t^3 + 5t^2 + 7$, $y = t^2 - 2$, at $t = 1$:

$$f_1'(1) = 9t^2 + 10t \Big|_{t=1} = 19$$

$$f_2'(1) = 2t \Big|_{t=1} = 2$$

$$\mathbf{R}'(1) = \frac{2}{19}.$$

- ▶ We can also find all of the horizontal and vertical tangent lines to this curve by a similar calculation.
- ▶ There are horizontal tangent lines when $f_1'(t) = 9t^2 + 10t = 0$.
- ▶ Factoring this shows that there are horizontal tangents when $t = 0$, $t = -\frac{10}{9}$.
- ▶ Plugging these values back into $x = 3t^3 + 5t^2 + 7$ gives horizontal tangents at $x = 7$ and $x = 9.058$.
- ▶ There are vertical horizontal lines when $f_2'(t) = 2t = 0$, which occurs only at $t = 0$, meaning $y = -2$.

Differentiation of a Vector Function

- ▶ The vector function $\mathbf{f}(t)$ is differentiable with domain t if the limit

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbf{f}(t + \Delta t) - \mathbf{f}(t)}{\Delta t}$$

exists and is bounded (finite) for all specified t .

- ▶ This is the same idea we saw for scalar differentiation, except that by consequence

$$\mathbf{f}'(t) = f'_1(t)\mathbf{i} + f'_2(t)\mathbf{j},$$

which means that the function can be differentiated by these orthogonal pieces.

- ▶ It follows also that if $\mathbf{f}'(t)$ meets the criteria above, then $\mathbf{f}''(t)$ exists, and so on.

- ▶ As a demonstration, let $\mathbf{f}(t) = e^{5t}\mathbf{i} + \sin(t)\mathbf{j}$, so that

$$\mathbf{f}'(t) = 5e^{5t}\mathbf{i} + \cos(t)\mathbf{j}.$$

Integration of a Vector Function

- ▶ Integration proceeds piecewise for the vector function just as differentiation was done.
- ▶ For $\mathbf{f}(t) = f_1(t)\mathbf{i} + f_2(t)\mathbf{j}$, the integral is

$$\begin{cases} \int \mathbf{f}(t)dt = [\int f_1(t)dt] \mathbf{i} + [\int f_2(t)dt] \mathbf{j} + \mathbf{K} & \text{for the indefinite form,} \\ \int_a^b \mathbf{f}(t)dt = \left[\int_a^b f_1(t)dt \right] \mathbf{i} + \left[\int_a^b f_2(t)dt \right] \mathbf{j} & \text{for the definite form.} \end{cases}$$

- ▶ We previously saw an arbitrary constant k for indefinite integrals of scalar functions, but that is replaced here with the more appropriate vector-valued form \mathbf{K} .
- ▶ This “splitting” of the integration process between the two dimensions can be tremendously helpful in simplifying difficult dimensional problems.

Integration of a Vector Function

- Consider the trigonometric function $\mathbf{f}(t) = \tan(t)\mathbf{i} + \sec^2(t)\mathbf{j}$. The integral over $[0 : \pi/4]$ is produced by

$$\begin{aligned}\int_0^{\pi/4} \mathbf{f}(t) dt &= \left[\int_0^{\pi/4} \tan(t) dt \right] \mathbf{i} + \left[\int_0^{\pi/4} \sec^2(t) dt \right] \mathbf{j} \\ &= \left[-\log(|\cos(t)|) \Big|_0^{\pi/4} \right] \mathbf{i} + \left[\tan(t) \Big|_0^{\pi/4} \right] \mathbf{j} \\ &= 0.3465736\mathbf{i} + 0.854509\mathbf{j}.\end{aligned}$$

- Indefinite integrals sometimes come with additional information that makes the problem more complete.

Integration of a Vector Function

- If $\mathbf{f}'(t) = t^2\mathbf{i} - t^4\mathbf{j}$, and we know that $\mathbf{f}(0) = 4\mathbf{i} - 2\mathbf{j}$, then a full integration starts with

$$\begin{aligned}\mathbf{f}(t) &= \int \mathbf{f}'(t) dt \\ &= \int t^2 dt \mathbf{i} - \int t^4 dt \mathbf{j} \\ &= \frac{1}{3}t^3 \mathbf{i} + \frac{1}{5}t^5 \mathbf{j} + \mathbf{K}.\end{aligned}$$

- Since $\mathbf{f}(0)$ is the function value when the components above are zero, except for \mathbf{K} we can substitute this for \mathbf{K} to complete

$$\mathbf{f}(t) = \left(\frac{1}{3}t^3 + 4 \right) \mathbf{i} + \left(\frac{1}{5}t^5 - 2 \right) \mathbf{j}.$$

Maximization

- ▶ In statistical work in the social sciences, a scalar-valued vector function is important for maximization and description.
- ▶ We will not go into the theoretical derivation of this process (maximum likelihood estimation) but instead will describe the key vector components.
- ▶ Start with a function: $y = f(\mathbf{x}) = f(x_1, x_2, x_3 \dots, x_k)$ operating on the k -length vector \mathbf{x} .
- ▶ The vector of partial derivatives with respect to each x_i is called the *gradient*:

$$\mathbf{g} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \partial y / \partial x_1 \\ \partial y / \partial x_2 \\ \partial y / \partial x_3 \\ \vdots \\ \partial y / \partial x_k \end{bmatrix},$$

which is given by convention as a column vector.

Maximization With A Hessian Matrix

- The second derivative for this setup is taken in a different manner than one might suspect; it is done by differentiating the complete gradient vector by each x_i such that the result is a $k \times k$ matrix:

$$\begin{aligned}
 \mathbf{H} &= \left[\frac{\partial}{\partial x_1} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right), \frac{\partial}{\partial x_2} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right), \right. \\
 &\quad \left. \frac{\partial}{\partial x_3} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right), \dots, \frac{\partial}{\partial x_k} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right) \right] \\
 &= \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_1 \partial \mathbf{x}_1} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_1 \partial \mathbf{x}_2} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_1 \partial \mathbf{x}_3} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_1 \partial \mathbf{x}_k} \\ \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_2 \partial \mathbf{x}_1} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_2 \partial \mathbf{x}_2} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_2 \partial \mathbf{x}_3} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_2 \partial \mathbf{x}_k} \\ \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_3 \partial \mathbf{x}_1} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_3 \partial \mathbf{x}_2} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_3 \partial \mathbf{x}_3} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_3 \partial \mathbf{x}_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_k \partial \mathbf{x}_1} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_k \partial \mathbf{x}_2} & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_k \partial \mathbf{x}_3} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_k \partial \mathbf{x}_k} \end{bmatrix} \\
 &= \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}}.
 \end{aligned}$$

Constrained Optimization

- ▶ Constrained optimization is a standard tool in some social science literatures, notably economics.
- ▶ The key feature of these methods is using the first derivative to find a point where the slope of the tangent line is zero.
- ▶ Usually this is substantively interesting in that it tells us where some x value leads to the greatest possible $f(x)$ value to maximize some quantity of interest: money, utility, productivity, cooperation, and so on.
- ▶ These problems are usually more useful in higher dimensions, for instance, what values of x_1 , x_2 , and x_3 simultaneously provide the greatest value of $f(x_1, x_2, x_3)$?

Constrained Optimization

- ▶ Now let us revisit the optimization problem but requiring the additional *constraint* that the values of x_1 , x_2 , and x_3 have to conform to some predetermined relationship.
- ▶ Usually these constraints are expressed as inequalities, say $x_1 > x_2 > x_3$, or with specific equations like $x_1 + x_2 + x_3 = 10$.
- ▶ The procedure we will use is now called *constrained optimization* because we will optimize the given function but with the constraints specified in advance.
- ▶ The constrained solution will never be a better solution than the unconstrained solution because we are requiring certain relationships among the terms.

Constrained Optimization

- ▶ Our task will be to maximize a k -dimensional function $f(\mathbf{x})$ subject to the arbitrary constraints expressed as m functions:

$$c_1(\mathbf{x}) = r_1, c_2(\mathbf{x}) = r_2, \dots, c_m(\mathbf{x}) = r_m,$$

where the r_1, r_2, \dots, r_m values are stipulated constants.

- ▶ The trick is to deliberately include these constraint functions in the maximization process.
- ▶ This method is called the *Lagrange multiplier*, and it means substituting for the standard function, $f(\mathbf{x})$, a modified version of the form

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= f(\mathbf{x}) + \lambda_1(c_1(\mathbf{x}) - r_1) + \lambda_2(c_2(\mathbf{x}) - r_2) + \dots + \lambda_m(c_m(\mathbf{x}) - r_m) \\ &= f(\mathbf{x}) + \underset{1 \times m}{\boldsymbol{\lambda}'} \underset{m \times k}{\mathbf{c}(\mathbf{x})}, \end{aligned}$$

where the second form is just a restatement in vector form in which the $-r$ terms are embedded ($\boldsymbol{\lambda}'$ denotes a transpose not a derivative).

- ▶ Commonly these r_1, r_2, \dots, r_m values are zero which makes the expression of $L(\mathbf{x}, \boldsymbol{\lambda})$ cleaner.
- ▶ The λ terms in this expression are called Lagrange multipliers.

Constrained Optimization

- ▶ Now we take two (multidimensional) partial derivatives and set them equal to zero just as before, except that we need to keep track of $\boldsymbol{\lambda}$ as well:

$$\frac{d}{d\mathbf{x}}L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{d}{d\mathbf{x}}f(\mathbf{x}) + \boldsymbol{\lambda}' \frac{d}{d\mathbf{x}}\mathbf{c}(\mathbf{x}) \equiv \mathbf{0} \quad \Rightarrow \quad \frac{d}{d\mathbf{x}}f(\mathbf{x}) = -\boldsymbol{\lambda}' \frac{d}{d\mathbf{x}}\mathbf{c}(\mathbf{x})$$

$$\frac{d}{d\boldsymbol{\lambda}}L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}(\mathbf{x}) \equiv \mathbf{0}.$$

- ▶ The derivative with respect to $\boldsymbol{\lambda}$ is simple because there are no λ values in the first term. The term $\frac{d}{d\mathbf{x}}\mathbf{c}(\mathbf{x})$ is just the matrix of partial derivatives of the constraints, and it is commonly abbreviated \mathbf{C} .
- ▶ Expressing the constraints in this way also means that the Lagrange multiplier component $-\boldsymbol{\lambda}' \frac{d}{d\mathbf{x}}\mathbf{c}(\mathbf{x})$ is now just $-\boldsymbol{\lambda}'\mathbf{C}$, which is easy to work with.
- ▶ It is interesting to contrast the final part of the first line $\frac{d}{d\mathbf{x}}f(\mathbf{x}) = -\boldsymbol{\lambda}' \frac{d}{d\mathbf{x}}\mathbf{c}(\mathbf{x})$ with unconstrained optimization at this step, $\frac{d}{d\mathbf{x}}f(\mathbf{x}) = \mathbf{0}$, because it clearly shows the imposition of the constraints on the function maximization.
- ▶ Finally, after taking these derivatives, we solve for the values of \mathbf{x} and λ that result.

Constrained Optimization, Numerical Example

- Suppose we have “data” according to the following matrix and vector:

$$\boldsymbol{\omega} = [1 \ 2 \ 3] \qquad \boldsymbol{\Omega} = \begin{bmatrix} 5 & 1 & 5 \\ 4 & 5 & 5 \\ 1 & 5 & 5 \end{bmatrix},$$

and we want to maximize the quadratic function:

$$f(\mathbf{x}) = \mathbf{x}'\boldsymbol{\Omega}\mathbf{x} - 2\boldsymbol{\omega}'\mathbf{x} + 5.$$

- This is great justification for using matrix algebra because the same function $f(\mathbf{x})$ written out in scalar notation is

$$f(\mathbf{x}) = 5x_1^2 + 5x_2^2 + 5x_3^2 + 5x_1x_2 + 6x_1x_3 + 10x_2x_3 + 2x_1 + 4x_2 + 6x_3 + 5.$$

- We now impose the two constraints:

$$2x_1 + x_2 = 0 \qquad x_1 - x_2 - x_3 = 0,$$

which gives

$$\mathbf{c}(\mathbf{x}) = \begin{bmatrix} c_1(\mathbf{x}) \\ c_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} 2x_1 & x_2 & 0 \\ x_1 & -x_2 & -x_3 \end{bmatrix}$$

Constrained Optimization, Numerical Example

- ▶ Taking the partial derivatives $\frac{d}{d\mathbf{x}}\mathbf{c}(\mathbf{x})$ (i.e., with respect to x_1 , x_2 , and x_3) produces the \mathbf{C} matrix:

$$\mathbf{C} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & -1 \end{bmatrix},$$

which is just the matrix-collected multipliers of the \mathbf{x} terms in the constraints since there were no higher order terms to worry about here.

- ▶ Specify the Lagrange multiplier version of the function:

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= f(\mathbf{x}) + \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x}) \\ &= \underbrace{\mathbf{x}' \boldsymbol{\Omega} \mathbf{x}}_{(1 \times 3)(3 \times 3)(3 \times 1)} - 2 \underbrace{\boldsymbol{\omega}' \mathbf{x}}_{(1 \times 3)(3 \times 1)} + 5 + \underbrace{\boldsymbol{\lambda}' \mathbf{C} \mathbf{x}}_{(1 \times 2)(2 \times 3)(3 \times 1)}. \end{aligned}$$

- ▶ Note that $m = 2$ and $k = 3$ in this example.

Constrained Optimization, Numerical Example

- ▶ The next task is to take the two derivatives and set them equal to zero:

$$\frac{d}{d\mathbf{x}}L(\mathbf{x}, \boldsymbol{\lambda}) = 2\mathbf{x}'\boldsymbol{\Omega} - 2\boldsymbol{\omega}' + \boldsymbol{\lambda}'\mathbf{C} \equiv \mathbf{0} \quad \Rightarrow \quad \mathbf{x}'\boldsymbol{\Omega} + \frac{1}{2}\boldsymbol{\lambda}'\mathbf{C} = \boldsymbol{\omega}'$$

$$\frac{d}{d\boldsymbol{\lambda}}L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}(\mathbf{x}) = \mathbf{C}\mathbf{x} \equiv \mathbf{0},$$

where $\mathbf{c}(\mathbf{x}) = \mathbf{C}\mathbf{x}$ comes from the simple form of the constraints here.

- ▶ These final expressions allow us to stack the equations (they are multidimensional anyway) into the following single matrix statement:

$$\begin{bmatrix} \boldsymbol{\Omega}' & \frac{1}{2}\mathbf{C}' \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\omega}' \\ \mathbf{0} \end{bmatrix}$$

Constrained Optimization, Numerical Example

- We used the transpose property such that $(\boldsymbol{\lambda}'\mathbf{C})' = \mathbf{C}'\boldsymbol{\lambda}$ and $(\mathbf{x}'\boldsymbol{\Omega})' = \boldsymbol{\Omega}'\mathbf{x}$ since we want the column vector $\boldsymbol{\omega}$:

$$\begin{aligned}\boldsymbol{\omega}' &= \mathbf{x}'\boldsymbol{\Omega} + \frac{1}{2}\boldsymbol{\lambda}'\mathbf{C} \\ \boldsymbol{\omega} &= (\mathbf{x}'\boldsymbol{\Omega} + \frac{1}{2}\boldsymbol{\lambda}'\mathbf{C})' = \boldsymbol{\Omega}'\mathbf{x} + \frac{1}{2}\mathbf{C}'\boldsymbol{\lambda} \\ &= \begin{bmatrix} \boldsymbol{\Omega}' & \frac{1}{2}\mathbf{C}' \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix}\end{aligned}$$

(the second row is done exactly the same way).

Constrained Optimization, Numerical Example

- ▶ This order of multiplication on the left-hand side is essential so that the known quantities are in the matrix and the unknown quantities are in the first vector.
- ▶ If we move the matrix to the right-hand side by multiplying both sides by its inverse (presuming it is nonsingular of course), then all the unknown quantities are expressed by the known quantities.
- ▶ So a solution for $[\mathbf{x} \ \boldsymbol{\lambda}]'$ can now be obtained by matrix inversion and then pre-multiplication:

$$\begin{aligned}
 \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\Omega}' & \frac{1}{2}\mathbf{C}' \\ \mathbf{C} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\omega} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 5 & 4 & 1 & 1.0 & 0.5 \\ 1 & 5 & 5 & 0.5 & -0.5 \\ 5 & 5 & 5 & 0.0 & -0.5 \\ 2 & 1 & 0 & 0.0 & 0.0 \\ 1 & -1 & -1 & 0.0 & 0.0 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \\ 0 \end{bmatrix} \\
 &= \begin{bmatrix} \frac{1}{18} & -\frac{1}{9} & \frac{1}{6} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{1}{9} & \frac{2}{9} & -\frac{1}{3} & \frac{4}{3} & -\frac{2}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{2} & -\frac{2}{3} & 0 \\ \frac{4}{9} & \frac{10}{9} & -\frac{2}{3} & -\frac{4}{3} & \frac{8}{3} \\ \frac{10}{9} & -\frac{20}{9} & \frac{4}{3} & \frac{10}{3} & -\frac{10}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \\ 0 \end{bmatrix} = \left[\frac{1}{3}, -\frac{2}{3}, 1, \frac{2}{3}, \frac{2}{3} \right]',
 \end{aligned}$$

meaning that $x_1 = \frac{1}{3}$, $x_2 = -\frac{2}{3}$, $x_3 = 1$, and $\boldsymbol{\lambda}' = \left[\frac{2}{3}, \frac{2}{3} \right]$.

Constrained Optimization, Numerical Example

- ▶ Since the λ values are nonzero, we know that the constraints changed the optimization solution.
- ▶ It is not essential, but we can check that the solution conforms to the constraints:

$$2 \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} -2 \\ 3 \end{pmatrix} = 0 \qquad \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \begin{pmatrix} -2 \\ 3 \end{pmatrix} - (1) = 0.$$

- ▶ Suppose we now impose just one constraint: $3x_1 + x_2 - x_3 = 0$. Thus the \mathbf{C} matrix is a vector according to $[3, 1, -1]$ and the $\mathbf{\Omega}$ matrix is smaller.
- ▶ The calculation of the new values is

$$\begin{bmatrix} 5 & 1 & 5 & -0.5 \\ 1 & 5 & 5 & 0.5 \\ 5 & 5 & 5 & -0.5 \\ 3 & 1 & -1 & 0.0 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} & -\frac{7}{20} & \frac{1}{4} & -\frac{3}{10} \\ -\frac{9}{25} & \frac{17}{25} & -\frac{2}{5} & \frac{26}{25} \\ \frac{6}{25} & -\frac{37}{100} & \frac{7}{20} & -\frac{43}{50} \\ \frac{4}{5} & -\frac{2}{5} & 0 & -\frac{6}{5} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ -\frac{1}{5} \\ \frac{11}{20} \\ 0 \end{bmatrix}.$$

- ▶ The $\lambda = 0$ here means is that our restriction actually made no impact: the solution above is the unconstrained solution.
- ▶ So we imposed a constraint that would have been satisfied anyway.