MAX-MIN PROBLEMS

MATH 152, SECTION 55 (VIPUL NAIK)

Corresponding material in the book: Section 4.5

Difficulty level: Moderate to hard. This is material that you have probably seen at the AP level, but it is very important and there will be many additional subtleties that you may have glossed over earlier.

What students should definitely get: The basic procedure for converting a verbal or real-world optimization problem into a mathematical problem seeking absolute maxima and absolute minima, solving that problem, and reinterpreting the solution in real-world terms.

What students should hopefully get: Important facts about area-perimeter optima. The idea that the maximum is determined by the minimum, or most binding, constraint. The intuiton of tangency (as seen in the tapestry problem). The multiple use heuristic. The idea of transforming a function into an equivalent function that is easier to optimize. The procedure for and subtleties in integer optimization. How single-variable optimization fits into the broader optimization context.

Executive summary

Words...

(1) In real-world situations, maximization and minimization problems typically involve multiple variables, multiple constraints on those variables, and some objective function that needs to be maximized or minimized.
(2) The only thing we know to solve such problems is to reduce everything in terms of one variable. This is typically done by *using up* some of the constraints to express the other variables in terms of that variable.
(3) The problem then typically boils down to a maximization/minimization problem of a function in a single variable over an interval. We use the usual techniques for understanding this function, determining the local extreme values, determining the endpoint extreme values, and determining the absolute extreme values.

Actions... (think of examples; also review the notes on max-min problems)

(1) Extremes sometimes occur at endpoints but these endpoints could correspond to degenerate cases. For instance, of all the rectangles with given perimeter, the square has the maximum area, and the minimum occurs in the degenerate case of a rectangle where one side has length zero.
(2) Some constraints on the variables we have are explicitly stated, while others are implicit. Implicit constraints include such things as nonnegativity constraints. *Some of these implicit constraints may be on variables other than the single variable in terms of which we eventually write everything.*
(3) After we have obtained the objective function in terms of one variable, we are in a position to throw out the other variables. However, before doing so, it is *necessary to translate all the constraints into constraints on the one variable that we now have.*
(4) When our intent is to maximize a function, it is sometimes useful to maximize an equivalent function that is easier to visualize or differentiate. For instance, to maximize $\sqrt{f(x)}$ is equivalent to maximizing $f(x)$ if $f(x)$ is nonnegative. With this way of thinking about equivalent functions, we can make sure that the actual function that we differentiate is easy to differentiate. The main criterion is that the two functions should rise and fall together. (Analogous observations apply for minimizing) Remember, however, that to calculate the *value* of the maximum/minimum, you should go back to the original function.
(5) Sometimes, there are other parameters in the maximization/minimization problem that are *unknown constants*, and the final solution is expected to be in terms of those constants. In rare cases, the nature of the function, and hence the nature of maxima and minima, depends on whether those

constants fall in particular intervals. *If you find this to be the case, go back to the original problem and see whether the real-world situation it came from constrains the constants to one of the intervals.*

(6) For some geometrical problems, the maximization/minimization can be done trigonometrically. Here, we make a clever choice of an angle that controls the *shape* of the figure and then use the trigonometric functions of that angle. This could provide alternate insight into maximization.

Smart thoughts for smart people ...

(1) Before getting started on the messy differentiation to find critical points, think about the constraints and the endpoints. Is it obvious that the function will attain a minimum/maximum at one of the endpoints? What are the values of the function at the endpoints? (If no endpoints, take limiting values as you go in one direction of the domain). Is there an intuitive reason to believe that the function attains its optimal value somewhere *in between* rather than at an endpoint? Is there some kind of trade-off to be made? Are there some things that can be said qualitatively about where the trade-off is likely to occur?

(2) Feel free to convert your function to an equivalent function such that the two functions rise and fall together. This reduces the burden of messy expressions.

(3) It is useful to remember the fact that the function $x^p(1-x)^q$ attains a local maximum at $p/(p+q)$. That's because this function appears in disguise all the time (e.g., maximizing area of rectangle with given perimeter, etc.)

(4) A useful idea is that when dividing a resource into two competing uses, and one use is hands-down better than the other, the *best* use happens when the entire resource is devoted to the better use. However, the *worst* may well happen somewhere in between, because divided resources often perform even worse than resources devoted wholeheartedly to a bad use. This is seen in perimeter allocation to boundaries with the objective function being the total area, and area allocation to surfaces with the objective function being the total volume.

(5) When we want to *maximize* something subject to a collection of many constraints, the most relevant constraint is the *minimum* one. Think of the ladder-through-the-hallway problem, or the truck-going-under-bridges problem.

1. Motivation and basic terminology

In the previous lecture, we discussed how to compute points and values of absolute maximum and absolute minimum. Our focus now shifts to using these tools and techniques for real-world (or pseudo-real-world) optimization problems. Because the techniques we have developed are so limited, we will be very selective about the nature of the real-world problems that we pick. Nonetheless, we'll see that even with the modest machinery we have built, we have ways of effectively understanding and tackling many real-world problems.

1.1. Notion of constraints and objective function.

In a typical real-world situation, we usually have multiple things interacting. Many of these items can be measured quantitatively, i.e., they can be measured using real numbers. The values of these real numbers may be subjected to further constraints. Those making decisions may have control over some of the variables. Those making decisions are also tasked with trying to maximize some kind of utility that is dependent on these variables, or minimize some kind of cost function that is dependent on these variables. The task is to *choose the variables subject to constraints in the manner that best maximizes that particular utility function or minimizes that particular cost function.*

For example, think of money management, something that you might be familiar with. You have a certain limited amount of money, and there are various things you want to buy with that money. Each thing that you buy gives you a certain amount of satisfaction; however, for most things, the amount of satisfaction varies with how much you buy it. The question is: how do you allocate money between the many competing things in the market so as to get the best deal for yourself? The number of variables in this case is just the number of different items that you can buy in variable quantities. At a broad level, you may choose to spend A on food, B on clothing, and C on extra books to study calculus. If the total money quantity with you is M, then, assuming that you're not one of those who likes to live on credit, you'll have the constraints $A + B + C \leq M$.

Now, there are going to be three functions f, g, and h, where $f(A)$ is the happiness that spending A on food gives you, $g(B)$ is the happiness that spending B on clothing gives you, and $h(C)$ is the happiness (?)

that spending on extra calculus books gives you. Assuming that happiness is additive, what you want to maximize is $f(A) + g(B) + h(C)$. Given specific functional forms for f, g, and h, we hope to use the tools of calculus to tackle this problem.

1.2. **The extremes and the middle path.** There are two schools of philosophical thought that shall contend for our attention here: the school that says that extremes are good, and the school that urges you to follow a middle path – a bit of this and a bit of that. Which of them is right? Depends.

The extremists would say that among the three things: food, clothing, and calculus books, one of them is the best value for money (for instance, in our case, it may be calculus books). This means that every additional unit of money that you spend should go on calculus books. Thus suggests that the way you'll be "happiest" is if you spend all your money on calculus books.

In reality, however, we know that that isn't how things work. The problem? We need a bit of food, a bit of clothing, and a bit of calculus, but beyond a point, more food, clothing and calculus is less helpful. This is obvious in the case of food – too much food at your disposal means that you either eat more than your body can handle or throw food away. It is also obvious in the case of clothing. It may not be that obvious in the case of calculus, but you'll have to take my word for it that there does come a point after which more calculus may not be worth it.

So, basically, this is a three-way trade-off game, and we need to figure out where to make the trade-offs between food, clothing, and calculus. This comes somewhere in between – a local maximum, where shifting resources from any one sector to any other sector reduces utility.

On the other hand, there are situations where extremes are better. Those are situations where it's just no contest between the two options – more of one thing is better than more of the other no matter how much of either you have. So the extremes could be the best option.

Thus, the maximum could occur at the endpoints, but it could also occur in between, as a local maximum, where diverting resources a bit in either direction makes things worse.

1.3. **Humbler matters: one-variable ambitions.** After suggesting that I could help you with managing money, I have to retreat to humbler ground. All the tools we have developed so far are tools that specifically deal with one variable – we've talked of *functions of one variable*, and developed concepts of limits, continuity, and differentiation all in this context. Thus, the kind of budgeting and allocation problems that we encounter in the real world, that involve a plethora of variables, are simply too hard for us to handle with this machinery. This is also a reason why you shouldn't just stop with the 150s, and should go on to study multivariable calculus, but let's now talk of what *can* be done using the one-variable approach.

A priori, you might expect that the one-variable approach can only work when there is only one variable involved. It's actually a little more general.

The one-variable approach can be used for situations where we can use some of the constraints to express all variables in terms of a single variable, wherein the optimization problem simply becomes a problem in terms of that variable. So, even though the problem has more than one variable, we are able to tackle it as a one-variable problem. Here is one example.

For instance, consider the following problem: For all the rectangles with diagonal length c, find the dimensions of the one with the largest area.

To solve this problem, we try to figure out what variables we have control over, and what constraints these variables satisfy. A rectangle is specified by specifying its length and breadth, i.e., the two side lengths. If we call these l and b, the goal is to maximize lb. Also, l and b are subject to the constraint $l^2 + b^2 = c^2$, and $l > 0, b > 0$.

The problem is that we have two variables, and we only know how to tackle situations with one variable. In order to solve the problem, we need to write one of the variables in terms of the other one. Note that the relation $l^2 + b^2 = c^2$, along with the fact that $l > 0$ and $b > 0$, allows us to write $b = \sqrt{c^2 - l^2}$. Thus, the area is given by a function of l, namely:

$$A(l) := l\sqrt{c^2 - l^2}$$

Note that $\sqrt{c^2 - l^2} > 0$ implies that $l < c$. Thus, the goal is to maximize this function on the interval $(0, c)$.

We compute the derivative:

$$A'(l) = \sqrt{c^2 - l^2} - \frac{l^2}{\sqrt{c^2 - l^2}}$$

Setting $A'(l) = 0$, we obtain that:

$$\sqrt{c^2 - l^2} = \frac{l}{\sqrt{c^2 - l^2}}$$

Simplifying, we obtain:

$$l = b = c/\sqrt{2}$$

and thus:

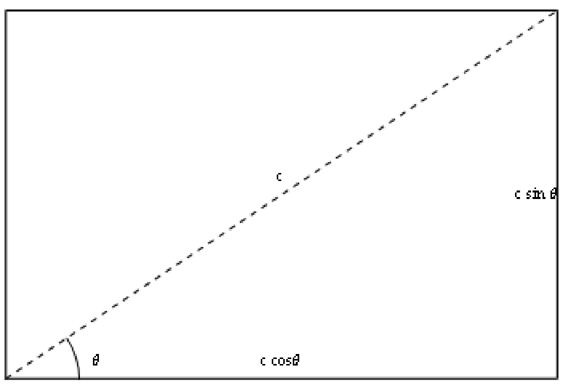
$$A(l) = c^2/2$$

Thus, the only critical point for the function is at $l = c/\sqrt{2}$. Note that for $l < c/\sqrt{2}$, we have $l < \sqrt{c^2 - l^2}$, so the expression for $A'(l)$ is positive, and for $l > c/\sqrt{2}$, the expression is negative. Thus, the point $l = c/\sqrt{2}$ is a point of local maximum.

To determine whether it is a point of absolute maximum, we need to verify that the value of the function at this point is greater than the limits at the two endpoints. It is easy to see that the limits at both endpoints are zero, so indeed, we have a local maximum at $l = c/\sqrt{2}$.

Here is an alternative approach, that involves a different way of choosing variables. Let θ be the angle made by the diagonal with the base of the rectangle. Then, $0 < \theta < \pi/2$, and the two sides of the rectangle have length $c\cos\theta$ and $c\sin\theta$. Thus, the area is given by the function:

$$f(\theta) = c^2 \sin\theta\cos\theta = (c^2/2)\sin(2\theta)$$



This attains an absolute maximum at $\theta = \pi/4$, where $2\theta = \pi/2$, so that $\sin(2\theta) = 1$. Note that we can solve the problem in this case even without using calculus, but if you don't notice that $\sin\theta\cos\theta = (1/2)\sin(2\theta)$, you can solve the problem the calculus way and obtain that the absolute maximum is at $\theta = \pi/4$. Thus, the area is $c^2/2$ and the two side lengths are $c/\sqrt{2}$.

Thus, the maximum occurs for a square.

1.4. **Geometrical and visual optimization.** In most of the situations that we'll be dealing with, it is helpful to draw a figure, label all the lengths and/or angles involved in the figure, and then write down the various constraints as well as the objective function that needs to be maximized. Then, try to get everything in terms of one variable, using the constraints, and finally, do the maximization for that variable. The book gives the following five-point procedure on Page 183:

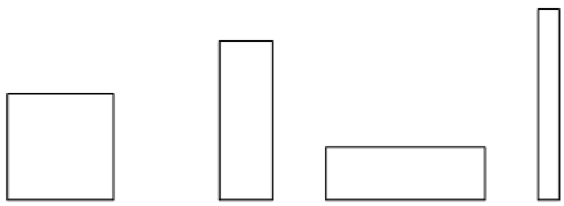
(1) Draw a representative figure and assign labels to the relevant quantities.
(2) Identify the quantity to be maximized or minimized and find a formula for it.
(3) Express the quantity to be maximized or minimized in terms of a single variable; use the conditions given in the problem to eliminate the other variable(s).
(4) Determine the domain of the function generated by Step 3.

(5) Apply the techniques of the preceding sections to find the extreme value(s).

One of the important things in this is to notice that we usually need to maximize the function with the input variable restricted to a certain domain. Thus, there are often situations where the absolute maximum or minimum occurs at an endpoint of the domain, i.e., it is an endpoint maximum/minimum.

Here are some examples that you should keep in mind, both in terms of the final results and the methods we use to get them:

(1) Of all the rectangles with a given perimeter, the square has the largest area. This boils down to maximizing $l((p/2) - l)$. There is no rectangle with minimum area – the minimum area occurs in the *degenerate rectangle* where one of the sides has zero length and the other side has length half the perimeter. The degenerate rectangle isn't ordinarily considered a rectangle. Here are pictures of a collection of rectangles with the same perimeter. It is visually clear that the square has the largest area.



(2) Conversely, of all the rectangles with a given area, the square has the smallest perimeter. This boils down to minimizing $2(l + (A/l))$. There is no rectangle with the largest perimeter – we can keep getting longer and thinner rectangles.

(3) Of all the rectangles with a given diagonal length, the square is the one with the largest area. This boils down to maximizing $l\sqrt{c^2 - l^2}$. Trigonometrically, it involves maximizing $\cos\theta \sin\theta$. The minimum again occurs for the degenerate rectangle, hence does not occur for any actual rectangle.

(4) Of all the rectangles with a given diagonal length, the square is the one with the largest perimeter. This boils down to maximizing $l + \sqrt{c^2 - l^2}$. Trigonometrically, it involves maximizing $\cos\theta + \sin\theta$. The minimum again occurs for the degenerate rectangle, hence does not occur for any actual rectangle.

1.5. **Applications to real-world physical situations.** We see a common concern that is apparent with all these maximization/minimization problems. Maximizing the area for a given perimeter, or minimizing the perimeter for a given area, is a concern that arises when trying to create containers with as little material used for the boundary as possible. Maximizing the area for a given diagonal length or constraints on lengths occurs in situations where concerns of space availability and fitting stuff are paramount.

Here are some of the quantities and formulas that are useful:

(1) For a right circular cylinder with base radius r and height h, the total volume (or capacity) is $\pi r^2 h$. The curved surface area is $2\pi rh$. Each of the disks at the ends has area πr^2. Thus, a right circular cylinder closed at one end has surface area $\pi r(2h + r)$ and a right circular cylinder closed at both ends has surface area $2\pi r(r + h)$. Based on the situation at hand, we need to figure out which of these three surface areas is being refered to.

(2) For a right circular cone with base radius r, vertical height h and slant height l, the volume is $(1/3)\pi r^2 h$. The curved surface area is πrl and the surface area of the base is πr^2, so the total surface area is $\pi r(r + l)$. Again, we need to figure out, based on the situation, which of the surface areas is being refered to. Also note that r, h, and l are related by the Pythagorean theorem: $l^2 = r^2 + h^2$.

(3) For a semicircle of radius r, the area is $(1/2)\pi r^2$. The length of the curved part is πr and the length of the straight part (the diameter) is $2r$, so the total perimeter is $r(\pi + 2)$. More generally, for a sector of the circle bounded by two radii and an arc, where the radii make an angle of θ, the perimeter is $r(2 + \theta)$ and the area is $(1/2)\theta r^2$.

(4) For a sphere, the surface area is $4\pi r^2$ and the volume is $(4/3)\pi r^3$. For a hemisphere, the surface area is $3\pi r^2$ ($2\pi r^2$ for the curved part and πr^2 for the bounding disk) and the volume is $(2/3)\pi r^3$.

Here are some important results on optimization in these various examples:

(1) For a right circular cylinder with volume V, there is no minimum and no maximum for the curved surface area. This is because for given radius r, the expression for the curved surface area is $2V/r$, which approaches ∞ as $r \to 0$ (smaller and smaller radius, larger and larger height) and approaches 0 as $r \to \infty$ (larger and larger radius, smaller and smaller height). If, however, we have additional boundary constraints on the radius or height, the maximum/minimum will occur at these boundaries.
(2) For a right circular cylinder with volume V, there is an absolute minimum for the surface area of the base plus curved part (i.e., only one bounding disk is included). The expression is $2V/r + \pi r^2$. As $r \to 0$ or $r \to \infty$, this expression tends to ∞. The absolute minimum occurs at the point $r = (2V/\pi)^{1/3}$. (see also Example 1 from the book).
(3) For a right circular cylinder with volume V, there is an absolute minimum for the total surface area (including both disks). The expression is $2V/r + 2\pi r^2$. As $r \to 0$ or $r \to \infty$, this tends to infinity. The absolute minimum occurs at the point $r = (V/\pi)^{1/3}$.

2. Important tricks in real-world problems

2.1. **The maximum is determined by the tightest constraint.**

Let me first state this mathematically (where it's obvious) and then non-mathematically (where again it's obvious).

Suppose x is a real number subject to the constraints $x \le a_1$, $x \le a_2$, and $x \le a_3$. What is the *maximum* value that x can take? Clearly, it is the *minimum* among a_1, a_2, and a_3, because that is the tightest, most limiting constraint on x.

Here are some non-mathematical formulations:

(1) A truck has to go on a highway. As part of its journey, the truck needs to negotiate three underpasses, with clearances of 10 feet, 9 feet, and 11 feet respectively. What is the *maximum* possible height of the truck? (Hint: You want to make sure you don't get into a problem anywhere).
(2) Hydrogen and oxygen combine in a ratio of 1 : 8 by mass to produce water. Assuming that we have 50 grams of hydrogen and 220 grams of oxygen, what is the *maximum* possible amount of water that can be produced from these? (Hint: Limiting reagent).

To repeat: *the maximum value that something can take is determined by the tightest of the upper bounds on it.* The importance of this idea cannot be over-emphasized. On the one hand, it is a staple of a whole branch of graph theory/network theory results called max-min theorems. All of them have the flavor that the upper end of what's possibility coincides with the lower end of the constraints. On the other hand, it is also the whole basis for the theory of least upper bounds and greatest lower bounds that we will see in 153 and that forms the basis for a rigorous study of the reals (which you might see if you proceed to study real analysis).

2.2. **Some random tricks.**

A real-world optimization problem is not usually given in a ready-to-solve form. Rather, some decisions and judgments need to be made about the procedure and the general form of the solution in order to obtain a mathematical setup.

The initial judgment may use general rules: for instance, the rule that straight line paths, where possible, are shorter than non-straight line paths. Thus, when asked to find a shortest path subject to certain constraints, we may be able to narrow it down to a straight line path and then do the optimization within that.

As a somwhat trickier example, consider the following problem, which appears on your homework:

> Two hallways, one 8 feet wide and the other 6 feet wide, meet at right angles. Determine the length of the longest ladder that can be carried horizontally from one hallway to the other.

Here, the significance of *horizontally* is simply that the ladder cannot be tilted vertically, a strategy that would enable one to carry a longer ladder. This problem is a hard one because the nature of the constraint is not clear. How does the width of the hallways constrain the length of the ladder that can be passed through?

We need to role-play the *process of carrying the ladder.* When a ladder is being carried along a corridor, it makes the most sense to align the ladder parallel to the walls of the corridor. When the direction of the corridor changes, the ladder needs to be rotated to align it with the new corridor. We must be able to rotate the ladder through every angle. This leads to the constraint: for every angle, the ladder must fit in. We then try to find, for every angle θ, the maximum length of ladder that can fit in at the junction between the two corridors. Each of these imposes a constraint on the length of ladder. The most relevant binding constraint is the *minimum* of these lengths.

2.3. **The intuition of tangency.** Let's now look at another problem that also appears on your homework:

> A tapestry 7 feet high hangs on a wall. The lower edge is 9 feet above an observer's eye. How far from the wall should the observer stand in order to obtain the most favorable view? Namely, what distance from the wall maximizes the visual angle of the observer?

Here's the intuition behind this problem. If you stand right under the tapestry, it seems *foreshortened*. If, however, you go very far, then it simply seems small. The quantity that measures how large the tapestry appears is the visual angle, or the angle between the lines joining your eyes to the top and bottom of the tapestry. This angle is zero if you are right under the tapestry, and it approaches zero as you go out far from the tapestry. Where is it maximum? Somewhere in between. But where exactly?

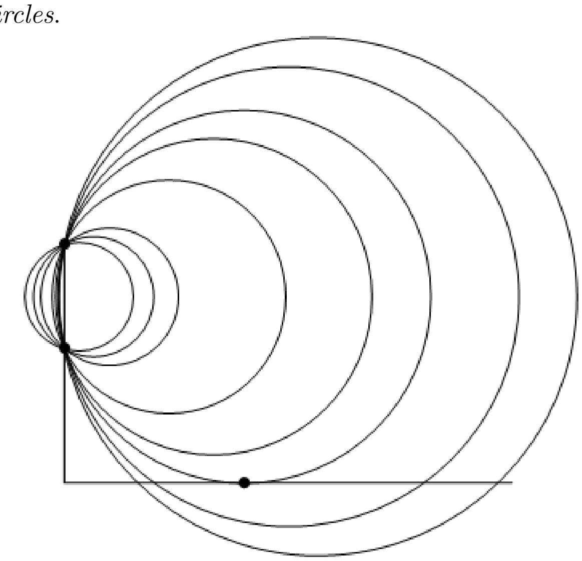
You can find this using calculus – which is what you are expected to do in this homework. But there is an alternative, related approach that is more geometric.

The main geometric fact used is that the angle subtended by a chord of a circle at any two points on the circle on the same side of the chord is the same.

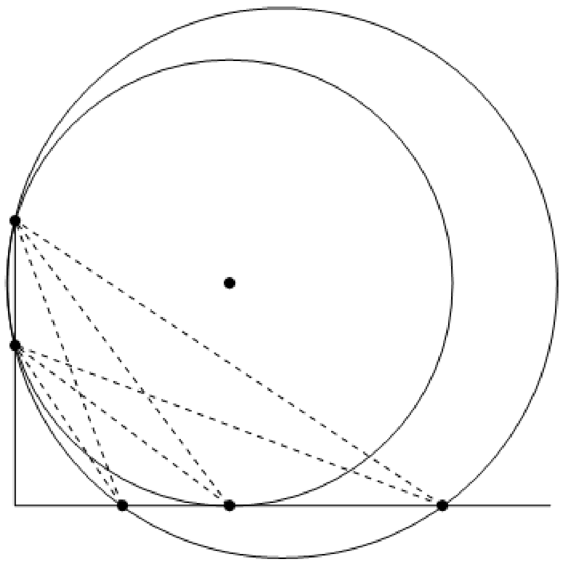
Now, consider a circle passing through the two ends of the tapestry. If this circle intersects the horizontal line of possible locations of your eye at two points P and Q, then by the result I mentioned, the visual angle at P equals the visual angle at Q. Note that for a very large circle, P is very close to the base of the tapestry and Q is very far away. This helps explain why small visual angles are achieved both very close and very far away from the tapestry.

We also see that the smaller the circle, the larger the visual angle. Thus, the goal is to find the smallest circle passing through the two ends of the tapestry that intersects the horizontal line of possible locations of the eye. A little thought reveals that this occurs when the circle is tangent to the horizontal line. If you imagine starting with a very large circle and shrinking it further and further, the circle that is tangent to the horizontal line is the one at which the circle just leaves the horizontal line. (Having deduced this, it is possible to determine the precise point using geometry and algebra, without any calculus. You can verify the answer you obtain using calculus via this method).

Here's the picture with lots of such circles drawn. Such a system of circles is called a *coaxial system of circles*.



Here's the same picture with just the circle of tangency and another circle drawn:



Note: We can use another result of geometry to calculate the distance of the point of tangency from the foot of the tapestry. Namely, the result says that if P is a point outside a circle, PT is a tangent to the circle with point of tangency T, and a secant line through P intersects the circle at A and B, then $PA \cdot PB = PT^2$. We can use this to calculate PT as the square root of the product of the distances from the base of the bottom and top of the tapestry.

2.4. **The heuristic of multiple uses.** Suppose a resource (such as fencing wire, which plays the role of perimeter) is to be divided among two alternative uses: say a square fence, and a circular fence. It is a fact that, of all possible shapes with a given perimeter, the circle encloses the largest area. (This is called the *isoperimetric problem*, and although we will not show it, it is useful to remember). In particular, devoting all the fencing to the circle yields a larger area than devoting all the fencing to the square. (This can be checked easily by algebra, and the fact that $\pi < 4$).

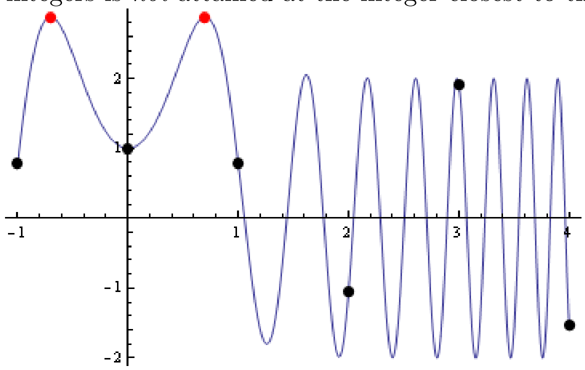
Given this, what is the way of allocating fencing so as to get the maximum and minimum possible total area? It turns out that for the *maximum possible*, we allocate all resources to the hands-down better use, which is in this case the circle. However, for the *minimum possible*, the strategy is *not* that of allocating everything to the square. Why not? It turns out that we can do even worse by providing some fencing to the square and some fencing to the circle? Why? Because there is some wastage that arises simply from having two fences. Even though a square is less efficient than a circle, devoting everything to the square is a little more efficient than devoting mostly to the square and a bit for the circle. A problem of this kind appears in Homework 5.

In the good old days when everybody farmed, each time a farmer with multiple sons died, his land was divided among the sons. As a result, fencing costs and wastage kept increasing. One solution to this was *primogeniture laws*, which stated that the eldest son was entitled to the land. While not fair, these laws helped combat the problem of fragmentation of land holdings.

2.5. **Integer optimization.** A few brief notes on integer optimization may be worthwhile. In many real-world situations, the possible values that a variable can take are constrained to be integers. For instance: *how many passengers can travel on this vehicle*? The optimization here thus requires one to optimize *subject to the integer value constraint on the variables.*

It may seem reasonable at first to believe that the best integer solution is the integer closest to the best real solution. This is not always the case. In fact, computer scientists have shown that even solving systems of linear equations and inequalities in integer variables has no general-purpose algorithm that runs quickly (subject to a long-standing conjecture called $P \neq NP$). This is despite the fact that the analogous problem is very easy to solve for real variables.

The problem here is that the value of a function can change very rapidly between a real number and the integers closest to it. See, for instance, this picture for a function where the maximum value among values at integers is *not* attained at the integer closest to the absolute maximum:



However, it is useful to look at the behavior of a function over all real numbers in order to determine the integers where it attains maxima and minima. For instance, if we can determine where the function is increasing and decreasing, we can use this information along with testing some values in order to find out the maxima and minima. Specifically, what we first do is *extend the function to all real numbers* (by considering the definition of the function applied to all real numbers) and find the intervals where the function is increasing and decreasing as a function with real inputs. Then:

(1) If f is increasing on an interval, then the minimum of f on that interval occurs at the smallest integer in the interval and the maximum occurs at the largest integer in the interval.
(2) If f is decreasing on an interval, then the minimum of f on that interval occurs at the largest integer in the interval and the minimum occurs at the smallest integer in the interval.
(3) If we need to find the absolute maximum of f over all integers, we can first break up into intervals where f is increasing/decreasing, find the maximum over each of those intervals, and finally compare the values of all these maxima to find which is the largest one.

Here are some simple examples:

(1) Consider a function that is decreasing on $(-\infty, 1.3]$ and increasing on $[1.3, \infty)$. Then, if viewed as a function on reals, this function has a unique absolute minimum at 1.3. As a function on integers, we know that the function is increasing from 2 onwards, so the value for any integer greater than 2 is greater than the value at 2. Similarly, we know that the value at any integer less than 1 is greater than the value at 1. So, there are two candidates for the absolute minimum among integers: the values at 1 and 2. We now calculate the values at 1 and 2 and find which one is smaller.
(2) Consider a function that decreases on $(-\infty, -1.1]$, increases on $[-1.1, 0.1]$, decreases on $[0.1, 0.9]$, and then increases on $[0.9, \infty)$. On the interval $(-\infty, -1.1]$, the minimum among integers is at -2. On the interval $[-1.1, 0.1]$, the minimum among integers is at -1. On the interval $[0.1, 0.9]$, there are no integers. On the interval $[0.9, \infty)$, the minimum among integers is at 1. Thus, the three candidates for the point of absolute minimum are -2, -1, and 1.

Note that it is *not* necessarily true that the integer where the absolute minimum among integers is attained is the closest integer to the real number where the absolute minimum among real numbers is attained. This is because the function can change very rapidly between a real number and the closest integer. For certain special kinds of functions (such as quadratic functions), it *is* true that the integer for absolute minimum is the closest integer to the real number for absolute minimum. But this is due to the symmetric nature of quadratic functions – the graph of a quadratic function with positive leading coefficient is symmetric about the vertical line through its point of absolute minimum.[1]

[1] For negative leading coefficient, the corresponding statement is true if we replace absolute minimum by absolute maximum.

3. Some notes from social and natural sciences

3.1. An important maximization: Cobb-Douglas, fair share, and kinetics.

Let's now go to a question considered by some economists in the early twentieth century. We'll then talk about how a similar question comes up in chemical reactions.

Suppose a factory is producing some goods using two kinds of inputs: labor and capital. For a given production process, if the factory spends L on labor and K on capital, the output of the factory is given by $L^a K^b$, where a and b are positive numbers. The goal of the factory is to maximize output for a given expenditure $(L + K)$. In other words, if the factory is spending a total E on labor and capital put together, how should it allocate E between labor and capital to obtain the maximum output?

Since E is fixed, we can choose L as the variable and write $K = E - L$. We thus get that the output is $L^a(E - L)^b$. If we further let $x = L/E$ (the fraction on labor), then the output is given by $E^{a+b}x^a(1 - x)^b$. Thus, in order to maximize output, we need to maximize $x^a(1 - x)^b$, where $x \in [0, 1]$.

A maximization of this sort appears on your homework, and we find there that the absolute maximum on the interval $[0, 1]$ occurs at the point $a/(a + b)$. Thus, the maximum occurs when $L = Ea/(a + b)$ and $K = Eb/(a+b)$. Thus, the labor-to-capital expenditure ratio L/K is a/b – the same as the ratio of exponents.

What this result shows is that the ratio of exponents on labor and capital represents the relative contributions of labor and capital to production. Optimization occurs when the allocation of resources is done according to these relative contributions: a fraction of $a/(a+b)$ to labor and a fraction of $b/(a+b)$ to capital. In hindsight, this makes intuitive sense: the larger the value of a, the more sense it makes to invest in labor, because the return on investment in labor is higher. However, after some point, it also makes sense to invest a bit in capital, otherwise that becomes a bottleneck. The proportion should have something to do with the ratio of a and b. Mathematically, we have shown that these two proportions in fact coincide.

This raises the question of what determines a and b in the first place. This has something to do with the nature of the production process. A *labor-intensive process* would be one where a dominates and a capital-intensive process would be one where b dominates.

All production functions do not look like the function above. However, it was the argument of Cobb and Douglas that assuming functions to be of the above form is a useful simplification and many phenomena of relative allocation of resources to factors of production can be understood this way. In many parts of economics and the social sciences, people wanting to do a simple analysis often begin by assuming that a given production function is Cobb-Douglas, in order to get a clear handle on the relative contribution of different factors.

Another place where a similar formulation pops up is chemical kinetics. Suppose we have a chemical reaction between two substances A and B, with equation of the form $mA + nB \to$ products. The theory of chemical kinetics suggests that, assuming this reaction is elementary, the rate of forward reaction is given by $k_f[A]^m[B]^n$ where k_f is a constant (with suitable dimensions), $[A]$ is the concentration of A and $[B]$ is the concentration of B.

Now, the question may be: for a given total concentration, how do you decide the proportions in which to mix A and B to get the fastest reaction? This is the same problem in a new guise, and it turns out that the maximum occurs when $[A] : [B] = m : n$. This is poetic justice, because this is precisely the *right* ratio from the *stoichiometric* viewpoint.

3.2. Frontier curves and optimal allocation.

An important concept, which you may first see in economics courses, but which also occurs elsewhere, is that of a *production possibility frontier* or *production possibility curve*. Let's understand these curves in the language of optimization.

Suppose you are running a farm that can produce only two things: wheat and rice. Now, let's say that you decide to produce $50,000$ bushels of wheat. Given this constraint, your goal is to produce as much rice as possible. This is now an optimization problem and you somehow solve it and find out that you can produce at most $25,000$ bushels of rice if you want to produce $50,000$ bushels of wheat.

Now, if you instead wanted to produce only $40,000$ bushels of wheat, it is possible that you can produce more – say $40,000$ bushels of rice. Thus, *for each quantity of wheat that you choose to produce*, there is a maximum quantity of rice you can produce with the given resources. We can thus define a *function* that takes as input the quantity of wheat and outputs the maximum quantity of rice that can be produced alongside. This is a *decreasing* function (the more wheat you produce, the less resources you can devote to producing rice) and its domain is from 0 to the maximum amount of wheat that you can produce. The largest value in the domain is the maximum amount of wheat you can produce if you devote all resources to wheat production, and the value of the function at 0 if the maximum amount of rice you can produce if you devote all your resources to rice production.

The graph of this function is called the *production possibility curve* or *production possibility frontier*. The important thing to note about this graph is that *every point on the graph is an optimal point in some sense* – there is no way to unambiguously improve from any of these points. Any point below a point on the curve, or on the inside of the curve, is achievable but non-optimal, in the sense that it is possible to increase the production of one or both the outputs without decreasing the other one. A point above or outside the production possibility frontier is a point that cannot be achieved, reflecting the *reality of scarcity* or the *limitations of current technology*, depending on your perspective.

3.3. **Can spontaneous processes solve optimization problems?** First, a little clarification on what the question means. In all the situations we have seen so far, there is a conscious agent that is using calculus to find an optimal allocation or optimal value by explicitly considering constraints. But optimization has been a goal for living creatures and for nature long before the advent of calculus. How did they do it?

For instance, bubbles tend to be spherical in order to minimize their surface area. More generally, the shapes that soap films can take are minimal surfaces – they minimize surface area. But are bubbles and soap films solving a complicated optimization problem by choosing a spherical shape? Do we need to posit a theory of consciousness and calculus ability every time we see such optimization in the physical or biological world?

No. Physical entities (and most primitive biological entities) are not trying to reach an optimal state – they simply *keep moving around until* they hit upon a *stable equilibrium*, which is *locally optimal*. In fact, the same is true for humans interacting in a large market. This point is extremely important.

For instance, here are some crude heuristics:

(1) In the world of physics, the reason why mechanical or physical systems tend to certain "optimal" configurations is that in these configurations, there are no forces rending them apart or causing them further change.
(2) In the world of chemistry, materials keep reacting until they reach a configuration where the push to the reaction in one direction equals the push to reaction in the other direction.
(3) In the world of biology, living creatures explore the space around them till they hit on something that's better than the stuff around it.
(4) In the word of economics, each individual keeps making changes in the variables under his/her economic control until reaching a situation where a change in either direction is not to his/her advantage.

The upshot is that local optima tend to be places of stability simply because there isn't a tendency to deviate either way, not because anybody did calculus. You can think of it like an ant moving along the graph of a curve and stopping when it gets to a peak and would need to go down both ways.

But this also has a flip side:

- Local optima need not be absolute optima. That was the whole point of our earlier lecture on the subject! But given their stability, natural systems may stay stuck at these local optima. To get to an even bigger global optima, a *push* may be needed. (For instance, activation energy in the context of a chemical reaction, or the entry of a new competitor in a stagnating and non-innovating industry).
- In some cases, there may be so many different local optima, or the situation may be so shaky, that there is never any place to settle down at. In some cases, inertia may prevent settling down. This causes such phenomena as *oscillatory* and *chaotic* behavior.