Khan Academy

Lagrange multipliers, introduction

The "Lagrange multipliers" technique is a way to solve constrained optimization problems. Super useful!

Background

- Contour maps
- Gradient
- Local maxima and minima

What we're building to:

- The Lagrange multiplier technique lets you find the maximum or minimum of a multivariable function $f(x, y, \dots)$ when there is some constraint on the input values you are allowed to use.

- This technique only applies to constraints that look something like this:

$$g(x, y, \dots) = c$$

Here, g is another multivariable function with the same input space as f, and c is some constant. *[Picture]*

- The core idea is to look for points where the contour lines of f and g are tangent to each other.

- This is the same as finding points where the gradient vectors of f and g are parallel to each other.

- The entire process can be boiled down into setting the gradient of a certain function, called the **Lagrangian**, equal to the zero vector. *[Specifically...]*

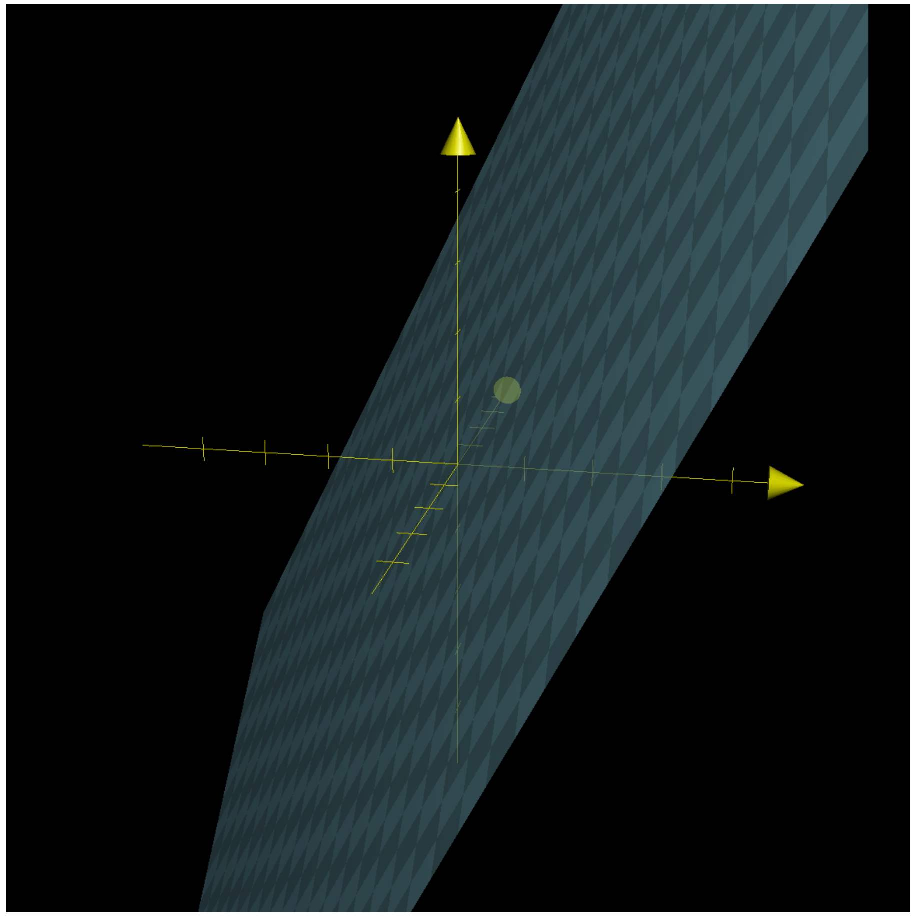
Motivating example

Suppose you want to maximize this function:

$$f(x, y) = 2x + y$$

Math › Multivariable calculus › Applications of multivariable derivatives › Constrained optimization (articles)

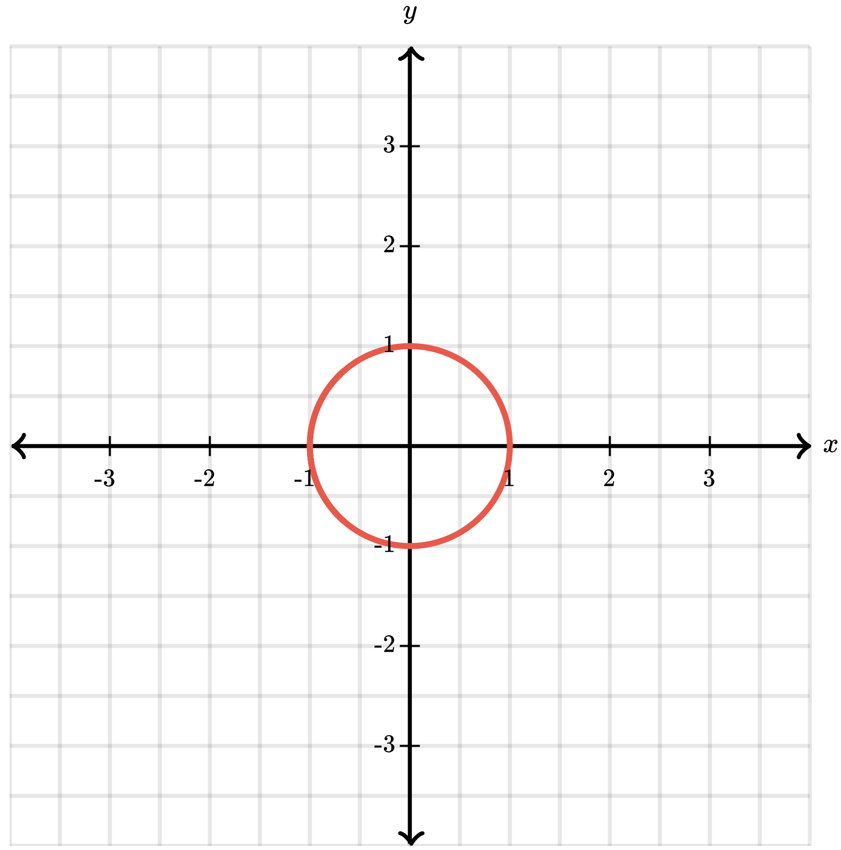
Constrained optimization (articles)



Plot of the function $f(x, y) = 2x + y$

But let's also say you limited yourself to inputs (x, y) which satisfy the following equation:

$$x^2 + y^2 = 1$$

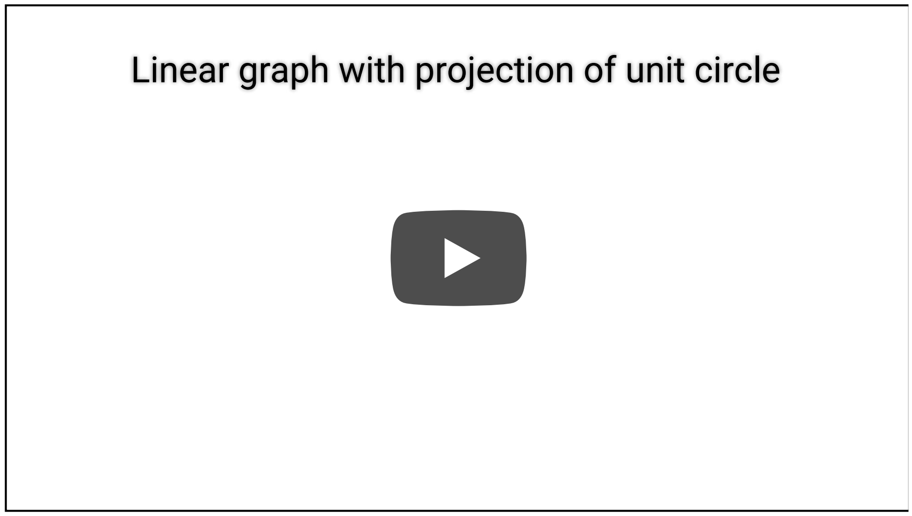


All points (x, y) satisfying $x^2 + y^2 = 1$, also known as the unit circle.

In other words, for which point (x, y) on the unit circle is the value $2x + y$ biggest?

This is what's known as a **constrained optimization problem**. The restriction to points where $x^2 + y^2 = 1$ is called a "constraint", and $f(x, y) = 2x + y$ is the function that needs to be optimized.

Here's one way to visualize this: First draw the graph of $f(x, y)$, which looks like a slanted plane since f is linear. Next, project the circle $x^2 + y^2 = 1$ from the xy-plane vertically onto the graph of f. The maximum we are seeking corresponds with the highest point of this projected circle on the graph.



See video transcript

More general form

In general, constrained optimization problems involve maximizing/minimizing a multivariable function whose input has any number of dimensions:

$$f(x, y, z, \dots)$$

Its output will always be one-dimensional, though, since there's not a clear notion of "maximum" with vector-valued outputs.

The type of constraints that the Lagrange multiplier technique applies to must take the form of some other multivariable function $g(x, y, z, \dots)$ being set equal to a constant c.

$$g(x, y, z, \dots) = c$$

Since this is meant to be a constraint on the input of f, the number of dimensions in the input of g is the same as that of f. For example, the example outlined above fits this general form as follows:

$$f(x, y) = 2x + y$$

$$g(x, y) = x^2 + y^2$$

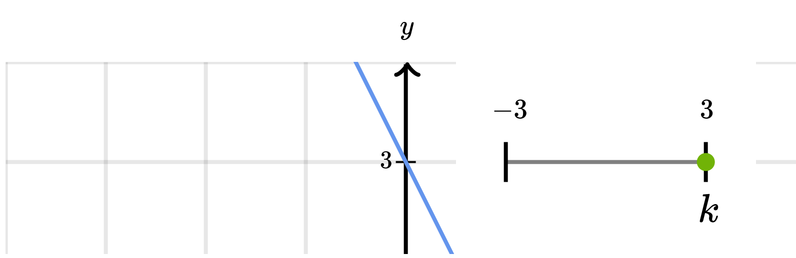
$$c = 1$$

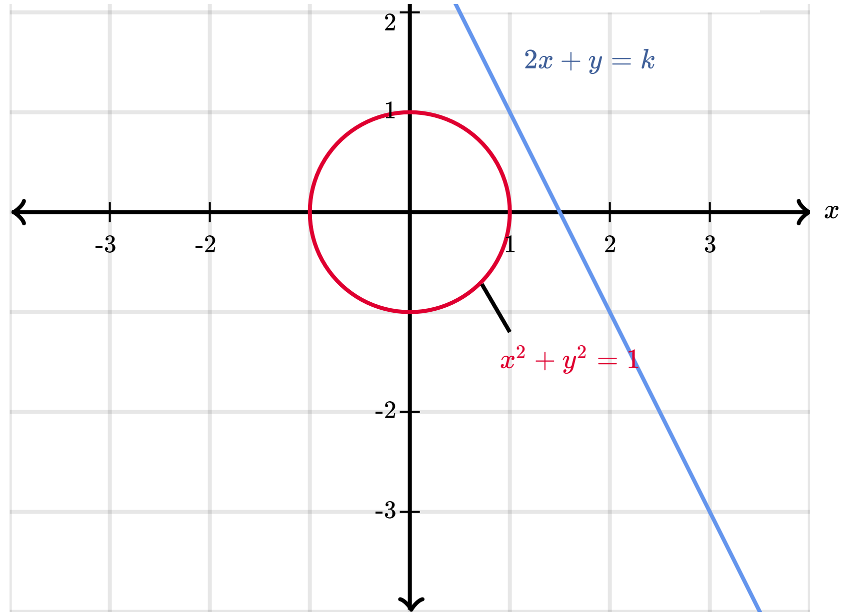
[Multiple constraints]

Using contour maps

Reasoning about this problem becomes easier if we visualize f not with a graph, but with its contour lines.

As a reminder, a contour line of $f(x, y)$ is the set of all points where $f(x, y) = k$ for some constant k. The following interactive tool shows how this line (drawn in blue) changes as the constant k changes. The circle $g(x, y) = 1$ is also shown (in red). Try to make k as big/small as possible while still allowing contour line of f to intersect the circle.





Concept check: What does it mean if for a particular value of k, the blue line representing $f(x, y) = k$ does **not** intersect the red circle representing $g(x, y) = 1$?

Choose 1 answer:

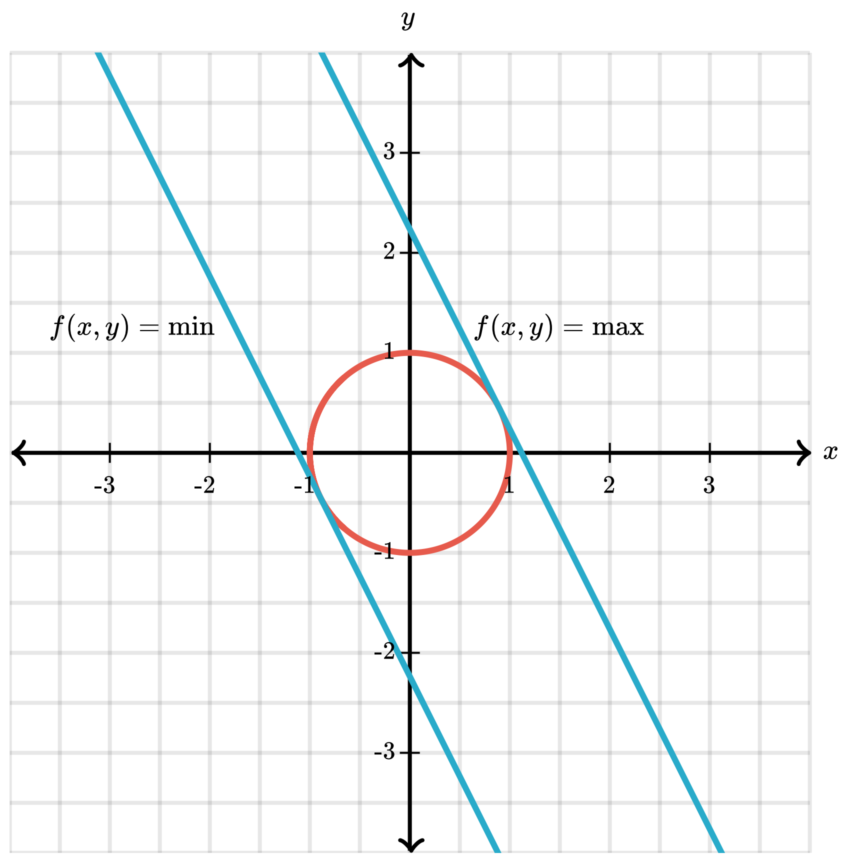
(A) There are no values of x and y satisfying both $2x + y = k$ and $x^2 + y^2 = 1$

(B) The given optimization problem has no solutions.

Check

Notice, the circle where $g(x, y) = 1$ can be thought of as a particular contour line of the function g. So with that, here's the clever way to think about constrained optimization problems:

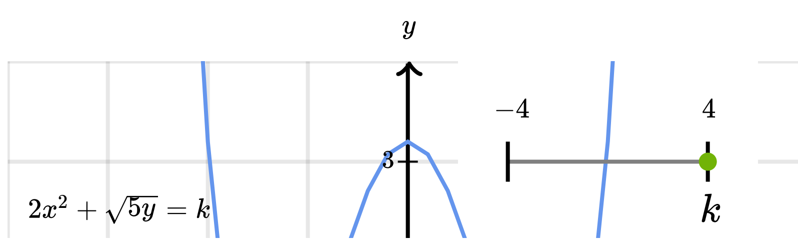
Key observation: The maximum and minimum values of f, subject to the constraint $g(x, y) = 1$, correspond with contour lines of f that are **tangent** to the contour representing $g(x, y) = 1$.

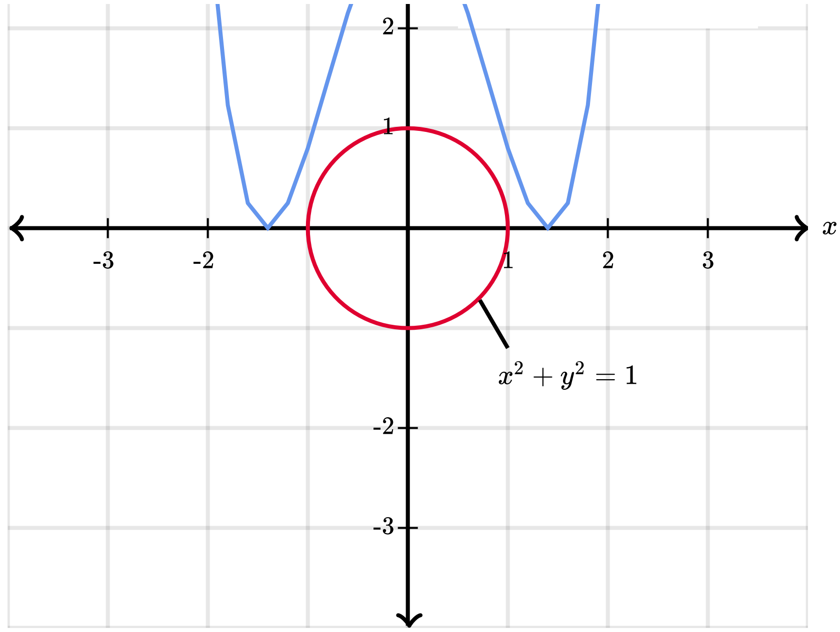


If f were a different function, its contours might not always be straight lines. This is unique to our example since f is linear. For example, take a look at this function:

$$f(x, y) = 2x^2 + \sqrt{5y},$$

Its contour lines look like this:





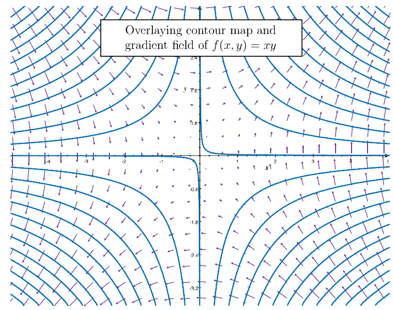
That said, the key observation still holds, and is worth repeating: When k is a maximum or minimum of f subject to the constraint, the contour line for $f(x, y) = k$ will be tangent to contour representing $g(x, y) = 1$.

Where the gradient comes into play

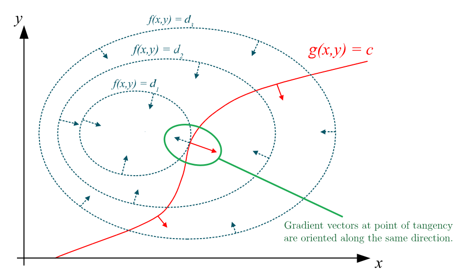
How do you put the idea of two contour lines being tangent into a formula you can solve?

To answer this, we turn to our loyal friend the gradient. There are many ways to interpret ∇f: The direction of steepest ascent, a tool for computing directional derivatives, etc. But for our purposes here, the property we care about is that **the gradient of f evaluated at a point (x_0, y_0) always**

gives a vector perpendicular to the contour line passing through that point.



This means when the contour lines of two functions f and g are tangent, their gradient vectors are parallel. Here's what that might look like for arbitrary functions f and g:



The fact that contour lines are tangent tells us nothing about the magnitude of each of these gradient vectors, but that's okay. When two vectors point in the same direction, it means we can multiply one by some constant to get the other. Specifically, let (x_0, y_0) represent a particular point where the contour lines of f and g are tangent (writing x_0 and y_0 with a 0 subscripts just indicates that we are considering constant values, and hence a specific point). Since this tangency means their gradient vectors align, here's what you might write down:

$$\nabla f(x_0, y_0) = \lambda_0 \nabla g(x_0, y_0)$$

Here, λ_0 represents some constant. Some authors use a negative constant, $-\lambda_0$, but I personally prefer a positive constant, as it gives a cleaner interpretation of λ_0 down the road.

Let's see what this looks like in our example where $f(x, y) = 2x + y$ and $g(x, y) = x^2 + y^2$. The gradient of f is

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial}{\partial x}(2x + y) \\ \\ \frac{\partial}{\partial y}(2x + y) \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and the gradient of g is

$$\nabla g(x, y) = \begin{bmatrix} \frac{\partial}{\partial x}(x^2 + y^2 - 1) \\ \\ \frac{\partial}{\partial y}(x^2 + y^2 - 1) \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Therefore, the tangency condition ends up looking like this:

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \lambda_0 \begin{bmatrix} 2x_0 \\ 2y_0 \end{bmatrix}$$

Solving the problem in the specific case

To sum up where we are so far, we are looking for input points (x_0, y_0) with the following properties:

- $g(x_0, y_0) = 1$, which for our example means

$$x_0^2 + y_0^2 = 1$$

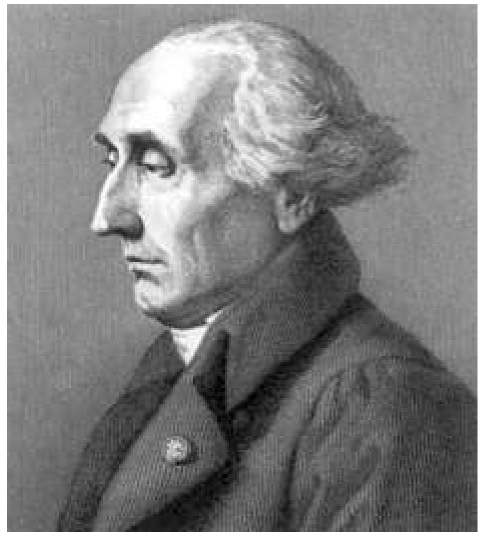
- $\nabla f(x_0, y_0) = \lambda_0 \nabla g(x_0, y_0)$ for some constant λ_0, which for our example means

$$\begin{aligned} 2 &= 2\lambda_0 x_0 \\ 1 &= 2\lambda_0 y_0 \end{aligned}$$

There are 3 equations and 3 unknowns, so this is a perfectly solvable situation.

[See the final solution]

The Lagrangian function



Joseph Louis Lagrange, looking peaceful, content, and sleepy, all at the same time. Wikimedia Commons

In the 1700's, our buddy Joseph Louis Lagrange studied constrained optimization problems of this kind, and he found a clever way to express all of our conditions into a single equation.

You can write these conditions generally by saying we are looking for constants x_0, y_0 and λ_0 that satisfy the following conditions:

- **The constraint**:

  $$g(x_0, y_0) = c$$

- **The tangency condition**:

  $$\nabla f(x_0, y_0) = \lambda_0 \nabla g(x_0, y_0).$$

  This can be broken into its components as follows:

- $f_x(x_0, y_0) = \lambda_0 g_x(x_0, y_0)$
- $f_y(x_0, y_0) = \lambda_0 g_y(x_0, y_0)$

Lagrange wrote down a special new function which takes in all the same input variables as f and g, along with the new kid in town λ, thought of now as a variable rather than a constant.

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c)$$

For example, consider our example above.

$$\begin{aligned} f(x, y) &= 2x + y \\ g(x, y) &= x^2 + y^2 \\ c &= 1 \end{aligned}$$

Here's how this new function would look:

$$\mathcal{L}(x, y, \lambda) = 2x + y - \lambda(x^2 + y^2 - 1).$$

Notice, the partial derivative of $\mathcal{L}$ with respect to λ is $-(g(x, y) - c)$:

$$\begin{aligned} \mathcal{L}_\lambda(x, y, \lambda) &= \frac{\partial}{\partial \lambda}(f(x, y) - \lambda(g(x, y) - c) \\ &= 0 - (g(x, y) - c) \end{aligned}$$

So we can translate the condition $g(x, y) = c$ as

$$\mathcal{L}_\lambda(x, y, \lambda) = -g(x, y) + c = 0$$

What's more, look at what we get when we set one of the other partial derivatives equal to 0:

$$\mathcal{L}_x(x, y, \lambda) = 0$$

$$\frac{\partial}{\partial x}(f(x, y) - \lambda(g(x, y) - c)) = 0$$

$$f_x(x, y) - \lambda g_x(x, y) = 0$$

$$f_x(x, y) = \lambda g_x(x, y)$$

That just so happens to be another one of our conditions! Almost identically, the condition $\mathcal{L}_y(x, y, \lambda) = 0$ unravels to become

$$f_y(x, y) = \lambda g_y(x, y)$$

Together, these conditions are the same as saying.

$$\nabla f(x, y) = \lambda \nabla g(x, y)$$

Therefore, the three conditions we need to solve to find x, y and λ come down to the various partial derivatives of $\mathcal{L}$ being equal to 0. This can be written extremely compactly by setting the gradient of $\mathcal{L}$ equal to the zero vector:

$$\nabla \mathcal{L} = \mathbf{0}$$

For example, using our specific functions from above, we see how this encodes the system of equations we need to solve:

$$\nabla \mathcal{L} = \begin{bmatrix} \frac{\partial}{\partial x}(2x + y - \lambda(x^2 + y^2 - 1)) \\ \frac{\partial}{\partial y}(2x + y - \lambda(x^2 + y^2 - 1)) \\ \frac{\partial}{\partial \lambda}(2x + y - \lambda(x^2 + y^2 - 1)) \end{bmatrix}$$

As a tribute to ol' Joey Lou, we call this function $\mathcal{L}$ the "**Lagrangian**", and the new variable λ that we introduce is called a "**Lagrange multiplier**". Imagine if someone added "-ian" the end of your last name and made it the name of a function everybody uses. Pretty sweet!

Warning: Some authors use a convention where the sign of λ is reversed:

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - c)$$

This doesn't make any difference when it comes to solving the problem, but you should keep it in mind in case the course you are taking or the text you are reading follows this convention.

[What if the constraint isn't so constraining]

Summary

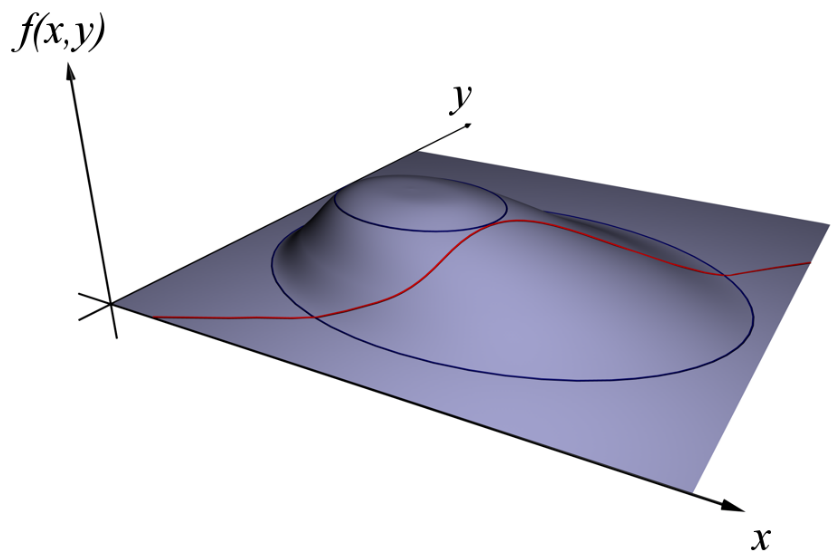


Image credit: By Nexcis (Own work) [Public domain], via Wikimedia Commons

When you want to maximize (or minimize) a multivariable function $f(x, y, \dots)$ subject to the constraint that another multivariable function equals a constant, $g(x, y, \dots) = c$, follow these steps:

- **Step 1**: Introduce a new variable λ, and define a new function $\mathcal{L}$ as follows:

$$\mathcal{L}(x, y, \ldots, \lambda) = f(x, y, \ldots) - \lambda(g$$

This function $\mathcal{L}$ is called the "Lagrangian", and the new variable λ is referred to as a "Lagrange multiplier"

- **Step 2**: Set the gradient of $\mathcal{L}$ equal to the zero vector.

$$\nabla \mathcal{L}(x, y, \ldots, \lambda) = \mathbf{0} \quad \leftarrow \text{Zero vect}$$

In other words, find the **critical points** of $\mathcal{L}$.

- **Step 3**: Consider each solution, which will look something like $(x_0, y_0, \ldots, \lambda_0)$. Plug each one into f. Or rather, first remove the λ_0 component, then plug it into f, since f does not have λ as an input. Whichever one gives the greatest (or smallest) value is the maximum (or minimum) point you are seeking.

Sort by: Top Voted

Questions Tips & Thanks

Want to join the conversation?

Log in

alimuldal 6 years ago

more

In the final "side note" example graph, the equation for the red diagonal line should be x - y = 0, not x + y = 0

Reply • Comment (34 votes) Upvote Downvote

Sam 6 years ago

more

Why is that the maximum or minimum value for f lies at the point where the contour lines of f and g are tangent? How did you prove it more rigorously?

Reply • Comment (7 votes) Upvote Downvote

Dahn Jahn 6 years ago

more

I won't prove this, but imagine them not being tangent, that would mean the function's contours cross the constraint at some point. If they cross, however, it means we can always shift the contour to a higher (or lower, depending of we're maximising or minimising) level and still be crossing the constraint. The time when you cannot shift it any higher is precisely at the level where they're only touching (tangent), not crossing. This is all assuming we have well-behaved functions/constraints

3 comments (13 votes) Upvote Downvote

Zaz Brown

Zaz Brown 6 years ago more

“There's a slight twist to this story, best illustrated with an example.” is repeated.

Reply • Comment (5 votes) Upvote Downvote

Molly Swenson 6 years ago more

Use the method of Lagrange Multipliers to determine the maximum and minimum of f(x,y,z) = x + y + z subject to the two conditions g(x,y,z) = x2 + y2 − 2 = 0 and h(x, y, z) = x + z − 1 = 0

Reply • Comment (2 votes) Upvote Downvote

Alexander Wu 5 years ago more

We haven't learned about multiple constraints yet. The Lagrangian Function for multiple constraints is 𝓛(x1,x2,...,xn,λ1,λ2,...,λM) = f(x1,x2,...,xn) - Σ λkgk(x1,x2,...,xn), where f(**x**) is the function to be maximized and g1, g2, ..., gM are the constraints (the Σ sum is from k = 1 to M). Note that here we use gk(**x**) = 0 instead of gk(**x**) = c.

We take the three gradients and get ∇f = **i** + **j** + **k**, ∇g = 2x **i** + 2y **j**, ∇h = **i** + **k**. ∇f = λ1∇g + λ2∇h, and g = h = 0, so we get: 1 = 2xλ1 + λ2, 1 = 2yλ1, 1 = λ2, x^2 + y^2 = 2, and x + z - 1 = 0. We see that λ2 = 1, so 2xλ1 = 0. because λ1 = 1/2y, x/y = 0, so x must be 0. Then y = +-√2, and λ1 = +-√2/4, and z = 1.

So we get two solutions: (0,√2,1) and (0,-√2,1). I think you can figure out which is the maximum.

If you can graph this, try it, and you'll see how stupid our computations were. After all this work the answers are on the y-z plane!

Comment (4 votes) Upvote Downvote

Kevin 5 years ago

more

Hi, thanks for the article. Question ... how does a parallel gradient tell us that the 2 contour lines are tangent to each other? Parallel gradients should be possible even if the 2 functions don't touch right? Thank you.

Reply • Comment (2 votes) Upvote Downvote

Alexander Wu 5 years ago

more

Actually, parallel gradients tell us that the contour lines are parallel. Contour lines are tangent when they are both touching and parallel at the place they're touching. Of course, we had to answer the question "Where are they parallel?" The direction of gradients and contour lines are different at every point.

Therefore, the gradients of the two functions need to be parallel *at the same point*. $\nabla f(0,1)$ being parallel to $\nabla g(1,0)$ wouldn't be much use. $\nabla f(0,1)$ must be parallel to $\nabla g(0,1)$.

Of course, this says nothing about where the functions f and g are, only what their gradients (slants or tilts) are. In fact, it doesn't matter whether $f(0,1) = g(0,1)$. After all, we are restricting g to $g(x,y) = c$, and c may be arbitrary. If the restraint is $x + y = 4$, we can let g(x,y) = x + y and c = 4, or we can let g(x,y) = x + y - 4 and c = 0, or g(x,y) = x + y + 100 and c = 104. So how "high" g is doesn't matter at all, it's just its gradient (slant or tilt) that we need.

Of course, sometimes you might be required to let c = 0, but there's no difference. The "height" of g doesn't matter.

Comment (3 votes) Upvote Downvote

PTNLemay 2 years ago

more

What does it mean when you have more than one constraint (g and h), so you have 5 equations, 5 unknowns, but the answer you get for the Lagrange multipliers don't come out as a nice scalars? For example lambda ends up being something that depends on x and y.

Reply • Comment (2 votes) Upvote Downvote

Alexander Wu 6 years ago

more

The contour graph for f(x,y) = 2x^2 + √(5y) doesn't seem right. For example, check the point (0,2). It should equal √10, yet in the contour graph, it has two values, √10 and -√10. And f(2,3) should be 8 + √15, about 12, but on the plot it seems it also has a value at about 4.

I think there is a bug in the program that displays this plot. Once you square everything and solve the equation, you get roots that weren't there before. For example, if you square both sides of x = √2, you get x^2 = 2, so x = +-√2. Though I'm not sure if that's what's wrong with the program.

I graphed the function with my calculator and there was only one value for (0,2) and (2,3). But then, f is a function, so the can only be one output for each pair of inputs. The contour plot above doesn't seem right.

Reply • Comment (2 votes) Upvote Downvote

menaker.maxim a year ago

more

Suppose that we have **two constraints** in **three dimensions**. How should we build the **bordered Hessian matrix** of the Lagrangian so that we will be able to determine whether the point we have got is a constrained maximum or a contained minimum?

Reply • Comment (1 vote) Upvote Downvote

festavarian2 2 years ago

more

All the interactive models I see show the two gradient vectors pointing in the SAME DIRECTION at the point of tangency. Couldn't they also point in the exact opposite directions.They would still be scalar multiples.

Reply • Comment (1 vote) Upvote Downvote

wrwrwr 6 years ago

more

In "Using contour maps", what exactly is drawn for the contour of 2x^2 + sqrt(5y), say for k = -2?

Reply • Comment (1 vote) Upvote Downvote

Emily H 6 years ago

more

The graph of 2x^2 + sqrt(5y) = k is drawn. We would normally write it a different form, though. Sqrt(5y) = k -

2x^2. 5y = (k - 2x^2)^2. Y= (k - 2x^2)^2 /5. Try to graph this y in a calculator for a value of k where k is between -4 and 4, and you will have the graph shown in "Using contour maps."

4 comments (1 vote) Upvote Downvote

Show more...

Lagrange multipliers, examples ›