Deep Dive: Applications of Differentiation

Deep Dive

1. Least Squares

We can treat this as an optimization problem, as we are trying to optimize for the best fit line to have the minimum squared-distance from all the points, and we will arrive at:

$$m\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$
$$m\sum_{i=1}^{n} x_i^2 + b\sum_{i-1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

What we know:

The vertical deviation: $d_i = y_i - (mx_i + b)$

Minimize: $\sum_{i=1}^{n} d_i^2 \to \sum_{i=1}^{n} (y_i - mx_i - b)^2$

Work we need to do:

Our objective function is: $f(m, b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2$

To find the extrema, we find partial derivatives with respect to m, b to 0, from the objective function.

$\partial m = 2(y_i - mx_i - b)(-x_i)$

$\partial b = 2(mx_i + b - y_i)$

Now that we have the partial derivatives, we can complete the gradient vector, and set it to 0.

$\nabla f(m, b) = \langle \sum_{i=1}^{n} 2(y_i - mx_i - b)(-x_i), \sum_{i=1}^{n} 2(mx_i + b - y_i) \rangle$

$\nabla f = \langle 2(y_i - mx_i - b)(-x_i), 2(mx_i + b - y_i) \rangle = 0$

From the ∂m:

$f_m = \sum_{i=1}^{n} 2(y_i - mx_i - b)(-x_i) = 0$

We can move the minus sign to -2 from $-x_i$:

$\sum_{i=1}^{n} -2(y_i - mx_i - b)(x_i) = 0$

We take the constant -2 out of the summation:

$-2 \sum_{i=1}^{n} (y_i - mx_i - b)(x_i) = 0$

We divide by -2:

$\sum_{i=1}^{n} (y_i - mx_i - b)(x_i) = \frac{0}{-2}$

$\sum_{i=1}^{n} (y_i - mx_i - b)(x_i) = 0$

We expand and divide the summation accordingly:

$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} mx^2 - \sum_{i=1}^{n} x_i b = 0$

We take the constants out of the summation again:

$\sum_{i=1}^{n} x_i y_i - m \sum_{i=1}^{n} x^2 - b \sum_{i=1}^{n} x_i = 0$

We have finally arrived to our first equation:

$m \sum_{i=1}^{n} x^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$

For ∂b we use the same procedure:

$f_b = \sum_{i=1}^{n} 2(mx_i + b - y_i) = 0$

$2\sum_{i=1}^{n}(mx_i + b - y_i) = 0$

$\sum_{i=1}^{n}(mx_i + b - y_i) = \frac{0}{2}$

$\sum_{i=1}^{n}(mx_i + b - y_i) = 0$

$\sum_{i=1}^{n} mx_i + \sum_{i=1}^{n} b = \sum_{i=1}^{n} y_i$

$m\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$

The last equation looks exactly like the first one that we need to derive, thus we have found both equations needed.

Interpretation:

We have used our mathematical tools to derive the equations that ensure that we have found the minimum squared distance between the best fit line and the points in the graph. That is achieved by treating the function as an optimization problem, which has constraints. In that case, we can set the partial derivatives of m, b to 0, and obtain the equations that we need to solve to get an extreme value. To ensure it is negative, we need to plug back the values that we find in the objective function and evaluate it for the objective function.

---

2. Breaking Lagrange

$f(x, y) = 2x + 3y$

$g(x, y) = \sqrt{x} + \sqrt{y} = 5$

Solving the problem by using Lagrange Multiplier

To form the gradient vector ∇f and ∇g, we need to use partial derivatives with respect to x and y.

$$\frac{\partial f}{\partial x} = \frac{d}{dx}(2x + 3y) = 2 \cdot \frac{d}{dx}x + \frac{d}{dx}(3y) = 2 \cdot 1 + 0 = 2$$
$$\frac{\partial f}{\partial y} = \frac{d}{dy}(2x + 3y) = \frac{d}{dy}(2x) + 3 \cdot \frac{d}{dy}(y) = 0 + 3 \cdot 1 = 3$$

$\nabla f = \langle 2, 3 \rangle$

$\frac{\partial g}{\partial x} = \frac{d}{dx}(\sqrt{x} + \sqrt{y}) = \frac{d}{dx}\sqrt{x} + \frac{d}{dx}\sqrt{y} = \frac{1}{2} \cdot x^{\frac{-1}{2}} + 0 = \frac{1}{2\sqrt{x}}$
$\frac{\partial g}{\partial y} = \frac{d}{dy}(\sqrt{x} + \sqrt{y}) = \frac{d}{dy}\sqrt{x} + \frac{d}{dy}\sqrt{y} = 0 + \frac{1}{2} \cdot y^{\frac{-1}{2}} = \frac{1}{2\sqrt{y}}$
$\nabla g = \langle \frac{1}{2\sqrt{x}}, \frac{1}{2\sqrt{y}} \rangle$

Now, we multiply the gradient vector of the constraint function with lambda, and that should be equal to the gradient vector of the main function:

$\nabla f = \lambda \nabla g$

$\langle 2, 3 \rangle = \lambda \langle \frac{1}{2\sqrt{x}}, \frac{1}{2\sqrt{y}} \rangle$

To find λ, x and y, we need to set up the equations from the above information:

$\lambda \frac{1}{2\sqrt{x}} = 2$

$\lambda \frac{1}{2\sqrt{y}} = 3$

As we have two equations, and three unknowns, we will get a parametric solutions. As we know that none of our variables is a parameter, we should make use of our constraint equation as well to solve the problem. Thus, we add our constraint function to our set of equations:

$\lambda \frac{1}{2\sqrt{x}} = 2$

$\lambda \frac{1}{2\sqrt{y}} = 3$

$\sqrt{x} + \sqrt{y} = 5$

We use Wolfram Alpha to solve this problem:

Input:

$$\left\{\lambda \times \frac{1}{2} \times \frac{1}{\sqrt{x}} = 2, \lambda \times \frac{1}{2} \times \frac{1}{\sqrt{y}} = 3, \sqrt{x} + \sqrt{y} = 5\right\}$$

Result:

$$\left\{\frac{\lambda}{2\sqrt{x}} = 2, \frac{\lambda}{2\sqrt{y}} = 3, \sqrt{x} + \sqrt{y} = 5\right\}$$

Alternate forms:

$$\left\{\lambda = 4\sqrt{x}, \lambda = 6\sqrt{y}, \sqrt{x} + \sqrt{y} = 5\right\}$$

$$\left\{\frac{\lambda}{\sqrt{x}} = 4, \frac{\lambda}{\sqrt{y}} = 6, \sqrt{x} + \sqrt{y} = 5\right\}$$

Solution:

$x = 9, \quad y = 4, \quad \lambda = 12$

Download Page

POWERED BY THE **WOLFRAM LANGUAGE**

Therefore, $x = 9, y = 4$ and $\lambda = 12$.

Now we use these values in the main function:

$f(9,4) = 2 \cdot 9 + 3 \cdot 4 = 30$

We use Lagrange multipliers to find extreme points, therefore $(9,4)$ is one of these points.

Does $f(25,0)$ give a larger value than $f(9,4)$?

$f(9,4) = 30$

$f(25,0) = 2 \cdot 25 + 3 \cdot 0 = 50$

As seen above, $f(25,0)$ gives a higher value than $f(9,4)$.

Graph solution:

The graph below represents 4 graphs:

1. Top-left: The constraint level curve in black, and the level curves from the objective function (colorized).
2. Bottom-Left: The two level curves of interest from f; one where the constraint function is tangent to, and the other representing two intersections with the constraint.
3. Top-right: $f = 50$, the constraint function touches the level curve of f twice.
4. Bottom-right: $f = 30$, the constraint function is tangent to the level curve of f.

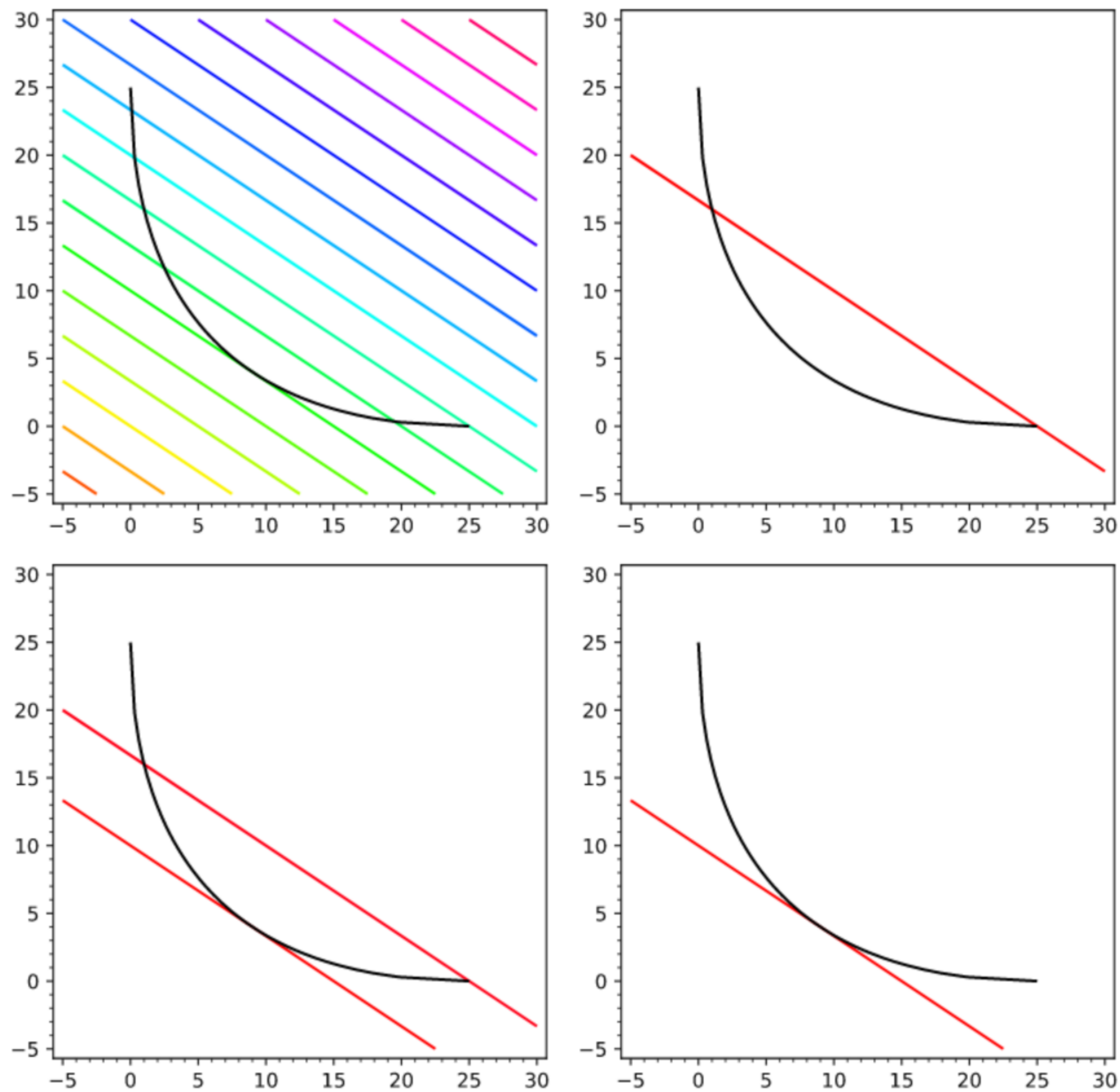

Figure 1. The graph on the top-left represents many level curves for the objective function, while displaying the constraint. The top-right graph represents the level curve of objective function touched by the constraint (f=50). The bottom-left represents both the level curve of objective function touched by the constraint and the level curve to where the constraint is tangent. And the final graph shows only the curve where the constraint is tangent (f=30) (Sage).

From the graphs we can note critical points, which will give us the maximas and the minimums. One of the critical points is found at $f(9,4)=30$, as this is the point of tangency. This is in agreement with the Lagrange Multipliers solution, which ensure that we find values of the functions for which the constraint function will be a tangent to the value. However, there are so many more critical points for which the function is increasing after the $f(9,4)$, one such $f(25,0)$, thus making our Lagrange Multiplier solution not a global maxima.

The failure of Lagrange Multipliers

There is an underlying assumptions once the equations are set, $x \neq 0$ and $y \neq 0$. This happens, since if x or y are 0, then λ cannot be found. This does not mean that there are no extremas at 0, but Lagrange equations are not meant to find them, as that would disable finding λ.

For instance,

$\langle 0, y_1 \rangle = \lambda \langle x_2, y_2 \rangle$

$0 = \lambda x_2$ → either λ or x have to be 0, thus leaving the other variable unsolved and not in a helpful stage for solving for y_2.

From the graph we see that the $f(9,4) = 30$ is actually an absolute minimum, instead of a maximum. That is because the other candidate solutions has one of the points equal to 0, respectively $y = 0$; $f(25,0)$. Lagrange gave us the wrong solutions because of not being able to account for 0.

Another case worth mentioning that Lagrange does not handle well is inequalities, for the same reasons of complications to find λ. Although in our example, λ seems to play the role of a scalar-parameter, in other examples it may contain more information, thus being able to finding λ value is important.

Reflection Questions:

1. #constraints: Although as an individual I was not constrained, I had to account for constraints in my problem-solving process for the Lagrange Multipliers. I have used the constraints towards narrowing the solution space, and I have used a method that uses constraints to find the solutions. That deepens my understanding of constraints as solution hints.

   #plausibility: I have checked the plausibility of two different solutions and explained their meaning to our problem. I found out that it is not plausible for points (9,4) to be a global maxima, after seeing the graph. Therefore, exploring different mediums and methods, I found new, meaningful plausibility checks for critical points.

2. **#scienceoflearning**: I have used several techniques of science of learning, such as summarization, spaced practice and dual codes, to finish the assignment on time and with the desired quality. As summarization is not an optimal technique to be applied, in terms of how much is retained after some time, I will try to reduce the application of that, and focus more on spaced practice, which is far more efficient.
3. Strengths:
    1. Good management of time; if this material was crammed, it would be difficult to digest or to be analyzed.
    2. Understanding of edge cases; I researched more on what Lagrange Multipliers cannot be applied, which resulted in deepening my knowledge about the topic.

- Weakness:
    1. Intuition: I still find it difficult to have a quick intuition or interpretation of mathematical concepts. For instance, I struggled conceptualizing λ, beyond it just being a parameter.
    2. Expression: I often feel motivated to write the solutions without explanations as I struggle to communicate the process clearly.