Lagrange Multipliers and Problem Formulation

Steven J. Miller*

Department of Mathematics and Statistics
Williams College
Williamstown, MA 01267

Abstract

The method of Lagrange Multipliers (and its generalizations) provide answers to numerous important tractable optimization problems in a variety of subjects, ranging from physics to economics to information theory. Below we discuss *five* different formulations for a military problem (which can be re-interpreted as a problem in business). We will show how the answer to this problem depends on our choice of measuring. This illustrates a very powerful and important lesson: depending on your evaluation criteria, you can end up with very different answers.

Contents

*E-mail: `sjm1@williams.edu`

1 Description of the Problem

Consider the following problem from Military Theory. Imagine that the Earth is sphere of radius 1, and that we are trying to determine where to place a military installation in order to minimize the average deployment time to three trouble spots, located at

$$P_1 = (5/13, 12/13, 0), \quad P_2 = (12/13, 5/13, 0), \quad P_3 = (3/13, 4/13, 12/13).$$

Assuming we may build our base anywhere on the Earth, where should we place it?

While stated as a military exercise, the above problem is very important in business. Instead of three trouble spots imagine we have three markets. Now our goal is to locate our factory or distribution center in order to minimize shipping costs.

2 Interpreting the problem

There are two competing influences in building a mathematical model. The first is that we desire a model which captures as many features as possible about the system of interest, while the second is that we wish the model to be mathematically tractable and hence solvable. In general, the more features the model captures, the harder it will be to solve it.

For this problem, we need to decide on how to measure the distance between points on the surface of the Earth. Assume our base is located at (x, y, z) with $x^2 + y^2 + z^2 = 1$. We give five different ways to measure the distance.

1. We use the sum of the distance-squared, where we compute distance by burrowing through the Earth:

$$D_1(x, y, z) = \sum_{i=1}^{3} ||(x, y, z) - P_i||^2.$$

   Expanding, we find

$$\begin{aligned} D_1(x, y, z) \quad = \quad & (x - 5/12)^2 + (y - 12/13)^2 + z^2 \\ & + (x - 12/13)^2 + (y - 5/12)^2 + z^2 \\ & + (x - 3/13)^2 + (y - 4/13)^2 + (z - 12/13)^2. \end{aligned}$$

   There are several advantages of using this measure of distance. The first is that we have a very nice formula (essentially the Pythagorean Theorem) for computing the distance between two points. The second is that since we are squaring the distances before adding, our distance function is a polynomial of degree two in x, y and z. This means ∇D_1 will be a nice function, and the algebra won't be too bad. The disadvantage of this method, of course, is that we cannot burrow through the Earth!

2. We use the sum of the distances, where we compute distance by burrowing through the Earth:

$$D_2(x,y,z) \;=\; \sum_{i=1}^{3} ||(x,y,z)-P_i||.$$

Expanding, we find

$$\begin{aligned} D_1(x,y,z) \;&=\; \sqrt{(x-5/12)^2+(y-12/13)^2+z^2} \\ &\quad + \sqrt{(x-12/13)^2+(y-5/12)^2+z^2} \\ &\quad + \sqrt{(x-3/13)^2+(y-4/13)^2+(z-12/13)^2}. \end{aligned}$$

There are two problems with this method. The first is that the square-roots lead to a function which will have a very messy derivative, and the second of course is that this is burrowing through the Earth.

3. We use the sum of $2(1-\cos\psi_i)$, where ψ_i is the angle between the point P_i and our base at (x,y,z):

$$D_3(x,y,z) \;=\; \sum_{i=1}^{3} 2(1-\cos\psi_i).$$

There are a lot of advantages of this. The first is that the cosine of an angle is easily computed using the dot product. As all vectors are unit vectors, we have

$$(x,y,z)\cdot P_i \;=\; ||(x,y,z)||\;||P_i||\;\cos\psi_i;$$

thus for the first point we have

$$\frac{5x}{13}+\frac{12y}{13} \;=\; \cos\psi_1.$$

The second advantage is that, for ψ_i small, $1-\cos\psi_i \approx \psi_i^2$; this follows from the Taylor series of cosine ($\cos\theta = 1-\theta^2/2!+\theta^4/4!-\cdots$). As we are multiplying by 2, we see that for small ψ_i each summand is about ψ_i^2; this is why we chose to look at $2(1-\cos\psi_i)$. Note that the closer two points are, the smaller their angle and thus the closer the cosine will be to 1; this shows our quantity is a reasonable measure of distance. For a circle of unit radius, the length of an arc equals the angle in radians, and thus this should be a reasonable approach. Explicitly, we have

$$D_3(x,y,z) \;=\; 6-2\left(\frac{5x}{13}+\frac{12y}{13}+\frac{12x}{13}+\frac{5y}{13}+\frac{3x}{13}+\frac{4y}{13}+\frac{12z}{13}\right).$$

Note that D_3 is a *linear* function in the unknown location (x,y,z); thus we expect this will be a very easy function to optimize.

4. We use the sum of the square of the sines of the angles ψ_i (with ψ_i as above). Explicitly,

$$D_4(x,y,z) \;=\; \sum_{i=1}^{3} \sin^2\psi_i.$$

From the Taylor expansion of sine ($\sin\theta = \theta - \theta^3/3! + \cdots$), we see for small ψ_i that $\sin^2\psi_i \approx \psi_i^2$, and thus we expect the answer to be similar to the previous method. Why are we using this, and more importantly, why are we squaring each sine? Unlike the dot product, our formula for sines is a bit more involved, coming from the cross product. We have

$$||(x,y,z) \times P_i|| \;=\; ||(x,y,z)||\;||P_i||\;|\sin\psi_i|;$$

as every vector has length 1, we have

$$||(x,y,z) \times P_i|| \;=\; |\sin\psi_i|.$$

The absolute value function is a horrible nightmare computationally. It is not differentiable, and thus if we use it we cannot apply any of the techniques of calculus. Thus we must remove the absolute value at all costs. We can do this by squaring the above, and obtain

$$||(x,y,z) \times P_i||^2 \;=\; \sin^2\psi_i.$$

While the formula for the square of the sine is a bit more involved, it will be a quadratic in x, y and z and thus its gradient won't be too bad. For example,

$$(x,y,z) \times (5/13, 12/13, 0) \;=\; \left(-\frac{12z}{13}, \frac{5z}{13}, \frac{12x}{13} - \frac{5y}{13}\right),$$

and thus

$$||(x,y,z) \times (5/13, 12/13, 0)||^2 \;=\; \left(\frac{12x}{13} - \frac{5y}{13}\right)^2 + z^2.$$

Adding the contributions from the other two points yields

$$D_4(x,y,z) \;=\; 329x^2 - 264xy + 322y^2 - 72xz - 96yz + 363z^2.$$

Note again how nice the above function looks.

5. We use the sum of the distances of the arcs on the surface of the Earth. We know from geometry that, for a circle of radius 1, the length of the arc equals the measure of the angle in radians. Thus the angle ψ_i between (x,y,z) and P_i satisfies

$$\cos\psi_i \;=\; (x,y,z)\cdot P_i,$$

or

$$\psi_i \;=\; \arccos\left((x,y,z)\cdot P_i\right).$$

Thus the distance is

$$D_5(x,y,z) \;=\; \sum_{i=1}^{3} \arccos\left((x,y,z)\cdot P_i\right).$$

The advantage of this formula is that it is the correct measure of distance; the disadvantage is that it involves the function arccosine. The derivative of $\arccos t$ is $-(1-t^2)^{-1/2}$. Similar to the second approach, this will lead to very messy derivatives and it won't be possible to get a nice, clean closed form expression for the solution. Explicitly, our function is

$$D_5(x,y,z) \;=\arccos\left(\frac{5x}{13}+\frac{12y}{13}\right)+\arccos\left(\frac{12x}{13}+\frac{5y}{13}\right)+\arccos\left(\frac{3x}{13}+\frac{4y}{13}+\frac{12z}{13}\right).$$

We now try and use the method of Lagrange Multipliers to find the optimal site for our base. As our constraint function is

$$g(x,y,z) \;=\; x^2+y^2+z^2 \;=\; 1,$$

we have

$$(\nabla g)(x,y,z) \;=\; 2(x,y,z).$$

Thus we are trying to solve

$$(\nabla D_k)(x,y,z) \;=\; \lambda(\nabla g)(x,y,z), \quad g(x,y,z) \;=\; 1$$

for $k \in \{1,2,3,4,5\}$, or

$$(\nabla D_k)(x,y,z) \;=\; 2\lambda(x,y,z), \quad g(x,y,z) \;=\; 1$$

for $k \in \{1,2,3,4,5\}$.

The natural question to ask is how our choice of measuring distance affects the final answer; do the five different methods lead to our base being in approximately the same place, or are there significant differences in its location?

Remark 2.1. Note that it is possible to generalize this problem by incorporating weights. For example, it may be the case that the three trouble spots are not equally unstable, and thus we are less concerned with one of them. Or perhaps the three points represent markets for our product, and one market is not as important as another and buys significantly less. Similar to our many choices for measuring the distance, we have many ways of including weights. One possibility is to have linear weights, such as

$$D_{2;\text{weighted}}(x,y,z) \;=\; \sum_{i=1}^{3} w_i||(x,y,z)-P_i||;$$

for example, we might take $w_1 = 6$, $w_2 = 3$ and $w_1 = 1$, indicating that the first location is significantly more important than the second which is significantly more important than the third. Such linear weights lead to trivial modifications in these functions, and similar arguments will lead to their solutions.

3 Solving the models

We now describe how to solve the five different models. Our goal is to see how significant these different formulations are. Remember the constraint is

$$g(x,y,z) \;=\; x^2+y^2+z^2 \;=\; 1, \quad (\nabla g)(x,y,z) \;=\; 2(x,y,z).$$

3.1 First Model

Recall

$$D_1(x,y,z) \;=\; \sum_{i=1}^{3} ||(x,y,z) - P_i||^2.$$

Expanding, we find

$$\begin{aligned} D_1(x,y,z) \;&=\; (x-5/12)^2 + (y-12/13)^2 + z^2 \\ &\quad + (x-12/13)^2 + (y-5/12)^2 + z^2 \\ &\quad + (x-3/13)^2 + (y-4/13)^2 + (z-12/13)^2. \end{aligned}$$

We thus find that

$$(\nabla D_1)(x,y,z) \;=\; (-40+78x, -42+78y, -24+78z).$$

Setting $\nabla D_1 = \lambda \nabla g$ yields

$$(-40+78x, -42+78y, -24+78z) \;=\; 2\lambda(x,y,z).$$

As happens for many problems like this, it helps to break the analysis into two cases: $\lambda = 0$ and $\lambda \neq 0$.

If $\lambda = 0$ we find

$$(-40+78x, -42+78y, -24+78z) \;=\; (0,0,0),$$

which yields

$$x \;=\; 40/78, \quad y \;=\; 42/78, \quad z \;=\; 24/78;$$

however, this point must also satisfy our constraint. Substituting we find

$$(40/78)^2 + (42/78)^2 + (24/78)^2 \;=\; 985/1521 \;\neq\; 1;$$

as this point does not satisfy the constraint, it is neither a maximum nor a minimum. Remember that the Method of Lagrange Multipliers finds *candidates* for maxima and minima; just because a point is a candidate does not ensure that it must be a local max or a local min.

We now turn to the other case, $\lambda \neq 0$. We thus have the following system of equations to solve:

$$\begin{aligned} -40+78x \;&=\; 2\lambda x \\ -42+78y \;&=\; 2\lambda y \\ -24+78z \;&=\; 2\lambda z \\ x^2+y^2+z^2 \;&=\; 1. \end{aligned}$$

As $\lambda \neq 0$, we may take ratios of the various equations. The advantage of doing this is that the λ's disappear. It's worth thinking about the seeming absurdity of our approach: we introduce λ to help solve the problem, and then proceed to eliminate it without ever finding its value! It's a placeholder whose existence helps us find the max/min. (Note: there are many problems where the value of λ is important, as it has a physical interpretation.)

Taking ratios yields

$$\frac{-40+78x}{-42+78y} \;=\; \frac{2\lambda x}{2\lambda y} \;=\; \frac{x}{y}.$$

When we cross multiply, we are quite fortunate:

$$-40y+78xy \;=\; -42x+78xy.$$

We have the same number of xy on both sides, which cancels, and we find

$$y \;=\; \frac{42x}{40}.$$

Taking the ratio of the first and the third yields

$$\frac{-40+78x}{-24+78z} \;=\; \frac{2\lambda x}{2\lambda z} \;=\; \frac{x}{z}.$$

Cross multiplying yields

$$-40z+78xz \;=\; -24x+78xz.$$

Canceling the $78xz$ from both sides yields

$$z \;=\; \frac{24x}{40}.$$

We now substitute into our constraint; as we can write y and z in terms of x, we will have one equation in one unknown:

$$x^2+y^2+z^2 \;=\; 1$$

becomes

$$x^2+\left(\frac{42}{40}\right)^2 x^2+\left(\frac{24}{40}\right)^2 x^2 \;=\; 1,$$

or

$$\frac{197}{80}x^2 \;=\; 1,$$

which implies

$$x \;=\; \pm\sqrt{\frac{80}{197}} \;=\; \pm\sqrt{\frac{400}{985}} \;\approx\; 0.637253,$$

where we have written our answer with the following denominator to facilitate comparisons with other methods.

Now that we have x it is easy to find y and z. While there are two different choices for x, a moments thought shows that the minimum distance has to be when we take the positive square root. The reason is our three trouble points are all in the first octant of the Earth (they all have non-negative coordinates), and thus the minimum distance will be from the point with three non-negative coordinates and not the point with three non-positive coordinates.

We thus find the optimal location, using the above method to measure distance, is

$$(x, y, z) \;=\; \left(\sqrt{\frac{400}{985}}, \sqrt{\frac{420}{985}}, \sqrt{\frac{240}{985}}\right) \;\approx\; (0.637253, 0.669116, 0.382352).$$

3.2 Second Model

We encourage the reader to find the gradient of $D_2(x, y, z)$. The first two terms of $\partial D_2/\partial x$ are

$$\frac{13x - 3}{13\sqrt{(x-3/13)^2 + (y-4/13)^2 + (z-12/13)^2}} + \frac{13x + 5}{13\sqrt{(x-5/13)^2 + (y-4/13)^2 + (z-12/13)^2}} + \cdots .$$

Using numerical approximation techniques, Mathematica computes that D_2 is minimized, subject to the constraint, when

$$(x, y, z) \;=\; (0.658882, 0.723174, 0.207108).$$

3.3 Third Model

This is perhaps the easiest of all cases, as we have

$$D_3(x, y, z) \;=\; 78 - 40x - 42y - 24z.$$

This leads to

$$(\nabla D_3)(x, y, z) \;=\; (-40x, -42y, -24z).$$

Setting $\nabla D_3 = \lambda \nabla g$ yields

$$(-40x, -42y, -24z) \;=\; 2\lambda(x, y, z) \quad \text{and} \quad x^2 + y^2 + z^2 \;=\; 1.$$

Again, in problems like this it is often convenient to break the analysis into two cases, when $\lambda = 0$ and when $\lambda \neq 0$. If $\lambda = 0$ then we quickly see $x = y = z = 0$, which clearly does not satisfy the constraint of being on the unit sphere. Thus we may assume $\lambda \neq 0$. We have

$$\begin{aligned} -40x &= 2\lambda x \\ -42y &= 2\lambda y \\ -24z &= 2\lambda z \\ x^2 + y^2 + z^2 &= 1. \end{aligned}$$

Taking the ratio of the first and the second equation gives

$$\frac{-40x}{-42y} \ = \ \frac{2\lambda x}{2\lambda y},$$

or

$$\frac{40}{42} \ = \ \frac{x}{y},$$

which yields

$$y \ = \ \frac{42x}{40}.$$

If we take the ratio of the first and the third we similarly find

$$\frac{-40x}{-24z} \ = \ \frac{2\lambda x}{2\lambda z},$$

or

$$z \ = \ \frac{24x}{40}.$$

Note that these are *exactly* the relation between x and y that we found in Model 1, and thus the solution will also be

$$(x, y, z) \ = \ \left(\sqrt{\frac{400}{985}}, \sqrt{\frac{420}{985}}, \sqrt{\frac{240}{985}}\right) \ \approx \ (0.637253, 0.669116, 0.382352).$$

As Models1 and 3 give the same location, it is worth thinking about why this is. We can use the law of cosines to determine the distance (burrowing through the Earth) from (x, y, z) to P_i. Letting ψ_i denote the angle between these two vectors and d_i the distance between them, and noting that the two vectors have unit length, we have

$$d_i^2 \ = \ ||(x, y, z)||^2 + ||P_i||^2 - 2||(x, y, z)|| \, ||P_i|| \cos\psi_i \ = \ 2 - 2\cos\psi_i$$

In other words,

$$d_i^2 \ = \ 2(1 - \cos\psi_i),$$

and

$$D_1(x, y, z) \ = \ D_3(x, y, z).$$

While the algebra after computing the gradients in Model 1 and Model 3 is the same, I think the algebra is significantly easier in Model 3 in the beginning.

3.4 Model 4

Recall

$$D_4(x, y, z) \; = \; 329x^2 - 264xy + 322y^2 - 72xz - 96yz + 363z^2,$$

which leads to the gradient

$$(\nabla D_4)(x, y, z) \; = \; (658x - 264y - 72z, -264x + 644y - 96z, -72x - 96y + 726z).$$

While we again have four unknowns in four equations, the algebra becomes a bit more tedious and involved here. If you know linear algebra, it is not that bad to find the solutions. As linear algebra is not a pre-requisite for the course, we will simply state what you would find

$$(x, y, z) \; \approx \; (0.659301, 0.688772, 0.301521)$$

(the closed form solution is a pain to write down).

3.5 Model 5

Due to the presence of the arccosine function, $D_5(x, y, z)$ has a very messy derivative, and we must resort to numerical approximations. This show that the optimal location is approximately

$$(x, y, z) \; \approx \; (0.659441`, 0.713781, 0.235913).$$

4 Comparing the Models' Solutions

We compare the five models:

	Model 1	Model 2	Model 3	Model 4	Model 5
x	.637	.659	.637	.659	.659
y	.669	.723	.669	.689	.714
z	.382	.207	.382	.302	.236

Note five different models returned four different answers! Two of the five models (numbers 2 and 5) had to be solved numerically; the others could be solved using elementary techniques. The fifth model is the best way to measure distance, but not the easiest to work with. Of the models where we can solve everything in closed form, it's clear that Model 4 is closer to the 'true' solution than Model 1 (which is the same as Model 3).

Again, the very important takeaway for this problem is that the answer to a problem can depend on how you choose to ask the question. Different notions of distance lead to different locations.

5 Connection to the rest of the course

This material is a natural stepping stone to the Method of Least Squares. The purpose of the Method of Least Squares is to find the best values of parameters in a theory given experimental observations. The main idea is that we have an error function which depends on the unknown parameters, and we must minimize it, with the minimum occurring when the gradient vanishes. We thus need to decide on what is the ‘best’ way to measure errors. We cannot do just the amount we are off, as that is a signed quantity. We could use the absolute value of the difference between theory and experiment, but the absolute value function is not differentiable. We thus use the square of the difference. This turns out to be mathematically tractable, and leads to very nice closed form expressions.