# The Method of Lagrange Multipliers

Zachary Yedidia

## Introduction

Suppose we would like to minimize (or maximize) a function of $n$ variables $x = (x_1, \dots x_n)$ subject to $p$ constraints of the form

$$\begin{aligned} &\text{minimize} \quad && f(x) \\ &\text{subject to} && \\ & && g_1(x) = c_1 \\ & && g_2(x) = c_2 \\ & && \dots \\ & && g_p(x) = c_p \end{aligned}$$

where $f, g_1, \dots, g_p$ are continuous and have continuous second derivatives. When these functions are linear, the problem is called a *linear program* and is solvable using algorithms such as the simplex algorithm (most commonly used) or interior point methods (for worst-case polynomial time). We wish to examine the case where these functions are nonlinear.

For example, consider the following problem where $p = 1$ and $f$ is quadratic

$$\begin{aligned} &\text{minimize} \quad && \sum_{i=1}^{n} x_i^2 \\ &\text{subject to} && \\ & && \sum_{i=1}^{n} x_i = 1. \end{aligned}$$

A first approach might be to differentiate $f$ with respect to $x_i$ and set the derivative to 0:

$$\frac{\partial f}{\partial x_i} = 0, \ 1 \le i \le n.$$

This gives the optimum of the function but ignores the constraint, and leads to $x_i = 0$ which is incorrect (the constraint is violated). Using Lagrange's method, we can solve the problem by introducing $p$ new variables called *Lagrange multipliers* and solve a larger system of equations.

## Lagrange's method

In order to solve the optimization problem, we define a Lagrangian

$$\mathcal{L} = f(x) + \sum_{i=1}^{p} \lambda_i(g_i(x) - c_i)$$

where $\lambda_1, \ldots, \lambda_p$ are new variables that we introduce into the system called Lagrange multipliers. If $\mathcal{L}$ is concave/convex, we can then find the solution $x$ by solving the system of equations

$$\frac{\partial \mathcal{L}}{\partial x_i} = 0, \ 1 \leq i \leq n$$
$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0, \ 1 \leq i \leq p.$$

This system has $n + p$ equations and $n + p$ unknowns. Fortunately this system is often not too complicated to solve.

## Single-constraint example

Let us apply this to our example from before. We define our Lagrangian

$$\mathcal{L} = \sum_{i=1}^{n} x_i^2 + \lambda(\sum_{i=1}^{n} x_i - 1)$$

and then calculate its derivatives and set them to 0

$$\frac{\partial \mathcal{L}}{\partial x_i} = 2x_i + \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{n} x_i - 1 = 0.$$

Thus from the first equation, we find $x_i = -\frac{\lambda}{2}$. If we plug this into the second equation we get

$$\sum_{i=1}^{n} x_i - 1 = 0$$
$$\sum_{i=1}^{n} -\frac{\lambda}{2} = 1$$
$$-\frac{n\lambda}{2} = 1$$
$$\lambda = -\frac{2}{n}.$$

Now we plug this value for $\lambda$ back into our equation for $x_i$, and it follows that $x_i = 1/n$ for $1 \leq i \leq n$.

## Multiple-constraint example

Solving optimization problems with multiple constraints is the same method as with single constraints except we get a larger system of equations to solve.

Suppose we wish to solve the following minimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{5} x_i^2 \\ \text{subject to} \quad & \\ & x_1 + 2x_2 + x_3 = 1 \\ & x_3 - 2x_4 + x_5 = 6. \end{aligned}$$

As with before, we set up our Lagrangian and find the derivatives

$$\mathcal{L} = \sum_{i=1}^{5} x_i^2 + \lambda_1(x_1 + 2x_2 + x_3 - 1) + \lambda_2(x_3 - 2x_4 + x_5 - 6)$$

$$\frac{\partial}{\partial x_i}\left(\sum_{i=1}^{5} x_i^2 + \lambda_1(x_1 + 2x_2 + x_3 - 1) + \lambda_2(x_3 - 2x_4 + x_5 - 6)\right) = 0$$

This tells us that $2x_1 + \lambda_1 = 0$, $2x_2 + 2\lambda_1 = 0$, $2x_3 + \lambda_1 + \lambda_2 = 0$, $2x_4 - 2\lambda_2 = 0$, and $2x_5 + \lambda_2 = 0$. We also know

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_1} &= x_1 + 2x_2 + x_3 - 1 = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= x_3 - 2x_4 + x_5 - 6 = 0. \end{aligned}$$

If we take the first three equations and plug them into the first constraint we get $2 + 6\lambda_1 + \lambda_2 = 0$. Taking the last three equations with the second constraint gives $12 + \lambda_1 + 6\lambda_2 = 0$. If we solve these two equations for $\lambda_1$ and $\lambda_2$ we get $\lambda_1 = 0, \lambda_2 = -2$. This implies $x_1 = x_2 = 0$, $x_3 = x_5 = 1$, and $x_4 = -2$.

Notice that $\frac{\partial \mathcal{L}}{\partial \lambda_i}$ will always just give back the constraints, so instead we can just plug into the constraints directly.

### General strategy

The general strategy for solving an optimization problem with Lagrange's method is outlined below:

- Write the Lagrangian $\mathcal{L} = f(x) + \sum_{i=1}^{p} \lambda_i(g_i(x) - c_i)$.
- Find $\frac{\partial \mathcal{L}}{\partial x_i}$
- Solve these equations for $x_i$ in terms of $\lambda_1, \ldots, \lambda_p$.
- Plug these values for $x_i$ into the $p$ constraints
- Solve the system of $p$ variables and $p$ unknowns for $\lambda_i, 1 \leq i \leq p$.
- Plug values for $\lambda_i$ back into the equations for $x_i$ that were written in terms of $\lambda_i$.

## Maximizing entropy

### Discrete example

Consider a random variable $X$ with $k$ discrete possible values $x_i, 1 \leq i \leq k$ and where the probability that the event $x_i$ occurs is $P(X = x_i) = p_i$. Recall that the entropy $H$ of such a random variable is defined as

$$H(X) = \langle -\ln p_i \rangle = -\sum_{i=1}^{k} p_i \ln p_i$$

Using the method of Lagrange multipliers we can find the probability distribution $p_i$ that maximizes the entropy given some constraints.

Consider the following problem: given a half-bounded discrete random variable whose state space consists of the non-negative integers and with mean $\langle X \rangle = \mu$, find the probability distribution that maximizes the entropy of $X$.

We first formulate the problem as an optimization problem where we maximize the entropy while demanding that the probabilities sum to 1 and the mean is $\mu$:

$$\begin{aligned} \text{maximize} \quad & H(X) \\ \text{subject to} \quad & \\ & \sum_{i=0}^{\infty} p_i = 1 \\ & \sum_{i=0}^{\infty} i p_i = \mu. \end{aligned}$$

Now we can write our Lagrangian:

$$\mathcal{L} = -\sum_{0}^{\infty} p_i \ln p_i + \lambda_1 \left( \sum_{i=0}^{\infty} p_i - 1 \right) + \lambda_2 \left( \sum_{i=0}^{\infty} i p_i - \mu \right)$$

We differentiate with respect to $p_i$ and set the expression to 0, which gives

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial p_i} &= -\ln p_i - 1 + \lambda_1 + \lambda_2 i = 0 \\ \ln p_i &= -1 + \lambda_1 + \lambda_2 i \\ p_i &= e^{-1+\lambda_1+\lambda_2 i}. \end{aligned}$$

Let us now plug this value back into the first constraint

$$\begin{aligned} \sum_{i=0}^{\infty} e^{-1+\lambda_1+\lambda_2 i} &= 1 \\ -\frac{e^{\lambda_1 - 1}}{e^{\lambda_2} - 1} &= 1 \\ e^{\lambda_2} &= -e^{\lambda_1 - 1} + 1. \end{aligned}$$

Now plug our value for $p_i$ into the second constraint

$$\begin{aligned} \sum_{i=0}^{\infty} i e^{-1+\lambda_1+\lambda_2 i} &= \mu \\ \frac{e^{\lambda_1+\lambda_2-1}}{\left(e^{\lambda_2} - 1\right)^2} &= \mu \\ e^{\lambda_1} &= \mu e^{1-\lambda_2} \left(e^{\lambda_2} - 1\right)^2. \end{aligned}$$

Plug in our value for $e^{\lambda_2}$ and solve and we get

$$e^{\lambda_1} = \frac{e}{\mu + 1}.$$

We can now also solve for $\lambda_2$ by plugging this in

$$e^{\lambda_2} = -\frac{e}{\mu + 1} e^{-1} + 1 = -\frac{1}{\mu + 1} + 1.$$

We now take these values for $\lambda_1$ and $\lambda_2$ and plug into our expression for $p_i$

$$\begin{aligned}
p_i &= e^{-1+\lambda_1+\lambda_2 i} \\
p_i &= e^{-1} e^{\lambda_1} (e^{\lambda_2})^i \\
p_i &= e^{-1} \left( \frac{e}{\mu + 1} \right) \left( \frac{1}{\mu + 1} + 1 \right)^i \\
p_i &= \left( \frac{1}{\mu + 1} \right) \left( -\frac{1}{\mu + 1} + 1 \right)^i \\
p_i &= \left( \frac{1}{\mu + 1} \right) \left( \frac{\mu}{\mu + 1} \right)^i \\
p_i &= \mu^i \left( \mu + 1 \right)^{-1-i}.
\end{aligned}$$

This is the final expression for the probability distribution with maximized entropy.

## Continuous example

Recall that the differential entropy $H$ of a continuous random variable $X$ whose support is $\mathcal{X}$ with probability density function $p(x)$ is defined as

$$H(X) = -\int_{\mathcal{X}} p(x) \ln p(x) dx.$$

Using the method of Lagrange multipliers we can find the probability distribution that maximizes entropy given certain constraints.

Consider the following problem: given a half-bounded continuous random variable $X$ defined for $x \in [0, \infty)$ with mean $\mu$, find the probability distribution that maximizes the entropy $X$.

We first formulate the problem as an optimization problem where we maximize the entropy while demanding that the density function sums to 1 and the mean is $\mu$:

$$\begin{aligned}
\text{maximize} \quad & H(X) \\
\text{subject to} \quad & \\
& \int_0^\infty p(x) = 1 \\
& \int_0^\infty x p(x) = \mu.
\end{aligned}$$

Now we can write our Lagrangian:

$$\mathcal{L}[p(x)] = -\int_0^\infty p(x)\ln p(x)dx + \lambda_1 \left(\int_0^\infty p(x)dx - 1\right) + \lambda_2 \left(\int_0^\infty xp(x)dx - \mu\right).$$

Variational calculus tells us that when we differentiate with respect to a function, we can remove the integrals. This makes sense intuitively because the integral is a sum and all the terms different from $p(x)$ vanish (see the discrete example above). Differentiating gives

$$\frac{\partial \mathcal{L}}{\partial p} = -\ln p(x) - 1 + \lambda_1 + \lambda_2 x = 0.$$

This gives

$$p(x) = e^{-1+\lambda_1+\lambda_2 x}.$$

When we apply the two constraints, we get an exponential function

$$p(x) = \frac{1}{\mu} e^{x/\mu}.$$

## Bibliographic notes

These notes are from MCB131 taught by Professor Haim Sompolinsky. The examples are from section, as well as the supplementary material on Lagrange multipliers by S. Sawyer.