Contact: `gmelvin@middlebury.edu`

May 4 Lecture

Textbook Reference:
- *Vector Calculus*, Colley, 4th Edition: §4.3, 4.4

Lagrange Multipliers; Linear Regression

Keywords: the method of Lagrange multipliers, linear regression, polynomial regression

---

Let $f(x,y)$, $g(x,y)$ be differentiable functions. We are interested in the **contstrained optimisation** problem

$$\begin{aligned} \text{maximise/minimise} \quad & f(x,y) \\ \text{subject to} \quad & g(x,y) = 0 \end{aligned}$$

Remark: An equation $h(x,y) = c$, c constant, is called a **constraint**. Note that any constraint $h(x,y) = c$ can be rearranged to a constraint of the form $g(x,y) = 0$ by letting $g(x,y) = h(x,y) - c$.

In the last lecture we saw that the solutions to this problem - the **constrained extrema** - came in two flavours:

(I) the points (x,y) satisfying $\nabla f(x,y) = 0$ and $g(x,y) = 0$;

(II) the points (x,y) satisfying $\nabla f(x,y) = \lambda \nabla g(x,y)$ and $g(x,y) = 0$, for some nonzero λ, called a **Lagrange multiplier**.

> Note that type (I) points can be considered to be type (II) points for the case $\lambda = 0$.

We generalise to the setting of several variables:

> ### Method of Lagrange multipliers: single constraint
>
> Let $f(\underline{x})$, $g(\underline{x})$ be differentiable functions of n variables. If $\underline{x}$ is a solution to the constrained optimisation problem
>
> $$\begin{aligned} \text{maximise/minimise} \quad & f(\underline{x}) \\ \text{subject to} \quad & g(\underline{x}) = 0 \end{aligned}$$
>
> then there exists some λ such that $(\underline{x}, \lambda)$ is a solution to the equation
>
> $$\nabla f(\underline{x}) = \lambda \nabla g(\underline{x})$$

Remark: The method of Lagrange multipliers can be extended to the case of multiple constraints $g_1(\underline{x}) = \ldots = g_k(\underline{x}) = 0$. In this case there are two flavours of constrained extrema:

(I) the points $\underline{x}$ satisfying $\nabla f(\underline{x}) = 0$ and $g_1(\underline{x}) = \ldots = g_k(\underline{x}) = 0$;

(II) the points $\underline{x}$ satisfying $\nabla f(\underline{x}) = \sum_{i=1}^{k} \lambda_i \nabla g_i(\underline{x})$ and $g_1(\underline{x}) = \ldots = g_k(\underline{x}) = 0$, for some $\lambda_1, \ldots, \lambda_k$ (not all equal to zero).

The gradient condition states that $\nabla f(\underline{x})$ is orthogonal to the tangent space of the space defined by $g_1 = \ldots = g_k = 0$. For details see p.284 of the textbook.

Example: Model the surface of the Earth by the unit sphere $x^2 + y^2 + z^2 = 1$. A satellite is orbiting the earth at a fixed height - in our model the satellite's orbit is constrained to lie in the sphere $x^2 + y^2 + z^2 = 9$. Assume we are standing at $(1, 0, 0)$ on the surface of the Earth. Let's use Lagrange multipliers to confirm the obvious(?) geometric fact: the satellite is closest to our position when the satellite is at $(3, 0, 0)$.

We model this problem as a constrained optimisation problem:

$$\begin{array}{rl} \text{minimise} & d(x, y, z) = (x-1)^2 + y^2 + z^2 \\ \text{subject to} & g(x, y, z) = x^2 + y^2 + z^2 - 9 = 0 \end{array}$$

Solution:

Applications of Extrema

Linear regression A set S of k points in the plane

$$S = \{(x_1, y_1), \ldots, (x_k, y_k)\}$$

can be intepreted as a data set relating two quantities x and y. For example, x could represent SAT scores and y could represent college grades.

We want to understand what the general linear correlation is between the quantities x and y i.e. we want to find the **line of best fit** $y = mx + b$. Mathematically, we want to solve the optimisation problem:

$$\text{minimise} \qquad D(m, b) = (y_1 - (mx_1 + b))^2 + \ldots + (y_k - (mx_k + b))^2$$

Diagram:

We need to find the extrema of the function D. We compute ∇D:

$$\frac{\partial D}{\partial m} = \underline{\hspace{8cm}}$$

$$\frac{\partial D}{\partial b} = \underline{\hspace{8cm}}$$

Setting both partial derivatives equal to zero gives the equations

$$(\sum x_i^2)m + (\sum x_i)b = \sum x_i y_i$$

$$(\sum x_i)m + nb = \sum y_i$$

This is a system of linear equations in the two variables m and b. We solve to obtain the single solution:

This approach can be used to solve more general **polynomial regression**. For example, we could try to determine a **parabola of best fit** $y = ax^2 + bx + c$. Then, we aim to minimise

$$D(a, b, c) = \underline{\hspace{8cm}}$$