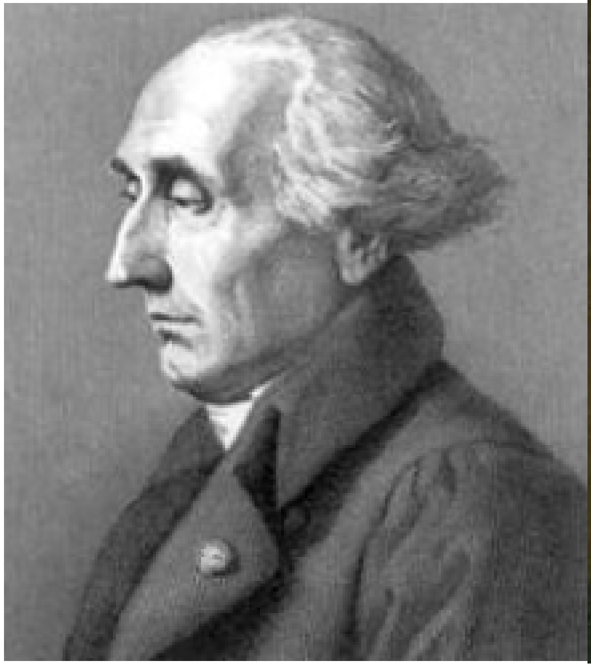
Lagrange, Laplace, Legendre and the French Revolution

The French Revolution

- The 18th century is often viewed as the century of revolutions. In America, 1776 marks the end of the American revolution. In France, 1789 marks the end of the French revolution.
- Both of these are in the context of the global Industrial revolution where scientific methods were being applied to increase productivity.
- In this period, three mathematicians stand out as leaders in their field: Lagrange, Laplace and Legendre.

Joseph Louis Lagrange

- Joseph Louis Lagrange (1736-1813) was actually Italian by birth and was the youngest of 11 children and the only one to have survived beyond infancy.
- The universities in France at that time were not as they are today. Thus, many mathematicians were educated privately and not in a formal sense at any university. Nor did they find jobs in universities but rather were patronized by royalty.
- Lagrange was instrumental in the French adoption of the decimal system in their system of weights and measures.
- He also developed analytic geometry and used it to develop a new branch of mathematics called the calculus of variations.

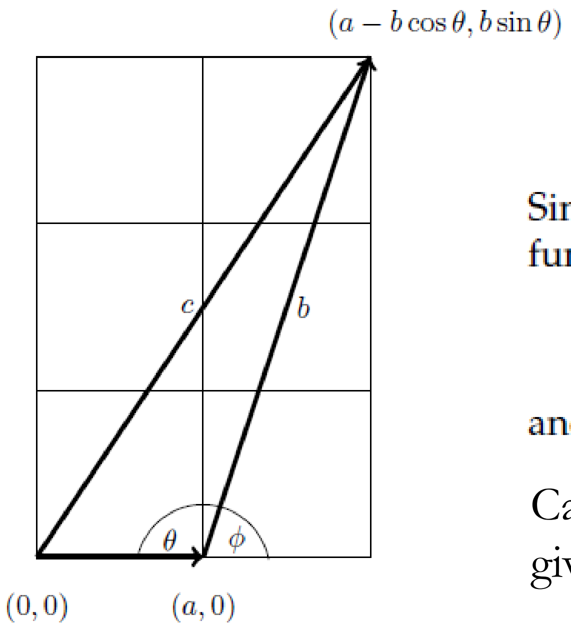
Analytic geometry and the cosine law

- Most students are familiar with the dot product of two vectors. However, they may not be familiar with how the concept arose from studying the cosine law using analytic geometry.

$$\mathbf{a} = (a_1, a_2, ..., a_n), \quad \mathbf{b} = (b_1, b_2, ..., b_n), \qquad \mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

Let us see how the cosine law motivates this definition of the dot product.

$c^2 = a^2 + b^2 - 2ab\cos\theta.$



The co-ordinates of a+b are easily calculated:

$$(a + b\cos\phi, b\sin\phi).$$

Since $\theta + \phi = \pi$, we find either by looking at the graph of the cosine and sine functions, or by using the addition formulas, that

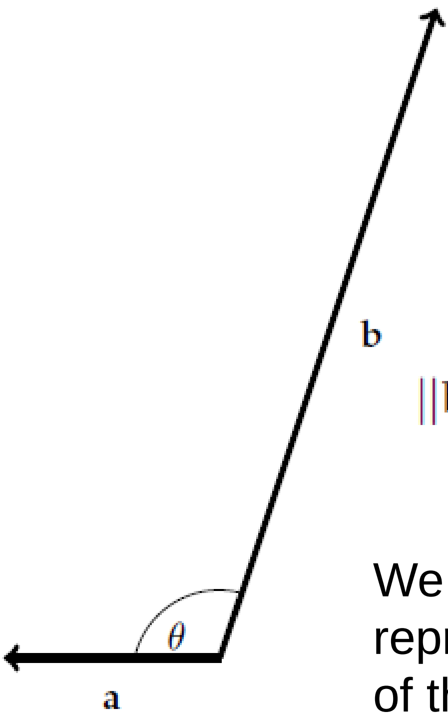
$$a + b\cos\phi = a - b\cos\theta, \quad b\sin\phi = b\sin\theta,$$

and our claim is now evident.

Calculating the length of the vector from the co-ordinates now gives us the cosine law.

The dot product and the cosine law

- Now we can relate the cosine law to the dot product. Consider the two vectors a and b as shown in the figure with an angle θ between them. The cosine law gives:



$$||\mathbf{b}-\mathbf{a}||^2 = ||\mathbf{b}||^2 + ||\mathbf{a}||^2 - 2||\mathbf{a}||\,||\mathbf{b}||\cos\theta$$

On the other hand, if $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$, then $\mathbf{b}-\mathbf{a} = (b_1 - a_1, b_2 - a_2)$ so that by the Pythagorean theorem, we have

$$||\mathbf{b}-\mathbf{a}||^2 = (b_1-a_1)^2 + (b_2-a_2)^2 = b_1^2 + b_2^2 + a_1^2 + a_2^2 - 2(a_1b_1 + a_2b_2).$$

The right hand side is: $||\mathbf{b}||^2 + ||\mathbf{a}||^2 - 2(\mathbf{a}\cdot\mathbf{b}).$

We therefore see, at least in two dimensions, a visual representation of the dot product in terms of the co-ordinates of the vectors using the cosine law.

Observe also that the dot product of two vectors is zero if and only if they are orthogonal to each other.

The cross product

- The cross product also affords a geometric meaning and students usually encounter it in a basic physics course.

It is convenient to recall here the **cross product** of two vectors in $\mathbb{R}^3$. Formally, given two vectors

$$\mathrm{a} = (a_1, a_2, a_3), \quad \mathrm{b} = (b_1, b_2, b_3)$$

in $\mathbb{R}^3$, the cross product is by definition

$$\mathrm{a} \times \mathrm{b} := (a_2 b_3 - a_3 b_2,\ a_3 b_1 - a_1 b_3,\ a_1 b_2 - a_2 b_1).$$

The formal definition lacks any hint of its importance and meaning. In physics, the concept arises to describe **torque**. If a represents the displacement of a particle from a fixed point to a movable point, and b is the force applied at the movable point, then the cross product $\mathrm{a} \times \mathrm{b}$ is the torque exerted by the force about the fixed point. The above opaque definition is better remembered using determinants. One writes symbolically,

$$\mathrm{a} \times \mathrm{b} = \begin{vmatrix} \mathrm{i} & \mathrm{j} & \mathrm{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

where $\mathrm{i}, \mathrm{j}, \mathrm{k}$ are the unit vectors $(1,0,0), (0,1,0)$ and $(0,0,1)$ respectively.

The area of a parallelogram in terms of the cross product

- We can formally expand the determinant:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Expanding the determinant using the first row leads to the earlier definition so that the "determinant expression" serves as a useful mnemonic for the cross product. A straightforward and tedious computation shows that

$$|\mathbf{a} \times \mathbf{b}|^2 = (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2$$

We recognize the last term as the square of a dot product

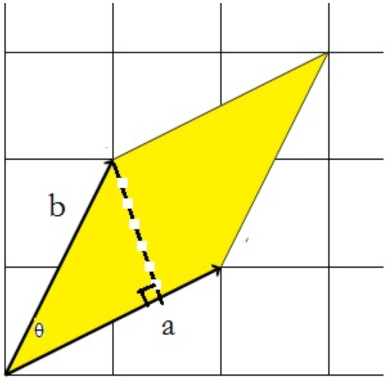
so that

$$|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2|\mathbf{b}|^2 - |\mathbf{a}|^2|\mathbf{b}|^2 \cos^2\theta.$$

Consequently,

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\theta,$$

where θ is the angle between the two vectors a and b. This formula implies a geometric interpretation of the magnitude of the cross product vector. It is the area of the parallelogram spanned by the two vectors a and b. The direction of the cross product is given by the familiar "right hand rule," where if you align the fingers of your right hand along the vector a and bend your fingers around in the direction of rotation from a to b, your thumb will point in the direction of a × b. We also note that the cross product of a and b is zero if and only if they are parallel.



Area of parallelogram = $|\mathbf{a}|\,|\mathbf{b}|\sin\theta$

The area of a parallelogram as a determinant

- Lagrange discovered that the area of a parallelogram can also be written as a determinant. Given two vectors v=(a,b) and w=(c,d) in R^2 we have from our understanding of the cross product that the area of the parallelogram determined by v and w is:

$$\sqrt{|\mathbf{v}|^2|\mathbf{w}|^2 - (\mathbf{v} \cdot \mathbf{w})^2}.$$

Computing this directly using vector co-ordinates leads to the square root of

$$(a^2 + c^2)(b^2 + d^2) - (ab + cd)^2$$

which is easily seen to be

$$(ad - bc).$$

This is the determinant of the 2×2 matrix whose columns are $\mathbf{v}$ and $\mathbf{w}$.

Areas and volumes as determinants

- Lagrange found general formulas for areas and volumes in terms of determinants. For instance, the area of a triangle with co-ordinates (x_1, y_1), (x_2, y_2), (x_3, y_3) is:

$$\frac{1}{2!}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

Similarly, the volume of a tetrahedron, with obvious notation is:

$$\frac{1}{3!}\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}$$

The gradient of a function of several variables

- Lagrange developed multivariable calculus in several directions. In dealing with functions of several variables, the concept of derivative is delicate and there are several ways of viewing it.
- The notion of gradient should be familiar. Given a continuous function $f : \mathbb{R}^n \to \mathbb{R}$, we define the gradient of f, denoted ∇f as a function from $\mathbb{R}^n$ to $\mathbb{R}^n$ given by

$$\nabla f = (D_1 f, \ldots, D_n f).$$

Here $D_1, \ldots, D_n$ denote partial differentiation operations with respect to the variables $x_1, \ldots, x_n$.

We also have the concept of a directional derivative: suppose u is a vector. Let us look at the case $f : \mathbb{R}^n \to \mathbb{R}$. Given a vector $\mathbf{u}$ we can define the directional derivative at $\mathbf{x}_0$, sometimes denoted $f_{\mathbf{u}}(\mathbf{x}_0)$ by the limit,

$$\lim_{h \to 0} \frac{f(\mathbf{x}_0 + h\mathbf{u}) - f(\mathbf{x}_0)}{h},$$

provided the limit exists.

The Lagrange multiplier method

The **Lagrange multiplier method** gives conditions for finding the maxima or minima of a scalar field subject to a side condition. Suppose we want to find the maximum or minimum of a function $f : \mathbb{R}^n \to \mathbb{R}$ subject to the side condition $g(\mathbf{x}) = 0$ for some differentiable $g : \mathbb{R}^m \to \mathbb{R}$. Let C be any curve given by $\mathbf{r} : [0, 1] \to \mathbb{R}^n$ lying on the hypersurface defined by $g(\mathbf{x}) = 0$. Thus, if $\mathbf{r}(t) = (x_1(t), \ldots, x_n(t))$, then $g(r(t)) = 0$. Now if f has an extremum at $\mathbf{x}_0$ (say), and C passes through $\mathbf{x}_0$, then setting $h(t) := f(\mathbf{r}(t))$, we see that h also has an extremum at t_0 where t_0 is such that $r(t_0) = \mathbf{x}_0$. Thus, by the chain rule, we deduce that

$$0 = h'(t_0) = \nabla f(\mathbf{r}(t_0)) \cdot \mathbf{r}'(t_0).$$

- As the dot product of these two vectors is zero, they must be orthogonal.

In other words, $\nabla f(\mathbf{x}_0)$ is orthogonal to the tangent vector $\mathbf{r}'(t_0)$ for every curve C lying on $g = 0$ passing through $\mathbf{x}_0$. But if $g(\mathbf{x}) = 0$, we see that for any $\mathbf{u}$,

$$0 = D_{\mathbf{u}} g = \nabla g \cdot \mathbf{u},$$

which means that ∇g is also orthogonal to $\mathbf{r}'(t_0)$. Therefore, ∇f and ∇g must be parallel at the extreme point. In other words, there is a λ such that

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0).$$

This is the Lagrange Multiplier method.

Laplace and variational methods

- Beginning with Lagrange, mathematicians began to study functions defined by integration, often encountered in the calculus of variations, a subject initiated by him.
- Laplace discovered the method of steepest descent to analyse the asymptotic behavior of integrals.
- This allowed him to re-derive Stirling's formula as well as prove what is now called the de Moivre-Laplace law of large numbers.

The method of steepest descent and Stirling's formula

We begin with an informal derivation of Stirling's formula using the **method of steepest descent**. Suppose that $\psi : [0, \infty) \to \mathbb{R}$ is a smooth function with a unique global maximum at $x_0 > 0$ and that $\psi''(x_0) < 0$ and $\psi(x) \to -\infty$ as $x \to \infty$. Then, using the identities

$$\begin{aligned} \psi(x) &\approx \psi(x_0) + \psi'(x_0)(x - x_0) + \frac{1}{2}\psi''(x_0)(x - x_0)^2 + O((x - x_0)^3) \\ &= \psi(x_0) + \frac{1}{2}\psi''(x_0)(x - x_0)^2 \end{aligned}$$

- Recall:

$$\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx = \sqrt{2\pi\sigma^2}$$

$$\begin{aligned} \int_0^{\infty} e^{n\psi(x)} dx &\sim \int_{x_0-\epsilon}^{x_0+\epsilon} e^{n\psi(x)} dx \\ &\sim e^{n\psi(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} e^{\frac{1}{2}n\psi''(x_0)(x-x_0)^2} dx \\ &\sim e^{n\psi(x_0)} \int_{-\infty}^{\infty} e^{\frac{1}{2}n\psi''(x_0)(x-x_0)^2} dx \\ &\sim e^{n\psi(x_0)} \sqrt{\frac{2\pi}{-n\psi''(x_0)}}. \end{aligned}$$

To derive Stirling's formula, we write

$$n! = \Gamma(n+1) = \int_0^{\infty} x^n e^{-x} dx = \int_0^{\infty} e^{n\psi(x)} dx$$

where the function

$$\psi(x) = \log(x) - \frac{x}{n}$$

has a unique global maximum at $x_0 = n$. Calculating $\psi(x_0) = \log(n) - 1$ and $\psi''(x_0) = -n^{-2}$, the method of steepest descents gives

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

The de-Moivre-Laplace theorem

- Laplace could give another proof of what is often called the law of large numbers.
- In the context of Bernoulli trials, or coin flipping, de Moivre discovered the limit distribution as the normal distribution.
- His proof was long and complicated.
- Laplace found a more direct proof using his new asymptotic analysis.
- We will sketch his argument now.

The law of large numbers

- Although I will use the terminology of probability theory, you do not need to have had a course in it to understand the essential idea.

Suppose that $X_1, X_2, \cdots$ are i.i.d. random variables with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ and let $S_n = X_1 + \cdots + X_n$. Then, since $S_{2n} = 2k$ if and only if exactly $n + k$ of the X_i's take the value $+1$ and the remaining $n - k$ X_i's take the value -1, we have

$$\mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-2n}.$$

Our aim is to study the asymptotics of this probability in the limit as $n \to \infty$ when $k \approx x\sqrt{n/2}$.

From Stirling's formula, we have

$$\begin{aligned}
\binom{2n}{n+k} &= \frac{(2n)!}{(n+k)!(n-k)!} \\
&\sim \left(\frac{2n}{e}\right)^{2n} \left(\frac{e}{n+k}\right)^{n+k} \left(\frac{e}{n-k}\right)^{n-k} \sqrt{\frac{4\pi n}{(2\pi(n+k)2\pi(n-k)}} \\
&= \frac{(2n)^{2n}}{(n+k)^{n+k}(n-k)^{n-k}} \sqrt{\frac{n}{\pi(n^2-k^2)}},
\end{aligned}$$

Putting this together

$$
\begin{aligned}
\mathbb{P}(S_{2n} = 2k) &\sim \left(1+\frac{k}{n}\right)^{-(n+k)} \left(1-\frac{k}{n}\right)^{-(n-k)} (n\pi)^{-1/2} \left(1-\frac{k^2}{n^2}\right)^{-1/2} \\
&= \left(1-\frac{k^2}{n^2}\right)^{-n} \left(1+\frac{k}{n}\right)^{-k} \left(1-\frac{k}{n}\right)^{k} (n\pi)^{-1/2} \left(1-\frac{k^2}{n^2}\right)^{-1/2} \\
&\sim \left(1-\frac{x^2}{2n}\right)^{-n} \left(1+\frac{x}{\sqrt{2n}}\right)^{-x\sqrt{n/2}} \left(1-\frac{x}{\sqrt{2n}}\right)^{x\sqrt{n/2}} (n\pi)^{-1/2} \left(1-\frac{x^2}{2n}\right)^{-1/2}.
\end{aligned}
$$

- We now recognize that we can use the familiar limits: $(1+x/n)^n$ tends to e^x as n tends to infinity. Thus:

$$
\begin{aligned}
\left(1-\frac{x^2}{2n}\right)^{-n} &\to e^{x^2/2} \\
\left(1+\frac{x}{\sqrt{2n}}\right)^{-x\sqrt{n/2}} &\to e^{-x^2/2} \\
\left(1-\frac{x}{\sqrt{2n}}\right)^{x\sqrt{n/2}} &\to e^{-x^2/2} \\
\left(1-\frac{x^2}{2n}\right)^{-1/2} &\to 1.
\end{aligned}
$$

If $2k/\sqrt{2n} \to x$, then

$$
\mathbb{P}(S_{2n} = 2k) \sim (\pi n)^{-1/2} e^{-x^2/2}.
$$

The final steps

$$\begin{aligned}
\mathbb{P}(a\sqrt{2n} \le S_{2n} < b\sqrt{2n}) &= \sum_{m\in[a\sqrt{2n},b\sqrt{2n})\cap 2\mathbb{Z}} \mathbb{P}(S_{2n}=m) \\
&\sim \sum_{x\in[a,b]\cap(2\mathbb{Z}/\sqrt{2n})} (n\pi)^{-1/2}e^{-x^2/2} \\
&= \sum_{x\in[a,b]\cap(2\mathbb{Z}/\sqrt{2n})} (2\pi)^{-1/2}e^{-x^2/2}\left(\frac{2}{n}\right)^{1/2} \\
&\to \int_a^b \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \;=\; \mathbb{P}(a \le \chi \le b),
\end{aligned}$$

- keeping in mind that the length of our interval is (b-a)√(2n) so that the penultimate sum is recognized as the Riemann sum converging to the final integral.
- This theorem is the beginning of probability theory.

Legendre and number theory

- In 1797-98, Legendre published his two volume treatise on number theory.
- The most important idea in it concerned solutions of quadratic congruences.
- If p and q are primes, when can we solve the congruence $x^2 = q \pmod{p}$?
- If we can solve this congruence, Legendre defined the symbol (q/p) to be 1. This is not a "fraction" but rather symbolic notation.
- Along with Euler, he discovered the relation that
- $(p/q)(q/p) = (-1)^{(p-1)(q-1)/4}$ for odd primes p, q.
- This is often called the law of quadratic reciprocity but neither Legendre or Euler gave a complete proof. This was done later by Gauss.

Legendre and the study of primes

- Following Legendre, let $\pi(n)$ be the number of primes up to n.
- In his two-volume treatise, he conjectured but could not prove that $\pi(n)$ is asymptotically n/(log n -1.08366)
- Now we know this conjecture is wrong, but it does come close to the truth in the sense that the correct term is n/log n.
- This was proved almost a century after Legendre conjectured it, in 1896, by Hadamard and de la Vallee Poussin and can be seen as the culmination of 19th century mathematics. We will discuss this later.

Legendre and his photograph

- Apparently, until 2005, many scholars were using a wrong photograph of Legendre.
- The photo of the politician Louis Legendre was erroneously used in most books.
- There is, as far as we know, no portrait of him except for the 1820 watercolor caricature of the mathematicians Legendre and Fourier by Julien Leopold Boilly.