# THE invariant


Hilary Term 2016

Fellow Mathematicians,

Hello and welcome to the new Hilary term. It has been 80 years since the *Invariants* was set up by a group of dedicated and brilliant Maths students aiming to promote the realm of the Queen of all Sciences not covered by the degree curriculum. Ever since, the *Invariants* have been there for the Oxford students to inspire them by encouraging learned discussions with distinguished Mathematicians invited to give dedicated talks for the society's members. To honour this tradition, the Society is proud to invite you all, fellow Invariants, to a Anniversary Dinner to be held at Balliol College on 19 February. Updates on this event will be showing up on the Society's website.

This Hilary Term issue of the *Invariant* sheds light on the—often underappreciated—usefulness of the divergent series which, if handled with care, can be an interesting and extension of the discussion of their convergent counterparts. We are also happy to include an article written by our colleagues from the Physics Department on the Noether's Theorem and its implications—a wonderfully mathematical way of looking at conservation laws in physical sciences. As a somewhat lighter respite from mathematically rigorous texts, you may also enjoy the "Mathematical Foundation of Classical Ballet" which takes an interesting look at the rigidly structured art of ballet dancing.

Wishing you all a gratifyingly challenging and successful term,

*Lawrence Kurowski*

Magazine Editor



Authors:

Nela Cicmil
*Dept Physiology, Anatomy & Genetics*

Matthew Elliot
*Magdalen College*

Matthew Steggles
*Magdalen College*

Benjamin Jarvis
*Keble College*

# Contents



## Term Card

Week 1 – Tuesday 19 January – L2

Tom Sanders

'Combinatorics and the Fourier transform'

Week 2 – Tuesday 26 January – L2

Event to be confirmed

Week 3 – Tuesday 2 February – L2

Vicky Neale

Week 4 – Tuesday 9 February – L2

Richard Thomas

'Curves—an introduction to algebraic geometry and topological string theory'

Week 5 – Tuesday 16 February – L2

Grace Stirling from ATASS Sports

'Smashing the racket: Detecting match-fixing in tennis via in-play betting irregularities'

– Friday 19 February – L2

*Anniversary Dinner*

Robin Wilson

Week 6 – Tuesday 23 February – L2

Diane Maclagan

'Tropical geometry'

Week 7 – Tuesday 1 March – L2

Charles Batty

'Transfinite Induction and Fourier Series'

Week 8 – Tuesday 8 March – Common Room

Annual General Meeting

# Divergent Series

Benjamin Jarvis

## The Invention of the Devil?

Consider the following fallacious "proof" that *Grandi's series* $1-1+1-1+\dots$ is equal to $\frac{1}{2}$:
Set

$$S_1 = 1 - 1 + 1 - 1 + \dots$$

Then,

$$\begin{aligned} S_1 &= 1 - 1 + 1 - 1 + \dots \\ &= 1 - (1 - 1 + 1 - 1 + \dots) \\ &= 1 - S_1 \end{aligned}$$

Thus, $2S_1 = 1$, and hence $S_1 = \frac{1}{2}$.

$$1 - 1 + 1 - 1 + \cdots = \frac{1}{2} \tag{1}$$

Using this we can now derive the even more bizarre equation $1-2+3-4+\cdots = \frac{1}{4}$:
Set

$$S_2 = 1 - 2 + 3 - 4 + \dots$$

Then,

$$\begin{aligned} 2S_2 &= 1 - 2 + 3 - 4 + \dots \\ &\qquad + 1 - 2 + 3 - \dots \\ &= 1 - 1 + 1 - 1 + \dots \\ &= S_1 = \tfrac{1}{2} \end{aligned}$$

Thus, $2S_1 = \frac{1}{2}$, and hence $S_2 = \frac{1}{4}$.

$$1 - 2 + 3 - 4 + \cdots = \frac{1}{4} \tag{2}$$

There is already clearly something wrong here, but it gets worse. Using much the same logic we can derive a truly fatal result:

Set

$$\begin{aligned} t &= 1 + 1 + 1 + 1 \dots \\ &= 1 + (1 + 1 + 1 \dots) \\ &= 1 + t \end{aligned} \tag{3}$$

Hence, $0 = 1$.

What's gone wrong here? I suspect the answer will be immediately obvious to anyone who has taken a basic first course in Real Analysis: the series considered here are *divergent*.

Recall the definition of *convergence* for an infinite series $\sum a_n$, due to Cauchy:

$\sum_{n=0}^{\infty} a_n$ *converges* to a limit $L$ if for all $\varepsilon > 0$ there exists $N(\varepsilon) > 0$ such that for all $n \geq N(\varepsilon)$: $|s_n - L| < \varepsilon$, where $s_n = a_0 + \cdots + a_n$ is the $n^{\text{th}}$ *partial sum* of the series. We then say that $L$ is the *sum* of the series $\sum a_n$.

Basic analysis tells us that convergent series can be manipulated much like finite sums, so that the proofs given above would be valid *if* these series were convergent. However, it is immediately clear that the series considered do *not* have sums under Cauchy's definition—they are divergent. This, then, is the logical flaw in the above proofs—they begin by assuming that each series converges to a sum. Given that the series are divergent this is false, and it is therefore no surprise that we arrive at absurd results.

In the light of the above it is no wonder that Abel famously declared that "divergent series are the invention of the devil, and it is shameful to base on them any demonstration whatsoever." This is what might be called the 'Abelian' view of divergent series, and there was a time when it was commonplace amongst mathematicians. I suspect it is also common amongst undergraduates who have recently encountered the theory of infinite series for the first time. It is however, a view that no trained mathematician nowadays would accept; over the course of the last century it has been demonstrated time and time again that there are in fact sensible ways of assigning values—generalized sums—to divergent series; ways which are not only logically consistent, but also highly useful. Under these definitions, the equations (1) and (2) are in a sense correct, as we shall see.

The emergence of a theory of divergent series happened, broadly speaking, in three stages:

Firstly, in the period before analysis was placed on a rigorous footing, many mathematicians—most notably Euler, who seems to have held the philosophical viewpoint that each series could be associated with a unique 'generalized sum'—heuristically manipulated divergent series to arrive at results that appeared to be correct.

Next, in the period after Cauchy and Weierstrass had laid the foundations for modern analysis, a small number of dissenters from the 'Abelian' viewpoint—including Cesàro and Borel—experimented with alternative ways of assigning values to infinite series which extended Cauchy's concept of convergence. Many of these ways proved to have significant applications throughout mathematics.

Finally, towards the end of the $19^{\text{th}}$ and beginning of the $20^{\text{th}}$ centuries, the development of *Abelian* and *Tauberian* theory—due to, amongst others, Frobenius, Tauber, and, perhaps most importantly, the Hardy–Littlewood partnership—created a unified theory of divergent series which could no longer be ignored by the mathematical mainstream.

In this brief introduction I plan to devote a section each to a 'snapshot' of each of these stages. My hope is that doing so will give an indication of how and why mathematicians came to view divergent series as legitimate mathematical objects, and will convince the reader that this is both a fascinating and important area of study.

## Euler's example

For a first glimpse of the power of divergent series, let's look at an example which was known to Euler:
It is well known that, for complex $x$:

$$1 + x + x^2 + \cdots = \frac{1}{1-x} \quad (|x| < 1) \tag{4}$$

and the sum diverges for $|x| \geq 1$. The right-hand side of this equation, however, makes sense for *all* complex $x \neq 1$. Abandoning all pretence of mathematical rigour, we might formally substitute $x$ in (4) for the complex number $e^{\theta i} = \cos\theta + i\sin\theta$ (for $0 < \theta < 2\pi$, so $e^{\theta i} \neq 1$), which has magnitude 1, to obtain:

$$\begin{aligned} 1 + e^{\theta i} + e^{2\theta i} + e^{3\theta i} + \ldots &= \frac{1}{1 - e^{\theta i}} \\ &= \frac{1}{2} + \frac{1}{2} i \cot\left(\frac{1}{2}\theta\right) \quad (0 < \theta < 2\pi) \end{aligned}$$

and, on taking real parts:

$$\cos\theta + \cos 2\theta + \cos 3\theta + \cdots = -\frac{1}{2} \quad (0 < \theta < 2\pi) \tag{5}$$

The series on the left-hand side of this equation is, unsurprisingly, divergent, so that the equation (5) is—according to the 'Abelian' view of divergent series—meaningless. Something peculiar happens, however, when we integrate this equation, term-by-term, between 0 and a new variable $\phi$; we obtain:

$$\sin\phi + \frac{1}{2}\sin 2\phi + \frac{1}{3}\sin 3\phi + \cdots = -\frac{1}{2}\phi + K \quad (0 < \phi < 2\pi)$$

where $K$ is a constant that we can evaluate to $\frac{\pi}{2}$ by setting $\phi = \pi$, giving us:

$$\sin\phi + \frac{1}{2}\sin 2\phi + \frac{1}{3}\sin 3\phi + \cdots = \frac{\pi - \phi}{2} \quad (0 < \phi < 2\pi) \tag{6}$$

The curious thing about this equation is that **it is true**: the series on the left-hand side is convergent to the expression on the right. We have thus manipulated a divergent series in such a way as to derive a correct result about a convergent series. (As an aside, the equation (6) can be used to obtain a remarkably slick solution to the Basel problem see [1] for details). At the end of this article, we shall be in a position to explain how this can be turned into a rigorous proof; for now, we can do much more with the series (5).

Consider, for complex $s$, the series

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \dots$$

This converges in the region $\mathrm{Re}(s) > 1$ and diverges everywhere else. In addition, in its region of convergence it defines an analytic function; we can extend this function uniquely to a function $\zeta(x)$—the *Riemann Zeta function*—which is defined and analytic everywhere in the complex plane except at $s = 1$, where it has a simple pole.

Due in part to a strong connection between its zeros and the distribution of prime numbers, established by Riemann in a seminal 1859 paper, the Zeta function is of unparalleled importance to modern analytic number theory. Despite its name it was actually first studied over 100 years before Riemann by Euler, who used arguments involving divergent series to make a number of remarkable conjectures about it, most notably conjecturing the functional equation

$$\zeta(1-s) = 2(2\pi)^{-s} \cos\left(\frac{1}{2}s\pi\right)\Gamma(s)\zeta(s)$$

which was eventually proven by Riemann.

A summary of how Euler did this can be found in [2]; to give a flavour of his methods we shall give a heuristic demonstration of the existence of its trivial zeros:

Recall that we have

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \text{ for } \mathrm{Re}\, s > 1 \tag{7}$$

Suppose that we wish to evaluate the Zeta function at negative even integers. The series (7) does not converge for $s = -2k$, but if we naively substitute this into the right-hand side we obtain the series

$$1 + 2^{2k} + 3^{2k} + 4^{2k} + \dots \tag{8}$$

Now, if we were to adopt Euler's philosophical viewpoint that each divergent series can be associated with a unique number—a generalized sum—and if we could find this number for the series (8), we might well expect that this number should be the value of $\zeta(-2k)$.

Let's look again at the series (5); the method that we used to derive it suggests that it should only be 'true'—whatever that means—for $0 < \theta < 2\pi$.
If we take yet another logical leap of faith, however, and set  to be the limit 0, we arrive at the bizarre equation:

$$1 + 1 + 1 + 1 + \dots = -\frac{1}{2} \tag{9}$$

We have already seen that assigning a value to this series is particularly fatal, but it turns out that the corresponding value of the Zeta function, $\zeta(0)$, actually is $-\frac{1}{2}$ . Already, this seems promising. Now, differentiating the series (5) $2k$ times, we obtain

$$\cos\theta + 2^{2k}\cos 2\theta + 3^{3k}\cos 3\theta + \cdots = 0$$

And, substituting $\theta = 0$:

$$1 + 2^{2k} + 3^{2k} + 4^{2k} + \cdots = 0$$

This leads us to conjecture that the Zeta function is zero at all even negative integers, and *again this turns out to be true*[1]

This example already strongly suggests that the Abelian view of divergent series is incomplete. In order to make sense of the above we shall have to adopt Euler's viewpoint that the sums of divergent series considered above are in some sense 'true'. To reconcile this with our modern rigorous approach we must extend our concept of the sum of an infinite series to include the divergent case.

## Summation Methods

In order to make progress we first have to define what we mean by the generalized sum of a divergent series. It turns out that there are a number of different possible *summation methods* which allow us to do this. A summation method $\mathcal{S}$ is a function from some set of series to the complex numbers, subject to some further conditions. For convenience (and to emphasise the connection with regular summation) we shall supress function notation and write $\sum a_n = L(\mathcal{S})$ or $a_0 + a_1 + a_2 + \cdots = L(\mathcal{S})$ for '$\mathcal{S}$ maps the series $\sum a_n$ to $L$'. We then say that the series is '$\mathcal{S}$-summable' with $L$ its $\mathcal{S}$-sum.

The axioms for a **summation method** $\mathcal{S}$ can now be stated as:

i.) **Linearity**: if $\sum a_n$ and $\sum b_n$ are $\mathcal{S}$-summable and $\lambda$ is a complex number, then $\sum(a_n + \lambda b_n)$ is $S$-summable and $\sum(a_n + \lambda b_n) = \sum a_n + \lambda \sum b_n$

ii.) **Regularity**: if $\sum a_n$ is a convergent series with sum $L$ then $\sum a_n$ is $\mathcal{S}$-summable with $\mathcal{S}$-sum $L$.

If, in addition, $S$ satisfies:

iii.) **Stability**: a series $a_0 + a_1 + a_2 + \ldots$ is $\mathcal{S}$-summable if and only if $a_1 + a_2 + \ldots$ is $\mathcal{S}$-summable and then $a_0 + (a_1 + a_2 + \ldots) = a_0 + a_1 + a_2 + \ldots (\mathcal{S})$

then we way that $\mathcal{S}$ is a **stable summation method.**

These are all properties that are satisfied by regular summation, which allow us to manipulate $\mathcal{S}$-sums much like convergent series; note that we do not require commutativity (i.e. permuting the terms of a series without altering the result)—it is a basic fact

[1]These are known as the trivial zeros of $\zeta(s)$; they are known to be the only zeros outside the critical strip $0 < \mathrm{Re}(s) < 1$, and conjectured to be the only zeros off the line $\mathrm{Re}(s) = \frac{1}{2}$.

that this does not hold for convergent series, so we certainly shouldn't expect it to hold for divergent series.

To begin to see the power of these axioms, note that we can easily convert the fallacious 'proofs' (1) and (2) into valid proofs that any generalized stable summation method which can sum these series must sum them to these values, and (3) into a proof that no stable method can sum $1 + 1 + 1 + \dots$ (this doesn't conflict with (9) as the summation method implicitly considered there is not stable).

It is worth noting briefly at this point that the axioms for a stable summation method imply that we can insert a finite number of zeros into a series without altering its sum. This is known as *finite interpolation.* As we shall see, we certainly do not have infinite interpolation in general.

We shall now describe two of the most important summation methods:

### Cesàro Summation

Our first method of assigning a sum to a divergent series is due to the Italian mathematician Ernesto Cesaro. We say that a series $\sum a_n$ is *Cesaro-summable* to $L$ (written $\sum a_n = L(\mathcal{C})$) if:

$$t_n = \frac{1}{n+1} \sum_{k=0}^{n} S_k \to L \text{ (where } S_k = a_0 + \cdots + a_k)$$

That is, the Cesaro-sum of a series is the limit of the averages of consecutive partial sums (known as the *Cesaro-means*) if this limit exists.

It is a basic exercise to show that this method is linear, regular and stable.

For a basic example of the use of this method, let's look again at Grandi's series. Recall that we know that *if* there is a stable summation method that sums this series then it must sum it to $\frac{1}{2}$. This does not imply that such a method exists, but it is easy to see that Cesaro-summation works:

$$a_n = \begin{cases} +1 & n \text{ even} \\ -1 & n \text{ odd} \end{cases}$$

So that

$$S_n = \begin{cases} 1 & n \text{ even} \\ 0 & n \text{ odd} \end{cases}$$

And

$$\sum_{k=0}^{n} S_k = \begin{cases} \frac{n}{2} + 1 & n \text{ even} \\ \frac{n+1}{2} & n \text{ odd} \end{cases}$$

Thus

$$t_n = \begin{cases} \frac{1}{2}\left(1 + \frac{1}{n+1}\right) & n \text{ even} \\ \frac{1}{2} & n \text{ odd} \end{cases}$$

This sequence clearly approaches $\frac{1}{2}$, so that we have obtained:

$$1 - 1 + 1 - 1 + \cdots = \frac{1}{2}(\mathcal{C})$$

We have thus shown that there is indeed a sensible way of associating the number $\frac{1}{2}$ with Grandi's series; the argument used in the introduction wasn't nonsense after all! (The reader might like to verify, however, that $1-2+3-4+\ldots$ is *not* Cesaro-summable; we shall need a stronger summation method for this series.)

We arrive at a serious difference between Cesàroand regular summation when we consider versions of Grandi's series interpolated with infinitely many zeros—so-called lacunary versions of the series. For example, consider the series:

$$1 + 0 - 1 + 1 + 0 - 1 + \cdots = \sum_{n=0}^{\infty} a_n$$

$$\text{where } a_n = \begin{cases} 1 & n \equiv 0 \mod 3 \\ 0 & n \equiv 1 \mod 3 \\ -1 & n \equiv 2 \mod 3 \end{cases}$$

Then

$$S_n = \begin{cases} 1 & n \equiv 0 \mod 3 \\ 1 & n \equiv 1 \mod 3 \\ 0 & n \equiv 2 \mod 3 \end{cases}$$

So that

$$\sum_{k=0}^{n} S_k = \begin{cases} \frac{2}{3}\left(n + \frac{2}{3}\right) & n \equiv 0 \mod 3 \\ \frac{2}{3}(n+2) & n \equiv 1 \mod 3 \\ \frac{2}{3} & n \equiv 2 \mod 3 \end{cases}$$

Hence

$$t_n = \begin{cases} \frac{2}{3}\left(1 + \frac{1}{2(n+1)}\right) & n \equiv 0 \mod 3 \\ \frac{2}{3}\left(1 + \frac{1}{n+1}\right) & n \equiv 1 \mod 3 \\ \frac{2}{3} & n \equiv 2 \mod 3 \end{cases}$$

But this series clearly converges to $\frac{2}{3}$, so that:

$$1 + 0 - 1 + 1 + 0 - 1 + \cdots = \frac{2}{3}(\mathcal{C})$$

Hence, in stark contrast with regular summation, interpolating an infinite number of zeros into a series can alter its Cesaro-sum. Even worse, consider *Hardy's series*:

$$1 - 1 + 0 + 1 + 0 + 0 + 0 - 1 + \cdots = \sum_{n=1}^{\infty} a_n$$

$$\text{where } a_n = \begin{cases} (-1)^k & \text{if } n = 2^k \\ 0 & \text{otherwise} \end{cases}$$

This series is in fact *not* Cesaro-summable; this can be verified directly or deduced as a consequence of a remarkable *Tauberian* theorem due to Hardy and Littlewood known as the **High-Indices theorem**. Roughly, this tells us that even in contexts more general than Cesàrosummation this problem always arises. A summable series can always be made into a non-summable series by interpolating so that the non-zero terms are spread sufficiently thinly. This is a first hint of the deep fact that we have no hope of ever 'taming' all series with a single summation method.

Why is Cesàrosummation a useful concept? In short, because it gives the next best thing after convergence: in addition to being linear, regular and stable, the Cesaro-sum can easily be computed to any degree of accuracy required, and it interacts well with the multiplication of series (better, in fact, than regular summation does) as well as with the operations of calculus. We can therefore use Cesàrosummation to manipulate a much larger class of series than the convergent series, using almost all of the properties that make convergence such a powerful notion. For example, it can be shown that the equation (5) is in fact true under the Cesàrosum; this hints at how we could make rigorous the heuristic derivations we based on this equation.

A famous application of Cesaro's method comes from the theory of Fourier series. In this field we are concerned with representing a given continuous periodic function $f(\theta)$ (presumed here to have period $2\pi$) by a trigonometric series of the form:

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta) \tag{10}$$

with coefficients:

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \cos(n\theta) d\theta$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \sin(n\theta) d\theta$$

This is known as the Fourier series for $f(\theta)$, and very often it converges to $f(\theta)$. We would like to know for which functions the series (10) converges (preferably uniformly) to $f(\theta)$, so that we have:

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta) \tag{11}$$

In 1807, Fourier announced—to widespread disbelief—that for *any* continuous periodic function $f(\theta)$ the series (11) converges. This is in fact not quite true, but something almost as good is. For a remarkably large class of functions the Fourier series does converge (for example the periodic functions with left- and right-hand derivatives everywhere), but in 1910 the Hungarian mathematician Lipot Fejer gave an explicit example of a continuous function with a divergent Fourier series. He also, however, managed to prove the remarkable **Fejer's theorem**, which states that for any continuous periodic function $f(\theta)$ the Fourier series *is* Cesaro-summable to $f(\theta)$ and its Cesaro-means converge uniformly. That is:

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta)\,(\mathcal{C})$$

This is useful for precisely the same reasons that a convergent Fourier series is: we can easily calculate the Cesaro-means and hence use the Fourier series to calculate $f(\theta)$ to any desired degree of accuracy; we can manipulate the series just as we would in the convergent case; and (due to uniformity) we can integrate and differentiate term-by-term. This illustrates a common theme in the theory of divergent series: if a generalized summation method preserves sufficiently many properties of regular summation then it often ends up being useful for the same reasons, whilst allowing us to work with a much larger class of series.

## Abel Summation

Our second method derives its name from **Abel's theorem**, a classical theorem of analysis which states:

i.) if $f(x)$ is an analytic function on the interval $(-1, 1)$ with power series expansion

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \quad (|x| < 1)$$

and

ii.) if the series

$$\sum_{n=0}^{\infty} a_n \tag{12}$$

converges to a limit $L$

then:

$$\lim_{x \to^- 1} f(x) = L \tag{13}$$

This result is often used to evaluate convergent series; for example, if we wish to evaluate the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$$

we first note that this series converges by the alternating series test, and that

$$\ln(1 + x) = x - \frac{1}{2}x + \frac{1}{3}x^2 - \frac{1}{4}x^3 + \ldots \quad (|x| < 1)$$

by the Taylor expansion. We then use Abel's theorem to deduce that

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \lim_{x \to ^- 1} \ln(1+x) = \ln(2) \tag{14}$$

The key insight that leads us to our next summation method is that even if the series (12) doesn't converge, the limit (13) may still exist; we may then say that this limit is in some sense the 'correct' generalized sum of (12).

Formally, we say that a series $\sum a_n$ is *Abel-summable* to $L$ (denoted $\sum a_n = L(\mathcal{A})$) if the power series

$$\sum_{n=0}^{\infty} a_n x^n$$

converges for all $|x| < 1$ and

$$\lim_{x \to ^- 1} \left( \sum_{n=0}^{\infty} a_n x^n \right) = L$$

It is now easy to prove that Abel summation is a stable summation method; linearity and stability follow immediately from the algebra of limits, and the statement of regularity is simply Abel's theorem.

As an example of the use of this method, we can sum both versions of Grandi's series that we Cesaro-summed in the previous section; for example:

$$1 - x + x^2 - x^3 + \cdots = \frac{1}{1-x} \quad (|x| < 1)$$

and

$$\lim_{x \to ^- 1} \left( \frac{1}{1+x} \right) = \frac{1}{2}$$

so that

$$1 - 1 + 1 - 1 + \cdots = \frac{1}{2}(\mathcal{A})$$

Further, we can now successfully sum the series $1 - 2 + 3 - 4 + \ldots$ which we evaluated heuristically to $\frac{1}{4}$ in the introduction:

$$\begin{aligned} 1 - 2x + 3x^2 - 4x^3 + \ldots &= (1 - x + x^2 - x^3 + \ldots)^2 \\ &= \frac{1}{(1+x)^2} \quad (|x| < 1) \end{aligned}$$

and

$$\lim_{x \to ^- 1} \left( \frac{1}{(1+x)^2} \right) = \frac{1}{4}$$

so that

$$1 - 2 + 3 - 4 + \cdots = \frac{1}{4}(\mathcal{A})$$

This already suggests that Abel summation is in a sense stronger than Cesàrosummation; this is true, and will be made precise in the next section. It still has its limitations, however. Just like Cesaro's method it cannot sum series which diverge 'too fast'; e.g. it is immediate from the definition that if the sequence $(a_n/2^n)$ doesn't approach zero then the series $\sum a_n$ is not Abel-summable. e.g. $1! + 2! + 3! + \dots$ is not Abel-summable.

We have already hinted at one of the reasons that this is a useful concept: our derivation of the equation (14) implicitly proceeded by showing that the series is Abel-summable to $\ln(2)$, before 'shifting down' to regular summability. This illustrates another theme in the theory of divergent series: often the easiest way to study *convergent* series is to shift up to some larger class of summable series where things are in some sense easier, and then shift back down. (This is somewhat analogous to finding real solution of a polynomial equation by finding all of the complex solutions (which we know exist) and then showing which, if any, are real.)

## Abelian and Tauberian theorems

The true value of the concept of generalized sums for divergent series only becomes apparent when we stop thinking only of individual summation methods and start to consider the ways that different methods relate to each other. This is made precise through the use of *Abelian* and *Tauberian* theory, which make up the core of the study of divergent series. Roughly speaking, an Abelian theorem tells us that some summation method $\mathcal{A}$ is stronger than another method $\mathcal{B}$, in that any $\mathcal{B}$-summable series is also $\mathcal{A}$-summable, and to the same value. A Tauberian theorem is a partial converse to this, telling us that given some 'size condition' on its terms, an $\mathcal{A}$-summable series is $\mathcal{B}$-summable, again to the same sum.

The prototype for an Abelian theorem is **Abel's theorem**, mentioned in the previous section, which tells us that any convergent series is Abel-summable; another example is the regularity result for Cesàrosums.

The first Abelian theorem relating two generalized summability methods to be discovered was found by Frobenius in 1880, which tells us that Abel-summation is stronger than Cesaro-summation.

**Frobenius' Theorem**: If $\sum a_n$ is Cesaro-summable then it is also Abel-summable, and to the same sum.

For an example of the power of this theorem, suppose that we wish to find

$$\lim_{x \to ^-1} \sum_{n=0}^{\infty} (-1)^n x^{n^2}$$

To do this, we forget completely about the power series structure and simply consider the corresponding divergent series:

$$\begin{aligned} & 1 - 1 + 0 + 0 + 1 + 0 + 0 + 0 + 0 - 1 + \dots \qquad (15) \\ &= \sum_{k=0}^{\infty} a_k, \text{ where } a_k = \begin{cases} (-1)^n & \text{if } k = n^2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

This is of course another lacunary version of Grandi's series, and it is fairly easy to show that it is Cesaro-summable to $\frac{1}{2}$ (consider the local maxima and minima of the Cesaro-means, at odd and even perfect squares respectively, and show that both these subsequences converge to $\frac{1}{2}$). Now we can use Frobenius' theorem to deduce that (15) is Abel-summable to the same sum. Hence:

$$\lim_{x \to^- 1} \sum_{n=0}^{\infty} (-1)^n x^{n^2} = \frac{1}{2}$$

We have thus managed to find a complicated limit without having to calculate a single value of this power series!

The prototypical Tauberian theorem was discovered by Tauber in 1897:

**Tauber's Theorem**: If $\sum a_n$ is Abel-summable to $L$ and $a_n = o\left(\frac{1}{n}\right)$ then $\sum a_n$ converges to $L$.

Typically of a Tauberian theorem, this result is important for two dual reasons. Firstly, in conjunction with Abel's theorem (to which it gives a partial converse) it allows us to work in the (much larger) class of Abel-summable series and then 'shift down' to the class of convergent series to give valid results. Thus we can use the machinery of power series of analytic functions (which is implicit in the concept of Abel-summability) to study convergent series. Secondly, and perhaps more interestingly from a theoretical perspective, the contrapositive of this theorem gives a limit on the applicability of this summation method: it tells us that a divergent series whose nth term is $o\left(\frac{1}{n}\right)$ (for example $\sum \frac{1}{n \ln n}$, $n \geq 2$, which diverges by the Cauchy Condensation test) cannot be Abel-summable. We already knew that a series could diverge too fast to be Abel-summable; now we know that a series can also diverge too *slowly*. A similar restriction exists for every useful notion of summability.

There are two more Tauberian theorems worth mentioning briefly here, both due to the Hardy-Littlewood partnership. The first gives a condition for a Cesaro-summable series to be convergent; the second gives a partial converse to Frobenius' theorem:

**Hardy's theorem**: If $\sum a_n$ is Cesaro-summable to $L$ and $a_n = \mathcal{O}\left(\frac{1}{n}\right)$ then $\sum a_n$ converges to $L$.

**Hardy–Littlewood theorem**: If $\sum a_n$ is Abel-summable to $L$ and the partial sums $S_k$ of the series are bounded then $\sum a_n$ is Cesaro-summable to $L$.

To give a final example of the power of the theory of divergent series, we return to the convergent series (6) discussed in the second section. Recall that we gave a heuristic argument for this result based on unjustified manipulations of divergent series. Using Tauberian theory we can now outline a scheme to turn this into a rigorous proof:

- Taking $0 < \theta < 2\pi$, consider the (convergent) geometric series:

$$1 + e^{\theta i} x + e^{2\theta i} x^2 + e^{3\theta i} x^3 + \cdots = \frac{1}{1 - e^{\theta i} x} \quad (|x| < 1)$$

- Take the limit as $x$ approaches 1 to obtain the Abel sum:

$$1 + e^{\theta i} + e^{2\theta i} + e^{3\theta i} + \cdots = \frac{1}{1 - e^{\theta i}} = \frac{1}{2} + \frac{1}{2} i \cot\left(\frac{1}{2}\theta\right) (\mathcal{A})$$

- Take real parts to obtain:

$$\cos\theta + \cos 2\theta + \cos 3\theta + \cdots = -\frac{1}{2}(\mathcal{A})$$

- Now use the Hardy-Littlewood theorem to show that this is in fact a Cesàrosum:

$$\cos\theta + \cos 2\theta + \cos 3\theta + \cdots = -\frac{1}{2}(\mathcal{C})$$

- Show that the Cesàromeans of this series converge uniformly, and use this to justify integrating the series term-by-term between 0 and a new variable $\phi$, giving:

$$\sin\phi + \frac{1}{2}\sin 2\phi + \frac{1}{3}\sin 3\phi + \cdots = -\frac{1}{2}\phi + K(\mathcal{C}) \quad (0 < \phi < 2\pi)$$

  where $K$ is constant.

- Use Hardy's theorem to deduce that this series is in fact convergent, so that we have:

$$\sin\phi + \frac{1}{2}\sin 2\phi + \frac{1}{3}\sin 3\phi + \cdots = -\frac{1}{2}\phi + K \quad (0 < \phi < 2\pi)$$

- Substitute $\phi = \pi$ to evaluate $K$ to be $\frac{\pi}{2}$, and so:

$$\sin\phi + \frac{1}{2}\sin 2\phi + \frac{1}{3}\sin 3\phi + \cdots = \frac{\pi - \phi}{2} \quad (0 < \phi < 2\pi)$$

Similar—though significantly more difficult—schemes can be devised to make our arguments on the Zeta function rigorous.

## Further Reading

[1] Divergent Series—G.H. Hardy: the classic book on the topic, written by one of the pioneers of the theory, includes everything covered in this article and much more.

[2] Divergent Series: Why $1+2+3+ = -1/12$—Bryden Cais: a brief article on the topic which contains a summary of Euler's heuristic derivation of the functional equation for the Zeta function.

[3] Invitation to Classical Analysis—Peter Duren: this book contains an excellent short section on Tauberian theory, which was the basis for my exposition of the subject.

[4] Summability of Alternating Gap Series—J.P. Keating and J.B. Reade: this short article on one of the more interesting aspects of Tauberian theory concerns itself with the summability of lacunary series, and contains an in-depth study of the fascinating Hardy's series and an introduction to the High-Indices theorem.

[5] Fourier Series—G.P. Tolstov: contains a chapter on the applications of generalized summation methods to harmonic analysis.

# Noether's Theorem: Symmetries, Invariance, and Conservation Laws

Matthew Steggles and Matthew Elliot

*The old guard at Göttingen should take some lessons from Miss Noether! She seems to know her stuff* - Albert Einstein

## Introduction

In the world around us, we are increasingly bombarded with the need for conservation—conservation of rainforests, conservation of endangered species, conservation of rare minerals and so on. The one key difference between these and the subject of this piece is the necessity in all of these for human intervention. There is, however, another form of conservation that enshrines itself above these human matters due to the total lack of intervention needed for them to apply—physical conservation laws. Energy, momentum, angular momentum; all commonly conserved quantities in elementary dynamics problems. Among others, the mere fact that these quantities do not change over time allows an obscene amount of insight into the dynamics of the world.

We (the authors) are of course being facetious, but it does raise a question: how do we know they are conserved? While the well known conservation laws can be shown from Newton's laws, doing it in this way would (as well as being tedious) lack a crucial insight into the underlying mathematics as well as the subtlety of identifying *when* these laws apply. The particular celebrity who's work we will examine here is Emmy Noether, German mathematician of the early twentieth century.

Noether was born in 1882 in the Bavarian city of Erlangen. The daughter of another mathematician, Max Noether, she spent her childhood with the desire to teach English and French, before turning to mathematics, studying at the university in her hometown. Despite her obvious talent, Noether experienced a good deal of the endemic sexism of the time—after completing her dissertation she worked in Erlangen without pay for seven years. Following this, she was invited to Göttingen by Hilbert and Klein, but rejected by the philosophical department, citing her sex as the reason. In spite of this, she proceeded to lecture in Göttingen for four years, under Hilbert's name, before finally being given a professorship in 1919.

While Noether is lauded among mathematicians for her work in abstract algebra, achieved later in life, we shall be concerning ourselves here with her earlier work on invariants. First proved in 1915, and published in 1918, what is now known as Noether's theorem elegantly links our conservation laws with invariance of the Lagrangian, a particular function of importance to physicists. Noether's theorem stated in full is: Any differentiable symmetry of the action has a corresponding conserved quantity. In order to understand what this means we will briefly explore first the framework of Lagrangian mechanics before demonstrating Noether's theorem. We finish by discussing a comparable phenomenon with deep physical insight—Ostrogradsky instability.

## The Euler-Lagrange equation

First, define the *action* $S = \int_{t_1}^{t_2} \mathcal{L}(q_i, \dot{q}_i, t)dt$, where $q_i$ are the $i$ generalised coordinates of the system and $\mathcal{L} = T - U$ is the Lagrangian, equal to the kinetic energy of the system take away the potential energy. A generalised coordinate can be any parameter that describes the configuration of the system, such as distance along the $x$ axis, or the azimuth $\theta$ in a spherical system. For example, an experiment with a single simple pendulum only needs one generalised coordinate to fully describe the motion of the system—the angle $\theta$ between the pendulum and the vertical. Next, we assert Hamilton's principle of stationary action, stating that the action integral is stationary to first order. With these in mind, we perturb $q$ in the following manner (in one dimension, without loss of generality):

$$q(t) \mapsto q'(t) \equiv q(t) + \varepsilon\eta(t) \qquad \eta(t_1) = \eta(t_2) = 0$$

Now differentiating the action with respect to $\varepsilon$ and noting Hamilton's principle, we find:

$$\frac{dS}{d\varepsilon} = \int_{t_1}^{t_2} \frac{d\mathcal{L}(q', \dot{q}', t)}{d\varepsilon} dt = \int_{t_1}^{t_2} \left( \eta \frac{\partial \mathcal{L}}{\partial q} + \dot{\eta} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) dt = 0$$

Next, by integrating the second term of the integrand by parts:

$$\frac{dS}{d\varepsilon} = \int_{t_1}^{t_2} \left( \eta \frac{\partial \mathcal{L}}{\partial q} - \eta \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) dt + \left[ \eta \frac{\partial \mathcal{L}}{\partial \dot{q}} \right]_{t_1}^{t_2}$$

By remembering the boundary condition $\eta(t_1) = \eta(t_2) = 0$, then we may write:

$$\int_{t_1}^{t_2} \left( \frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \eta dt = 0 \implies \frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} = 0$$

This is the Euler-Lagrange equation, and it is entirely equivalent to Newton's laws, only far more elegant and in many ways easier to use as no awkward consideration of forces is necessary; the information is contained within the energies (The other solution to making the integral zero; $\eta = 0$; merely refers to the coordinate $q$ being totally unperturbed on its path). The generalisation to $i$ coordinates is trivial—there is simply one Euler-Lagrange equation for each of the coordinate. For an intuitive reason as to why the Lagrangian, a seemingly arbitrary function, is of such importance, read the short casual paper "The Origin of the Lagrangian" by Matt Guthrie.

# Noether's theorem

### Each differentiable symmetry of the action has a corresponding conserved quantity

Now, we arrive on the subject of Noether's theorem, armed with the Euler-Lagrange equation. Reading Noether's theorem again, we see we are looking at differentiable symmetries of the action, that is, transformations that leave $\delta\mathcal{L} = 0$. Before we proceed further, we must define a few more expressions. First, we define the *conjugate momentum* of each coordinate, defined as

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$$

If we use a cartesian coordinate system, and a potential that depends only on position, then it is not so difficult to see that the conjugate momenta are simply the linear momenta of the system. The next thing to define is the *Hamiltonian*, defined to be

$$\mathcal{H} = \sum_i p_i \dot{q}_i - \mathcal{L}$$

The Hamiltonian has a strong relationship to the total energy of the system, as we will see later. Now, as mentioned earlier, we shall be focusing on transformations leaving $\delta\mathcal{L} = 0$. Assuming we have such a transformation:

$$\begin{aligned}\delta\mathcal{L} &= \sum_i \frac{\partial \mathcal{L}}{\partial q_i}\delta q_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i}\delta \dot{q}_i = \sum_i \underbrace{\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_i}}_{\text{from E-L equation}} \delta q_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i}\delta \dot{q}_i \\ &= \sum_i \dot{p}_i \delta q_i + p_i \delta \dot{q}_i \qquad = \frac{d}{dt}\sum_i p_i \delta q_i = 0\end{aligned}$$

This tells us something very interesting indeed: provided we have some transformation to $q_i$ that leaves $\mathcal{L}$ untouched, then it automatically tells us that the quantity $\sum_i p_i \delta q_i$ does not vary with time, that is to say it is conserved! This proves Noether's theorem. Here are a few examples, to illustrate the power of this theorem:

### Conservation of linear momentum

First, let us rewrite $\delta q_i$ as $f(q_i)\delta \equiv f_i\delta$, i.e. that $q_i' = q_i + f_i\delta$, and our conserved quantity is $\sum_i p_i f_i$. Now, in a cartesian coordinate system, we decide to translate $x$;

$$x \mapsto x + \delta$$

We clearly see that in this case, $f = 1$, so our conserved quantity is simply $p_x$. If translation leaves the Lagrangian unchanged, then linear momentum is necessarily conserved!

### Conservation of angular momentum

Now, in a 3-dimensional system, apply an anticlockwise rotation about the $z$ axis, and allow this to preserve the Lagrangian as is:

$$x \mapsto x\cos\delta - y\sin\delta = x - y\delta \quad , \quad y \mapsto y\cos\delta + x\sin\delta = y + x\delta$$

In the limit that $\delta$ is small. We see now that $f_x = -y$ and $f_y = x$, so our conserved quantity is:

$$\sum_i p_i f_i = xp_y - yp_x = (\mathbf{r} \times \mathbf{p})_z$$

Which is the $z$ component of angular momentum.

### Conservation of energy

In order for a continuous transformation of the Lagrangian through time to be symmetric, then the Lagrangian can not itself be a function of time.

$$\frac{d\mathcal{L}}{dt} = \sum_i \frac{\partial\mathcal{L}}{\partial q_i}\dot{q}_i + \frac{\partial\mathcal{L}}{\partial\dot{q}_i}\ddot{q}_i = \frac{d}{dt}\sum_i p_i\dot{q}_i$$

This substitutes directly into our definition of the Hamiltonian to give

$$\frac{d}{dt}\left(\sum_i p_i\dot{q}_i - \mathcal{L}\right) = \frac{d\mathcal{H}}{dt} = 0$$

So the Hamiltonian is conserved through time. We identify the Hamiltonian as being equivalent to the total energy of the system, and indeed it holds that in any system where the Lagrangian does not depend explicitly on time, then energy is conserved. This is fairly intuitive when you step back and think about it - the Lagrangian is a function of the kinetic and potential energies, and one can reason that having either of these depend on time implies that energy is being added or taken from the system, for example by a motor.

However, the true physical significance of this has been (intentionally) glossed over until now - remember that the Euler-Lagrange equation entirely contains Newton's laws? This is the same as saying it governs the equations of motion for the system. When we say *"the Lagrangian is unchanged under some transformation"*, in the physical world it means *"when we do an experiment, the result is unchanged"*. Combining this with Noether's theorem, this provides an incredibly powerful set of tools to determine what in a given circumstance is or is not conserved; if an experiment can simply be moved to a different place, and return the same results, then that and that alone is enough for us to say that linear momentum is conserved. If we may rotate our apparatus and redo the experiment and gain similar results, then angular momentum is conserved, and finally, if we may simply *go away and come back again later* and do the same experiment, and once again it gives us the same results as before, then energy too is conserved! This is the beauty and immediate significance of Noether's theorem.

## Ostrogradsky instability

Here, we will explore a consequence of having Lagrangians of higher orders, using Hamiltonian dynamics—a similar set of tools to Lagrangian dynamics. Ostrogradsky's theorem states that *a non-degenerate Lagrangian dependent on time derivatives higher than first order leads to a linearly unstable Hamiltonian.* We begin by deriving Hamilton's equations, from $d\mathcal{H}$ and $d\mathcal{L}$. Notice that at the beginning we tacitly left off a $\frac{\partial \mathcal{L}}{\partial t}dt$ term, as we did not assert that $\mathcal{L}$ did not depend on $t$. It was not necessary then, and will not be here either, so we shall again leave it off, purely to avoid clutter.

$$d\mathcal{L} = \sum_i \left( \frac{\partial \mathcal{L}}{\partial q_i} dq_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i} d\dot{q}_i \right) = \sum_i \left( p_i d\dot{q}_i + \dot{p}_i dq_i \right) )$$

Now we may substitute this into a similar expression for the Hamiltonian:

$$d\mathcal{H} = \sum_i \left( p_i d\dot{q}_i + \dot{q}_i dp_i \right) - d\mathcal{L} = \sum_i \left( \dot{q}_i dp_i - \dot{p}_i dq \right)$$

Shuffling this expression around yields Hamilton's equations, and comparing the above two lines gives another important relation we shall use.

$$\frac{\partial \mathcal{H}}{\partial p_i} = \dot{q}_i, \quad \frac{\partial \mathcal{H}}{\partial q_i} = -\dot{p}_i = -\frac{\partial \mathcal{L}}{\partial q_i}$$

A final expression that we shall assert without proof is the Euler-Lagrange equation generalised to second order.

$$\frac{\partial \mathcal{L}}{\partial q_i} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_i} + \frac{d^2}{dt^2}\frac{\partial \mathcal{L}}{\partial \ddot{q}_i} = 0$$

Now, if you're sitting comfortably, we may begin. Suppose we have a Lagrangian of higher order (in one dimension, again without loss of generality), such that $\frac{\partial \mathcal{L}}{\partial \ddot{q}}$ depends on $\ddot{q}$. This necessarily implies that our coordinate is governed by a $4^{th}$ order differential equation, meaning it is a function of time, and of the initial conditions $[q_0, \dot{q}_0, \ddot{q}_0, \dddot{q}_0]$. The presence of 4 initial conditions implies that we can transform to a set of four *canonical coordinates.* These play a similar role in Hamiltonian dynamics as the generalised coordinates in Lagrangian mechanics. Where a Lagrangian is defined by the 2n coordinates $[q_n, \dot{q_n}]$, in Hamiltonian dynamics we use instead the 2n coordinates $[Q_n, P_n]$. We will choose the coordinates to be $Q_1 = q$, $Q_2 = \dot{q}$. Now, using Hamilton's first equation:

$$\frac{\partial \mathcal{H}}{\partial P_1} = \dot{Q}_1 = Q_2 \implies \mathcal{H} \sim P_1 Q_2$$

Similarly, for $P_2$;

$$\frac{\partial \mathcal{H}}{\partial P_2} = \dot{Q}_2 = \ddot{q} \implies \mathcal{H} \sim P_2 \ddot{q}$$

This enables to construct our Hamiltonian:

$$\begin{aligned}
\frac{\partial \mathcal{H}}{\partial Q_1} &= \frac{\partial \mathcal{L}}{\partial q} = -\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}} + \frac{d^2}{dt^2}\frac{\partial \mathcal{L}}{\partial \ddot{q}} = -\frac{d}{dt}P_1 && \implies P_1 = \frac{\partial \mathcal{L}}{\partial \dot{Q}_1} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \ddot{Q}_2} \\
\frac{\partial \mathcal{H}}{\partial Q_2} &= P_1 - \frac{\partial \mathcal{L}}{\partial \dot{q}} = -\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \ddot{q}} = -\frac{d}{dt}P_2 && \implies P_2 = \frac{\partial \mathcal{L}}{\partial \dot{Q}_2}
\end{aligned}$$

We now have expressions for all our canonical coordinates. The crucial final step is to notice now that $P_2$ depends on $\ddot{q}$, so the term featuring it in the Hamiltonian is quadratic with respect to $P_2$–that is, it has a well defined extremum. However, the other term does not satisfy this condition—it is *linear* in $P_1$. This means it can have no well defined minimum, which is a huge problem, considering we have identified the Hamiltonian as being closely related to the energy. This linear relationship leads to instabilities in the system, hence the name.

The implications of this are the real significance of the theorem. Instabilities tend to lead to runaway solutions that, while potentially mathematically sound, do not make good physical solutions. Hence, Ostrogradsky instability has been on several occasions put forward as an attempt to explain why is it so rare to find any differential equations of higher than second order in physics, within the classical limit.

# Mathematical Foundations of Classical Ballet

Nela Cicmil

Everyone knows that music and mathematics go well together. There are myriad examples of this, from Albert Einstein who loved to play the violin to John F. Nash who listened to Bach and Mozart while solving mathematical problems. It is no coincidence that Marcus du Sautoy and Douglas Hofstadter named their brilliant books *The Music of the Primes* and *Gödel, Escher, Bach*, respectively. I argue, however, that classical ballet is in fact *the* truly mathematical art - more so than even mathematicians or dancers might appreciate[1]. Ballet incorporates core principles from theoretical computer science, linear algebra, and geometrical aesthetics. When a ballet performance combines these mathematical fundamentals with graceful expression, evocative music, and sublime visual arts, it brings mathematics to life. Let me explain what I mean.

## Choreography as automata theory

Every classical ballet dance, such as the Dance of the Sugar Plum Fairy in the Nutcracker, or the Black Swan solo in Swan Lake, is constructed by stringing together individual components from the alphabet of ballet. These individual components are called the steps of the dance. The process of inventing new dances is called choreography. When ballet was invented in the Baroque court of King Louis XIV in 1665, in France, it was decreed that every dance step in ballet should begin and end in one of the so-called Positions of the Feet (Fig. 1)[2].

These positions form the basis of a 'ballet automaton' as defined by automata theory. In brief, a deterministic finite automaton (DFA) is an abstract machine that processes input strings, either 'accepting' them or not. The automaton has a finite number of *states*, a finite set of *input symbols*, and a *transition function* that specifies the next state that the automaton will move to, given its current state and an input symbol. The automaton also has a set of *final* or *accepting* states. An automaton therefore reads in a string of input symbols, moving from state to state as it processes each symbol in a manner determined by its transition function. If the automaton finds itself in an accepting state after the final symbol, it 'accepts' the string, otherwise it does not. In this way, automata can decide whether a given string is a member of some particular

[1] But see Wasilewska, K. (2012) Mathematics in the World of Dance. *Bridges 2012: Mathematics, Music, Art, Architecure, Culture*, 453:456.

[2] As ballet developed, further positions of the feet were used, including a sixth position, a seventh position, and various positions of standing on only one foot. However, the positions shown in Figure 1 remain the core positions.

language. As strings can code for logical expressions, graphs, integers, and much more, automata can decide answers to some interesting mathematical problems[3].

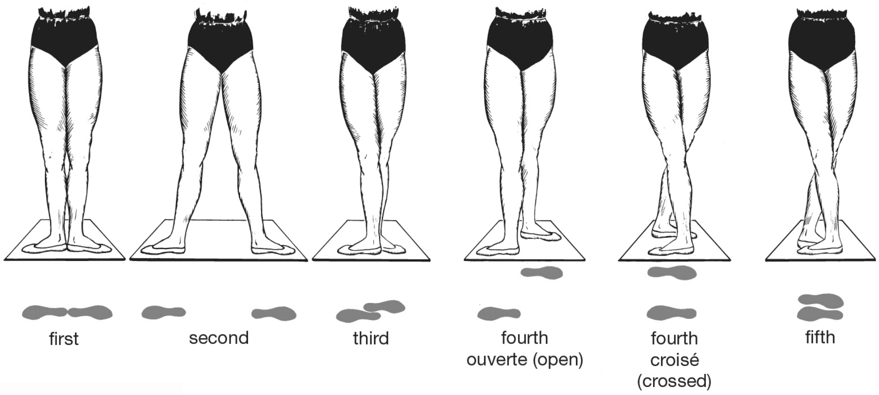


Figure 1: The basic positions of the feet in ballet. In fact, there are two versions of 'fourth position', open (*ouverte*) and closed (*croisé*). Generally speaking, a ballet step will begin and end in one of these positions.[4]

We observe that the process of choreographing a ballet dance is equivalent to finding an input string that is accepted by the ballet automaton[5]. Each of the positions of the feet are states of the automaton, and are also accepting states, since a dance can validly end in any of these positions. We define an additional 'impossible' state, which is not an accepting state, to represent a physically impossible choreographical request. Each ballet step is a symbol in the ballet alphabet. The transition function maps these symbols to a specific movement between the possible states (foot positions of the dancer). The process of choreographing a dance by sequencing steps together is then, at a first approximation, equivalent to creating an *input string* that this automaton will accept, which dictates a specific path to take through the states. If the input string is valid, it will move the automaton from state to state, that is, foot position to foot position, ending on one of the final positions. If the input string is invalid, by having two consecutive steps where the end position of the first step is not the same as the start position of the next step, then the automaton will move into the 'impossible' state, which is not a final state, and the transition function will ensure that the automaton remains in that state to the end of the string. In this case, the string will not be accepted. The set of strings, or *language*, accepted by the ballet automaton therefore defines the set of all possible valid

[3]For more information see the excellent textbook *Introduction to Automata Theory, Languages, and Computation* by John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman.

[4]Figure adapted from "ballet: five basic positions. Art. Britannica Online for Kids. Web. 12 Dec. 2015. `http://kids.britannica.com/comptons/art-167043`."

[5]For other ideas along similar lines, see Schaffer, K. Mathematics and the Ballet Barre. *Bridges 2011: Mathematics, Music, Art, Architecture, Culture*, 529:523, and LaViers, A. & Egerstedt, M. The Ballet Automaton: A Formal Model for Human Motion. *Proceedings of the 2011 American Control Conference*, 3837:3842.

ballet dances.

Of course, the above automaton is a simplification. It would be too complicated to describe an automaton that models the entirety of classical ballet, at least in the present article. But, we can define and consider a small example in the final section, below.

## 'En avant et en arrière': linear transformations and inverse functions

Another founding principle of ballet is the linear transformation of three-dimensional Euclidean vector space $\mathbb{R}^3$. First, most ballet steps can be translated in space, that is, they can be performed going forwards (*en avant*), backwards (*en arrière*), to the left, and to the right. For example, the step known as a *chassé*, a sliding step, can be performed in each of these four ways. Secondly, many steps involve rotation, either of the body or of the step itself. A wonderful example is the pirouette, where the dancer spins on one leg. Thirdly, many steps undergo scaling. For example, the basic *jeté* (thrown leap) comes in three sizes: a *petit jeté*, a small hop; a standard *jeté*, which is a medium-sized jump; and a *grand jeté*, a huge leap (Fig. 3a). To a mathematical mind, this organisation of ballet steps according to principles of linear transformation is very pleasing.

Many ballet steps are asymetric, or chiral. Every such step can be reflected along the main vertical axis of the body, that is, they can be performed both on the left side and on the right side. A wonderful example is the *attitude*, a held pose that can be performed on either the left leg or the right leg (Fig. 2). Ballet also makes much use of repetition: an out-and-out application of the identity function. Most interestingly, many ballet steps have an *inverse* - the step can literally be danced backwards from its finishing position to its starting position - an almost perfect example of an inverse function. This means that a dance made up only of these invertible steps can technically be performed *backwards* - equivalent to reversing the equivalent input string and thereby tracing backwards the path through our ballet automaton defined in the previous section! In fact, in ballet classes it is not uncommon for the teacher to set a short exercise, and after the students have performed it, to ask the students to perform it backwards, giving them little time to think. This speaks to the somewhat unrecognised mathematical ability of dancers. See the final section, below for a detailed example of such an inversion.

In the first section, we considered a theoretical ballet automaton that defines all possible ballet dances, that is, those strings of ballet steps that are physically possible to perform. However, not all of these possible ballet dances will necessarily be beautiful or even interesting to watch. Those input strings, or dances, that incorporate the elements discussed above, such as translations, reflections, repetitions, or inversions, tend to be used the most in classical ballet. For example, in the Black Swan solo in Swan Lake, the choreography makes much use of reflection (performing the same sequence to the left and to the right) and repetition (most famously in the 32 fouetté turns at the very end of the solo).

Figure 2: An *attitude* pose in ballet, performed here by Yamamoto Mashiko in the ballet *Le Corsaire*. *Attitude* can be performed either on the left or right foot, perfectly reflected.[6]

## The beauty of lines, angles and symmetry

Many mathematical objects, such as groups, geometric structures or topological spaces, are considered beautiful when we 'see' them in our mind's eye or in an illustration. For example, there is a fundamental, pleasing beauty to a perfect sphere or a tetradedron. Just as in mathematics, fundamental to the core beauty of ballet is the composition of shapes, angles and structures in visual space.


Figure 3: Caption: Illustrations of curves, symmetry, angles and extension in the ballets (a) *Don Quixote*, (b) *Swan Lake*, and (c) *Apollo*.[7]

Figure 3 illustrates some of these geometrically aesthetic ideas. In the *grand jeté* (Fig. 3a), the soft ellipse created by the arms stands in sharp contrast to the 180° angle created by the legs. The principle of symmetry, made possible by the fact that every

[6]Note: Image adapted from cc Fanny Schertzer `http://creativecommons.org/licenses/by-sa/2.5/`

[7]Note: Images adapted from (a) cc Peter Gerstbach, (b) cc Fanny Schertzer `http://creativecommons.org/licenses/by-sa/2.5/`, (c) Andrea Mohin/The New York Times.

ballet step can be reflected (as discussed in the previous section), is utilised to full effect by the *corps de ballet* in the First Act of *Swan Lake* (Fig. 3b). A obsession with repeated angles, in addition to symmetry, is seen in this iconic moment from the ballet *Apollo*, in which three female dancers create the illusion of multiple legs radiating out from a central origin (Fig. 3c). The radiating lines, generated by the concept of *extension* of the limbs in ballet, seem as though they could go on forever, evoking the impression of infinity. These are just a few examples of the general principle that classical ballet choreography often searches for moments of profound geometric perfection.

## Conclusion

This paper argued that ballet is based upon mathematical foundations, both at a constructional and aesthetic level. Choreographing and dancing ballet requires fast mathematical problem-solving, consciously or unconsciously. Moreover, recognising the computational and geometrical under-pinnings of ballet increases one's appreciation of any ballet performance. For these reasons, it would be natural for ballet to become the art of choice for mathematicians!

## A Deterministic finite automaton for "*petit allégro*"

We present a deterministic finite automaton (DFA) $A$ to exemplify *petit allégro* exercises, as may be found in ballet class. *Petit allégro* refers to the part of a ballet lesson or dance that involves small jumps. For simplicity, we limit our exercise to the steps *changement* and *entrechat*, and various *sautés* (jumps) between the foot positions (Table 1).

The DFA $A$ is the five-tuple $A = (Q, \Sigma, \delta, q_0, F)$, where $Q$ is the set of states, $\Sigma$ is the set of input symbols, $\delta$ is the transition function, $q_0$ is the start state, and $F$ is the set of final accepting states.

For our simplified automaton $A$, the set of states $Q$ is limited to first position (1st), second position (2nd), fifth position-right-foot-in-front (5thR), fifth position-left-foot-in-front (5thL) (Fig. 1), and the physically impossible state $I$. That is, $Q = \{$1st, 2nd, 5thR, 5thL, $I\}$. We designate 5thR as the start state $q_0$ as the majority of ballet class exercises start in this position. The set of input symbols $\Sigma$ consists of 6 *petit allégro* steps, specified in Table 1. Transition function $\delta$, which controls how each step moves the dancer from one state (foot position) to another, is specified in Table 2. Note that if the choreography asks for an impossible sequence of steps, that is, the end position of one step is not the start position for the following step, the automaton moves to non-accepting state $I$ and remains there for the rest of the input.

We define valid input strings, or valid choreographies for a dance exercise based upon these *petit allégro* steps, as those strings that would be accepted by $A$. An interesting valid exercise to consider is string1 = *ch-ch-S2-S1-S2-S2-S5L-ec*. This string, which starts at the start state 5thR and ends in the state 5thL, will be accepted by the automaton. Since the step *S5L* is chiral, we can also define the reflection of this exercise as string 2 = *ch-ch-S2-S1-S2-S2-S5R-ec*. By concatenating string1 and string2 to form new string3 = *ch-ch-S2-S1-S2-S2-S5L-ec-ch-ch-S2-S1-S2-S2-S5R-ec*, we can create a new exercise in which string1 leads straight into string2. We can do this because step *ch* is valid from the state 5thL, that is, it does not at any point send the automaton into the non-accepting state $I$. Since the exercise defined by string3 starts and ends on the same position (5thR), the exercise can be repeated consecutively as many times as you like - the only limit being the dancers collapsing from exhaustion!

Moreover, since the exercise encoded by string3 consists of steps that each have an inverse (see Table 1), the entire exercise can be danced backwards by stringing together the inverses of the steps in string 3, in inverse order, to form string4 = *ec-S2-S2-S1-S2-S5L-ch-ch-ec-S2-S2-S1-S2-S5R-ch-ch*. Necessarily, the inverse of a valid string, if it exists, will be valid. Both the original exercise (string3) and the inverse exercise (string4) can be traced in the diagram of automaton $A$ in Figure 5.

Automaton $A$ does not accept every possible string constructed from the set $\Sigma$. For example, a string that includes the following consecutive steps *...-S2-ec-...* will certainly not be valid: From any state, transition function $\delta$ sends automaton $A$ to state "2nd" when S2 is read in. But then when symbol *ec* is read in when $A$ is in state "2nd", $\delta$ defines that $A$ moves to state $I$. Since all input symbols read in state $I$ keep automaton $A$ in state $I$, which is not an accepting state, the string will not be accepted as valid. In the dance, this represents the situation in which the choreography requires the dancer to perform an *entrechat* after a *sauté* to second position - which is physically impossible since the *entrechat* can only be danced by the feet starting in a fifth position.

| symbol | step | description | inverse step |
|---|---|---|---|
| *ch* | *changement* | Jump up from, and land, in fifth position, having swopped which foot is in front. | *changement* |
| *ec* | *entrechat* | Jump in fifth position, crossing the legs twice in the air and landing in same foot position (Fig. 4). | *entrechat* |
| *S1*<br>*S2*<br>*S5R*<br>*S5L* | *sauté to first*<br>*sauté to second*<br>*sauté to fifth (right foot front)*<br>*sauté to fifth (left foot front)* | Jump from any foot position to land in the specified foot position. | A *sauté* back to the foot position from which the original *sauté* started. |

Table 1: Set $\Sigma$ of input symbols for automaton $A$, corresponding to certain steps of *petit allégro*, with explanations. Note that the *changement* is in fact a special case of the *sauté* to the opposite fifth position.

| | ***ch*** | ***ec*** | ***S1*** | ***S2*** | ***S5R*** | ***S5L*** |
|---|---|---|---|---|---|---|
| **1st** | *I* | *I* | 1st | 2nd | 5thR | 5thL |
| **2nd** | *I* | *I* | 1st | 2nd | 5thR | 5thL |
| **5thR** | 5thL | 5thR | 1st | 2nd | 5thR | 5thL |
| **5thL** | 5thR | 5thL | 1st | 2nd | 5thR | 5thL |
| ***I*** | *I* | *I* | *I* | *I* | *I* | *I* |

Table 2: Table of transition function $\delta$.

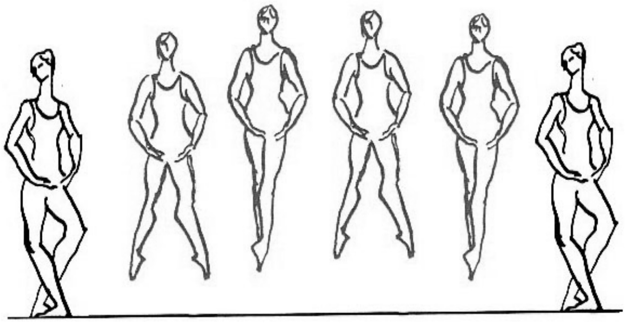

Figure 4: The *entrechat* steps begins and ends in the same fifth position (in this case, fifth position with right foot in front). While in the air, the legs are crossed with the right foot behind, and then crossed again to return the right foot to the front.[8]

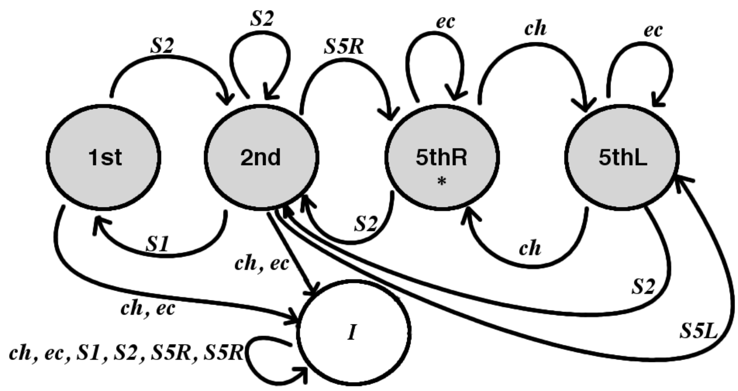


Figure 5: Diagram of automaton $A$, showing the five states (foot positions) and the ways the different symbols (steps) in Table 1 move the dancer between these states. For clarity, only the transitions used in string3 are depicted (see main text).

[8]Image adapted from Ortia at Wikimedia Commons.

We are a magazine entirely created by the members of the University. If you have an interesting mathematical idea that you would like to share with other Mathematicians, Natural Scientists and mathematical hobbyists, we would love to work with you and help you publish it in print on the pages of the "Invariant". For more information, contact the editor at
`magazine@invariants.org.uk`

The Invariants Society is Oxford University's student society for Mathematics. We're here to promote Maths and we host weekly informal lectures often given by leading mathematicians. To find out more, please see our website
`http://www.invariants.org.uk/` or find us on Facebook at
`https://www.facebook.com/oxford.invariants/`

$$\sqrt[\omega^{\pi}]{i\text{nvariants}}$$