THE SPATIAL INFINITY OF THE UNIVERSE: THE NEGLECTED PROBLEM OF COSMOLOGY*

Rafael Andres Alemañ Berenguer  
Department of Physics, Systems Engineering and Signal Theory  
University of Alicante (Spain)  
rafael.aleman@ua.es

Abstract: The finitude or infinity of the universe divided the ancient philosophers fueling a debate intertwined with the subtleties of the very concept of infinity and the plausibility of its realization in the physical world. While the 19th century took the first steps in the formal domain of mathematical infinity, the question of the size of the cosmos remained open pending better empirical evidence. At the end of the 20th century, the observational data of cosmology seemed to favour an infinite volume, although the delicate physical and metaphysical problems that such an option implies have rarely been highlighted.

Keywords: Infinity, actual, potential, space, cosmology, geometry, topology

1. Introduction

The possibility of infinite quantities in nature has inspired some of the most intense scientific and philosophical debates, especially about the extent of the universe. It is well known that Aristotle distinguished between potential infinity, which is conceived through a process capable of generating unlimitedly increasing magnitudes, and the actual, which is given entirely in itself at once [Oppy (2006)]. Recall the Stagirite’s reply to Zeno's famous aporias against movement, in which he remarked to the Eleata that although space and time are infinitely divisible potentially, they cannot be infinitely divided actually. Aristotle also denied the existence of infinitely large objects, unlike Anaxagoras and Epicurus, who did accept it. For his part, Architas of Tarentum starred in one of the most famous arguments in defense of spatial infinity, imagining himself in the sky of the fixed stars, where it is always possible to stretch out your hand or stick and thus demonstrate that there is no limit to the places you could occupy. This very intuitive argument was repeated with various nuances by Eudemus, Cicero, Lucretius, Giordano Bruno, Hery More, and John Locke. Since the medieval scholastics reserved infinity for the power of God, Galileo, and Descartes remained prudently ambiguous about it, chastened by Bruno's disastrous end.

Newton, however, found no impediments to reconcile divine omnipotence with the infinity of space, although he was unable to explain the gravitational stability of a cosmos thus constituted. The English genius did not succeed because the mathematical concepts necessary for the handling of infinity were only fully developed many years after his death. In fact, it was in the second half of the 19th century when the work of

* To appear in Teorema. Revista Internacional de Filosofía (Spain)
Cauchy and Weierstrass would definitively establish the consistency of the infinitesimal calculus. At the same time, infinite sets would be brilliantly systematized by Georg Cantor, opening up a new and fertile field in mathematical research whose fruits continue to germinate to this day without ceasing to abound in new challenges.

2. Infinity in natural science

For the purposes of our discussion, it is convenient to separate the role played by infinite quantities in physical theories into two large families, namely, those that are adopted for merely instrumental simplifying purposes, and those that are provisionally admitted on account of a future and more developed theory of which our present formalisms would be just crude approximations. The first modality is evident in idealized cases such as the infinite plate condenser of elementary electrostatics, the free particle moving inertially in infinite space, or the plane waves without beginning or end.

A different matter is the one that involves infinite quantities related to elementary particles or space-time itself. In both cases they are usually associated with the so-called "singularities", certain regions in which our physical equations lose their validity and yield infinite results. This concept acquired a new perspective when in 1915 Albert Einstein formulated a new theory of gravitation, called general relativity. Einstein's theory conceives of gravity as an effect of the space-time curvature, whose geometry is pseudo-Riemannian. Such a curvature is given by the mass-energy content of the material systems located in a certain space-time region, as given by the famous gravitational equations of general relativity.

By studying the evolution of stars depending on the relationship between their mass and their radius, the Indian scientist Subrahmanyan Chandrasekhar warned of the possibility that in certain extreme cases gravitation would overcome any stabilizing force and the star would collapse under its own weight, shrinking to a point. [Chandrasekhar (1931a, 1931b, 1932); Penrose (1996)], the singularity. Eddington and Einstein publicly rejected this possibility, but in 1939 Oppenheimer and Volkoff presented a paper [Oppenheimer and Volkoff (1939)] in which, by using the Schwarzschild metric, they demonstrated the perfect theoretical validity of the hypothetical gravitational collapse suggested by Chandrasekhar.

The Schwarzschild metric described the space-time curvature around a non-rotating mass with spherical symmetry [Schwarzschild (1916), Wald (1984)]. In spherical coordinates it is written (eq. (2.1):

$$ds^2 = \left(1 - \frac{R_s}{r}\right)c^2dt^2 - \frac{dr^2}{\left(1 - \frac{R_s}{r}\right)} - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)$$

(2.1),

where \(r\) is the radial distance from the central body, considered at the origin of coordinates, \(\theta\) and \(\phi\) are the spherical polar angles, and \(R_s = 2GM/c^2\) is the so-called “Schwarzschild radius” (or “gravitational radius”). When \(r = R_s\), no physical signal can escape from the region \(r < R_s\) in order to influence events in the region \(r > R_s\). We have therefore an “event horizon”, the border where the condition \(r = R_s\) is met, and inside it nothing –not even light– can escape.

The concept of event horizon was introduced in 1958 when it was understood that in the Schwarzschild metric, the case \(r = R_s\) (“coordinate singularity”) could be solved with a coordinate change [Filkenstein (1958)]. But if the body was dense enough, it became impossible to avoid an overwhelming gravitational contraction until \(r = 0\), and
then the theory lost all its predictive power. In 1970 Stephen William Hawking and Roger Penrose proved that physical singularities are a characteristic of any acceptable solution of general relativistic equations, regardless of quantum considerations [Hawking and Penrose (1970)]. Furthermore, their topological reasoning concluded that the formation of the singularity implied at least one point of discontinuity in the fabric of space-time, although there is always an event horizon hiding the singularities formed in the future of a surface of regular initial data [Penrose (1969)]. Perhaps the incorporation of quantum effects [Mottola and Mazur (2004)] prevents the appearance of genuine singularities, nevertheless solid conclusions are still lacking in this regard.

The singularities associated with particle physics are also related to null volumes. Usually the elementary character of the particles is conceived by identifying them with dimensionless points. This creates daunting problems in both classically and quantum regimes. In classical physics, an electron, for example, should generate an infinite electric field around it if its size is zero, since the intensity of the field depends on the inverse of the squared distance. Quantum physics aggravates the situation by maintaining that around each elementary particle there is always a cloud of other virtual particles –creating and annihilating one another in perpetual mutual interaction– whose total energy, when rigorously calculated, also yields an infinite result, usually called “divergence” [Weinberg (1979)].

The problems derived from virtual particles become even worse, since the number of virtual particles turns out to be itself divergent, that is, it tends to infinity by itself. The solution chosen to date has been to assume that, if the accompanying cloud of virtual particles has an infinite total energy, the proper mass of the electron –or any other particle– must be an infinite amount of opposite sign, so that both quantities cancel almost exactly and cause us to measure experimentally the mass of the electron that we actually did measure. This technique is called “renormalization” and well-founded suspicions continue to hover on its legitimacy, both formal and conceptual, only cushioned by the indisputable empirical success of its predictions.

3. The cosmological scenario

There is, however, a branch of natural sciences in which the existence of the actual infinite is admitted: such is the case of modern cosmology. Einstein was the first to apply his gravitational theory to the universe as a whole, an endeavour that began in 1917 by assuming for simplicity a homogeneous, isotropic and static cosmos [Einstein (1917)]. For this purpose, he had to introduce a new term in his equations, the cosmological constant, whose effect would only manifest itself on a large scale in the universe to compensate for the universally attractive character of ordinary gravitation and thus prevent the collapse of all matter in the universe. The German scientist did not take long to notice that his model would be radically unstable, which prompted the Dutchman William De Sitter, also in 1917, to propose another cosmological model in which the universe would be practically empty and expanding [Weinberg (1971)]. As a consequence, it was demonstrated that general relativity admitted a dynamic vacuum solution.

This opened the door for others to consider the variety of possible solutions to general relativity equations, everyone representing a different type of universe. This is what Alexander Friedman did by assimilating the material content of the universe to a dynamic, homogeneous and isotropic gas. In this way he obtained cosmological models in which the universe could be expanding or contracting depending on the initial conditions [Misner et al (1973)].
The further we go back in time, the more the cosmic volume decreases, while the value of certain physical magnitudes—such as temperature or density—grows without limit, tending to infinity. And, in fact, in the singularity these properties would become supposedly infinite. But if we thoughtfully review the process in the ordinary chronological order, it seems that we are starting from a bunch of infinities that gradually reduce until they reach their current, obviously finite, values over time. Then it is worth asking when these properties stopped being infinite, if this question itself makes any sense.

The answer is that at no time do temperature, density, or any other physical property acquire infinite values, which is why there is no instant in which they cease to be so to become finite. In the case of the series of instants that converge—they are as close as we want—to the origin of the universe (that is, $t_0 = 0$), admitting as usual the continuity of time and space, it turns out that there is no instant immediately prior to $t_0$, since between every two instants there are always an infinity of them, as it is typical for a continuous magnitude. Every instant subsequent to $t_0$, regardless how small it may be, has a finite value of the temperature, no matter how large it may be.

In $t_0$ we would not have infinite values because, generally speaking, the limit of a series is not necessarily one of the terms that constitute the series. Only in some specific cases the limit value of a convergent series belongs to the set of terms that form the series. In our case, $t_0$ is not part of the decreasing series of time instants that converge at the origin of the universe. Surprisingly, this cosmological approach allows us to conclude that, even without going back to an infinite past, the cosmos did not have a beginning in time, if by "beginning" we understand the instant $t_0$.

Perhaps Friedman's most lasting legacy was the metric that bears his name (eq. 3.1), today expanded to Friedmann-Lemaître-Robertson-Walker (FLRW). It is an exact solution of the relativistic equations of gravity involving the curvature $K$ and the scale factor, $a(t)$, which expresses the time dependence of the spatial component of the metric. In spherical coordinates:

$$ds^2 = c^2 dt^2 - a^2(t) \left[ \frac{dr^2}{1-Kr^2} + r^2 \left( d\theta^2 + \sin^2 \theta d\varphi^2 \right) \right]$$

Likewise, there happens to be of a high importance an equation (eq. 3.2) that expresses the spatial curvature of the universe $K$ as a function of the scale factor $a(t)$, the average cosmic density $\rho$ and the critical density $\rho_c$, in whose definition is included the gravitational constant $G$ and the well-known Hubble parameter $H$ (wrongly called “constant”, since it varies with time), which relates the rate of distance between two points in the universe with their mutual separation.

$$K = \frac{1}{c^2} \left( \frac{da}{dt} \right)^2 \left( \frac{\rho}{\rho_c} - 1 \right)$$

The mentioned critical density acts as a threshold value to discuss the possible spatial geometries of the universe and its evolution over time. When $\rho > \rho_c$, the curvature is positive, the geometric properties of the universe resemble those of a sphere, and the spatial volume of the cosmos has a finite value. On the contrary, if $\rho < \rho_c$, the curvature is negative, its geometry corresponds to the hyperbolic type (as in the surface of a horse saddle) and the spatial volume is infinite. Finally, under the condition, $\rho = \rho_c$, the
geometric characteristics of the cosmos would be given by the typical Euclidean geometry of the ordinary flat plane.

The relationship between density and its critical value is so important that it is usually expressed as a quotient with its own symbol, \( \Omega = \rho/\rho_c \). The case \( \Omega > 1 \) implies a closed cosmos with a beginning in time (the so-called Big Bang) and a sufficient material stuff to stop the expansion at some future instant depending on the specific value of \( \rho_c \). From that moment the expansion will become a contraction and the universe will end in a total gravitational collapse (Big Crunch). But if \( \Omega < 1 \) there is not enough stuff to reverse the process and the cosmological expansion will go on forever.

This very direct link between density, geometry and evolution over time is broken when we introduce the cosmological constant into our considerations. This new magnitude contributes to the total curvature of the universe in a way opposite to that of ordinary matter, that is, as if it were a repulsive force (antigravity). In fact, current estimates suggest that the cosmological constant would be responsible for an accelerated expansion of the universe with zero or, perhaps, very slightly negative curvature [Perlmutter et al. (1998), Goldhaber (2009)].

4. The spatial volume problem

The possible spatial infinity of the universe had been a real headache even since Newton offered the first scientifically sound formulation of gravity as the dominant force on a cosmic scale. At the end of the 19th century the German astronomer Hugo von Seeliger exposed the well-known paradox that bears his name [Seeliger (1895)]. Briefly expressed, it tells us that, in an infinite and static cosmos with a uniform distribution of matter, it follows that the Newtonian gravitational potential is undefined and, therefore, the total gravitational force on any object in the universe is also undefined. The controversy, which lost its strength when general relativity entered the stage, never failed to cast serious doubt on the coherence of classical Newtonian cosmology [Norton (1999); Vickers (2008)].

The advent of Einstein's theory of gravity altered the terms of the debate, but the problem of infinity resurfaced in another guise. When we ask ourselves if the volume of the universe is infinite, we must remember that in general relativity any division of space-time in a space volume—or hypersurface—and a dimension of time is equally valid (as long as the condition that no pair of points on the spatial hypersurface can causally influence each other is obeyed). Therefore, we must first clarify how we define the spatial portion of the universe whose size we wish to calibrate.

The matter becomes more complicated when we include proposals as exotic as the multiple variants of inflationary models (eternal, self-generated, fractal, etc.). In such an intricate framework, it hardly makes sense to wonder about spatial infinity when the very notion of the universe is diluted in a profusion of bubble-universes that come from other bubble-universes and in turn engender their own offspring in a process without beginning or end (as its followers say). Even so, it is difficult to understand how the problem could be solved by admitting that from one of these bubble-universes of infinite size another, also infinite, could break off in a process repeated an infinity of times [Albrecht (2004), Steinhardt and Turok (2007), Linde (2008)].

For all these reasons, from now on we will limit our discussion to the traditional Friedman models, as mentioned in the preceding section. Unlike closed cosmological models—whose positive curvature is represented by the surface of a sphere—the case of an open universe (negative or null curvature) is more difficult to visualize. And it is noteworthy that the inconveniences of attributing an infinite extension to the universe are
not usually discussed. For instance, defining the density of matter or the global temperature over an infinite spatial volume is a particularly delicate matter.

With an infinite spatial extent, regression in time would not imply the reduction of the three-dimensional volume until it collapsed to a point. In such a situation we would progressively approach a curious state in which, preserving spatial infinity, the matter density would grow as we get closer to the initial instant, tending to infinity at all points in space. Whether such a state has any physical meaning is quite another matter.

Obviously, the limit of this retrospective sequence would be found again at a null value for time, a state in which the density would be infinite at all points of an infinite spatial volume. Unlike the case of positive curvature, however, we cannot now say that infinite values of physical quantities correlate with a zero-tending decrease in space-time volume. In the previous case, this procedure had allowed us to consider that these infinities corresponded to a limit that did not belong to the series itself, expelling them so to speak of space-time. But that possibility does not exist when we are forced to maintain an always infinite spatial volume, given the close imbrication between space and time required by relativity, and nothing guarantees that such a state has some kind of legitimate physical interpretation.

Not less important is the unnoticed inconsistency that creeps in when we try to fix the physical meaning of an infinite spatial extension where all its points happen to be singular. The difficulty will become more apparent if we look at what happens when we go back in time with a magnitude like, say, the density of matter. An infinite cosmos would contain an equally infinite number of particles distributed more or less uniformly within it. That quantity of material particles would correspond to a countable infinity, unlike the points in space if accepting the usual assumption of continuity which form an infinite uncountable set. It seems difficult, then, to explain how we could have an initial state in which the density was infinite at all points of an infinite spatial volume (uncountable set) if we started our countdown from a countably infinite quantity of material particles.

Looking to the future, instead of going back to the origin, the inconvenience does not disappear, especially if we intend to compose a model like the one offered by Roger Penrose in his conformal cyclical cosmology. To face the fact that all available data rules out that the universe stops expanding and implodes, Penrose suggests that the end of one stage (with a supposedly infinite spatial volume) is connected to the beginning of the next (with a volume tending to zero) through the convenient re-scaling associated with a conformal transformation [Penrose (2006)]. Leaving aside other obstacles such as, for example, the need to respect the conservation of electric charge the formal legitimacy that mathematical coherence grants does not in itself grant the necessary physical plausibility that cosmological models aspire to possess. At the moment, no process is known that can give physical meaning to the infinite re-scaling of distances defended by Penrose in his proposal, without forgetting that the empirical evidence collected for the time being does not support it either.

5. Proposals to overcome the issue

In addition to geometry, to deal with this problem we have to count on topology, a branch of mathematics that deals with the properties that remain invariant under deformations that do not involve adding or removing points from the object (or, in general, from a variety) in question. Whether the size of the cosmos is finite or infinite, it may depend on its topological properties, which, in turn, are linked to some extent to the geometry of space-time. Stated very succinctly and without excessive loss of rigor, the topology of a manifold is defined by the type of n-dimensional figure, or “fundamental
domain”, whose repetition—as in a tile—can cover the entire space considered [Lachieze-Reya and Luminet (1995), Flapan (2010)]. When the curvature is not zero, the radius of curvature sets a natural scale of distances for the minimum size of the fundamental domain.

A model of the universe with positive curvature necessarily implies for any instant a finite volume to which we can associate many—infinite—simple or multiply connected topologies. Models with null or negative curvature enrich the possibilities, since in both cases the volume of the cosmos they represent can be finite or infinite, depending on the topology chosen for them. And between these two options, the hyperbolic cosmos—with negative curvature—seems to be the most promising because, in addition to the freedom to assign it a finite or infinite volume, the fact that its radius of curvature is finite makes it possible to establish a harmonious link between its geometry and its topology.

Furthermore, according to the pioneering work of the American mathematician William Thurston, almost all three-dimensional manifolds admit a metric with negative curvature, as in hyperbolic spaces [Thurston (1988, 1997)]. The special attractiveness of this class of models increases thanks to the rigidity theorem, in virtue of which geometric magnitudes as characteristic as volume, length or geodesic trajectories, turn out to be topological invariants.

This leaves us at a triple crossroads:

(A) the 3-dimensional hypersurface of our universe has an infinite spatial volume, and what we call the “observable universe” is only an infinitesimal portion of its extension;

(B) the spatial volume is finite due to peculiar topological connectivity, but the fundamental domain is larger than the observable universe;

(C) the same case as in (B), although the fundamental domain of the topology is smaller than the observable universe and, therefore, we could detect its characteristic features.

Option (A), although not impossible, is as untestable as any hypothesis involving the measurement of an infinite quantity. Depending on the size of the fundamental domain—that basic figure capable of “tiling” the entire space—case (B) could be confused with (A), unless there were some topologically significant inhomogeneity manifesting itself in the observable universe. And case (C) would be desirable for any researcher with the appropriate technical means.

The empirical procedure to decide between these possibilities is usually directed towards the search for specific patterns or regularities in two areas: one is the microwave background radiation—which, as far as we know, permeates all outer space—and the other one concerns the distribution of measurable radiation sources (pulsars, quasars, supernovae, active galaxies, etc.). Currently we have an appreciable amount of data provided by the Hubble (1990) and Herschel (2009) space telescopes, as well as the COBE (1989), WMAP (2001) and PLANCK (2009) space probes, among many others.

Based on such evidence, with almost absolute certainty we have to rule out a closed cosmological model with positive curvature. The evidence points rather towards a flat universe or perhaps with a slight negative curvature [O’Raifeartaigh et al. (2018)]. There are also no traces of multiple topological connectivity which would allow a finite spatial volume to be associated with one of these two cosmic geometries. However, it is true that a fundamental domain larger than the observable universe—option (B) mentioned above—would be compatible with the cumulative data.
To increase the complexity of this subject, in the first two decades of the 21st century a theorem was proven that excluded the possibility of knowing the global structure of space-time even assuming that the physical laws valid in our environment also rule any other space-time region [Manchak (2009, 2020), Smeenk, and Wuthrich (2020)]. It seems that we are dealing with an inevitable underdetermination of reality by the theory that describes it: general relativity provides us with too many possibilities, in principle empirically equivalent, for the global shape of the universe. However, it is not known for sure if this overabundance of observationally indistinguishable models would be solved by including the presence of matter in the aforementioned theorem.

The publication in June 2021 of a new study on the possibility of attributing a finite volume to the universe fuelled interest in the cosmological implications of topology [Aurich et al (2021)]. A joint Franco-German team from the Universities of Ulm and Lyon compared the observable distribution of background radiation perturbations at different scales with the results of a computer simulation designed for a universe whose topology constrained its three-dimensional volume. The observational data appeared to fit the finite volume computer simulation better than the usual model with a simple topology and infinite volume. If this is confirmed in future work in this regard, it could be assumed that the cosmos has the topological properties of a kind of three-dimensional doughnut (toroidal shape), with a volume about three or four times greater than that of the visible universe, whose radius is about 45,000 million light-years. It's too early to tell, but perhaps an emerging gravitational-wave astronomy will allow us to explore deep space and discover some characteristic patterns for the topology of our universe.

6. Conclusions

The most pertinent way of dealing with infinity has been a source of confusion in science and philosophy from the very origins of these disciplines. Mathematicians, as so many other times, demonstrated sufficient ability to establish themselves as pioneers in the task of mastering the formal management of infinite sets, although their work, more than a century and a half later, is far from finished. However, the elegance and depth of the mathematical discoveries about infinite quantities left their possible existence in the natural world unclear. That was a question that could only be answered empirically and, in the absence of data to settle the query, this problem remained to be resolved.

The development of modern science—so-called from Galileo and Newton onwards—had to deal with the same ambiguities as the ancient philosophers due to a lack of formal tools with which to bridle the concept of infinity. This was, other than being identified with divine omnipotence, associated with ideal models very useful for the simplification of calculations. Infinity was as well taken as a distinctive sign of the limits of applicability of the theory applied (and perhaps also as a warning of a future theory that would replace it). In other words, and using a more Aristotelian language, for theoretical reasons in the natural sciences it was accepted—and taken advantage from—potential infinity with the same serenity with which actual infinity was rejected on empirical grounds.

Modern cosmology, that started in 1917 with the first applications of general relativity to the universe, came to completely change such a quiet landscape. The cosmic geometry went forth, opening up three possible scenarios in two of which the spatial volume of the universe appeared to be necessarily infinite. When astronomical observations ruled out the possibility of a cosmos with positive curvature, at the end of the 20th century, the remaining options invited us to think of an infinite extension. This circumstance contained not a few inconveniences, both theoretical (What sense does the
actual infinity make in the physical world?) and empirical (How could the value of a magnitude that by definition is incommensurable be calibrated?),

However, the difficulties implied by open cosmological models refused to go away. Spatial infinity would entail an infinite amount of uniformly distributed matter, otherwise we would have an infinite amount of absolutely empty space, something with very little physical meaning. But then the problem of temporal evolution would reappear: going back to the initial moment of the Big Bang, the type of countable infinity associated with the amount of matter would not correspond to the non-countable infinity that characterizes the continuity of space.

The only reasonable way out seems to direct us towards a cosmological model with null or slightly negative curvature, as indicated by the observations, which at the same time presents a finite spatial volume to avoid the aforementioned problems. This could only be achieved, in the current framework, either by relaxing some of the conditions that lead to the FLRW family of geometries, such as homogeneity in the distribution of matter (a requirement, on the other hand, not infrequently discussed), or by resorting to considerations that are not merely geometric, such as topological properties. These new research directions take us into a vast territory of possibilities, the vast majority of which we can barely glimpse today.

Be that as it may, in the cosmological field, theoretical work will have to wait for new and more precise observational data to guide its path. Meanwhile, the question of the actual infinity, clearly expressed in the volume of our universe, will remain open and no one knows how long it will be unsolved or what future consequences for our image of the cosmos the solution will entail.

BIBLIOGRAPHIC REFERENCES


