

A Robust Algorithm to Determine the Topology of Space from the Cosmic Microwave Background Radiation

JEFFREY R. WEEKS (*)

dedicated to the memory of Marco Reni

SUMMARY. - *Satellite measurements of the cosmic microwave background radiation will soon provide an opportunity to test whether the universe is multiply connected. This paper presents a new algorithm for deducing the topology of the universe from the microwave background data. Unlike an older algorithm, the new algorithm gives the curvature of space and the radius of the last scattering surface as outputs, rather than requiring them as inputs. The new algorithm is also more tolerant of errors in the observational data.*

1. Introduction

Since ancient times humans have speculated on whether the universe is finite or infinite. Upcoming satellite measurements of the cosmic microwave background radiation will finally provide an opportunity to test whether the universe is multiply connected. The central idea is Cornish, Spergel and Starkman's *circles in the sky* method. The

(*) Author's address: 15 Farmer Street, Canton NY 13617, USA, e-mail: weeks@northnet.org

AMS Subject Classification: 85A40.

Keywords and phrases: cosmic topology, microwave background radiation.

following paragraph introduces the basic idea. In the literature the reader will find more complete elementary [6], intermediate [2], and advanced [1] expositions.

The cosmic microwave background (CMB) radiation is the radiation remaining from the big bang. It fills space uniformly, travelling in all directions. Furthermore, all CMB photons have been travelling for approximately the same length of time (since the so-called *decoupling*, when the primordial plasma condensed to a neutral gas roughly 300000 years after the big bang), and all at the same speed (the speed of light), so therefore all CMB photons have travelled approximately the same distance. This implies that the CMB photons now arriving on Earth began their journey on the surface of a huge geocentric sphere, the *last scattering surface* (LSS). When we observe CMB photons, we are literally looking back in time and seeing the LSS as it was at the time of decoupling. The early universe was very homogeneous, but there were small density variations on the order of a few parts in 10^5 . CMB photons coming from slightly denser regions in the early universe have done a little extra work against gravity, and therefore arrive slightly cooler. Conversely, CMB photons arriving from less dense regions have done less work against gravity, and arrive slightly warmer. These temperature variations allow us to “see” density variations in the early universe. If the universe is multiply connected and sufficiently small, then the LSS reaches all the way around the universe and overlaps itself. The near isotropy of the CMB implies that the observable universe has constant curvature to about 1 part in 10^4 [7], so the overlap (the self-intersection set of the LSS) is a circle. Subjectively we see what appear to be two different circles in two different parts of the sky, but really they are the same circle of points in space. Thus the two circles in the sky will (ignoring various sources of noise) display the same pattern of temperature variations.

Observationally, one begins with a temperature map of the CMB and searches for matching circles. If matching circles are found, the next task is to use them to deduce the topology of space. A first solution to this problem is to use the matching circles to construct a Dirichlet domain [9]. Thinking in the universal cover, it is clear that the plane of each matched circle lies exactly half way between one

preimage of the observer and a nearby preimage of the same observer. The face of a Dirichlet domain also lies exactly half way between one image of the basepoint and a nearby image. So, roughly speaking, the planes of the matched circles determine the faces of the Dirichlet domain. Unfortunately this simple algorithm has two weaknesses.

Weakness 1. The algorithm requires as input the geometry of space (spherical, flat, or hyperbolic) and the radius of the last scattering surface.

Weakness 2. The algorithm is insufficiently robust in the face of the three kinds of errors that real data will surely include:

Error A. Missing circle pairs. Microwave emissions from our own Milky Way galaxy contaminate the CMB in the plane of the galactic equator, so for this reason alone some pairs of matching circles are likely to be missed.

Error B. Falsely matched circle pairs. Some unrelated circles are likely to have similar temperature patterns just by chance.

Error C. Noise. CMB temperature fluctuations reflect not only density variations in the early universe, but also the doppler shift due to plasma motion and other sources of noise. Thus even the correctly matched circles will suffer some error in their sizes and locations.

All three types of errors will occur at a much higher rate than was believed at the time [1] and [9] went to press.

A new algorithm, described in detail in the remainder of this article, avoids both these weaknesses. The new algorithm requires as input only the set of matched circles; it produces as output the geometry of space and the radius of the last scattering surface, as well as the topology of space. The new algorithm tolerates much greater errors because it makes use of the group structure. It is especially effective in the spherical and hyperbolic cases, where the group of covering transformations is completely rigid; if the original data is good enough to determine the combinatorics of the group, then the algorithm can refine the data to produce an exact discrete group. Testing has shown that the new algorithm can handle large

numbers of missing and/or false matches along with moderate noise, and it can also handle large amounts of noise along with moderate numbers of missing and/or false matches. It fails only in the case that the missing/false matches and the noise are both severe, because in these cases there is not enough information available to recognize the combinatorics of the group of covering transformations.

The balloon-based CMB observations of the BOOMERanG [4] and MAXIMA [5] projects have found that the observable universe is approximately flat. A determination of the topology of space could provide independent confirmation of this result. Alternatively, in the unlikely event that the upcoming satellite observations fail to resolve the so-called “missing second peak” in the balloon data, a determination of a spherical or hyperbolic topology would establish that space is curved.

I thank Neil Cornish for providing sets of realistically noisy data with which the new algorithm was tested, and for his excellent advice at all stages of this project. Source code for the algorithm is available upon request.

2. Conventions

2.1. Curvature units

Throughout this article the radius R_{LSS} of the surface of last scattering is expressed in curvature units, that is, $R_{LSS} = r_{LSS}/r_{curvature}$, where r_{LSS} is the radius of the last scattering surface in meters and $r_{curvature}$ is the radius of curvature of the universe in meters. In spherical geometry, curvature units are the same as radians. In hyperbolic geometry, curvature units are the standard hyperbolic units, which may also be thought of as radians in the Minkowski space model of hyperbolic 3-space. Euclidean geometry is exceptional: it has no radius of curvature and R_{LSS} is arbitrarily set to 1.

2.2. Matrices

When a matrix acts on a vector, mathematicians write the product as $(matrix)(columnvector)$, while many computer graphics professionals write $(rowvector)(matrix)$. The two matrices are, of course,

the transpose of each other. Here we follow the mathematicians' convention.

Isometries of spherical, Euclidean, and hyperbolic space are handled in a uniform way by 4×4 matrices. An isometry of the 3-sphere is represented by an orthogonal matrix in $O(4)$. Similarly, an isometry of hyperbolic space is represented by a matrix in $O(1, 3)$, with the convention that the first coordinate is the "timelike" one. Euclidean 3-space is modelled as the set of points (w, x, y, z) in R^4 satisfying $w = 1$. The group of isometries of Euclidean 3-space (including translations) is thus a subgroup of $GL(4)$. In all three geometries the observer is at the basepoint $(1, 0, 0, 0)$.

3. Algorithm

This section explains the new algorithm. Implementation details are documented in the source code.

Input: A set of matched circles. Typically the input data will suffer from the three types of errors listed in Section 1.

Outputs: (1) the geometry of space (spherical, flat, or hyperbolic), (2) the radius of the last scattering surface in units of the curvature radius (if space is not flat), and (3) the topology of space, expressed both as a Dirichlet domain with face identifications and as a triangulation.

In the flat case, the topology may be recognized as one of the ten closed flat 3-manifolds by computing its orientability and first homology group. In the spherical and hyperbolic cases the volume also helps to recognize the topology as that of a known manifold. In the spherical case we compare to the known classification of closed spherical manifolds. In the hyperbolic case we use the computer program SnapPea [8] to compare to a data base of known low-volume manifolds.

Step 1. Determine the curvature of space. Assume for a moment that we know both the geometry of space and the radius R_{LSS} of the last scattering surface. It is then a straightforward matter to convert each pair of matched circles C, C' to a covering transformation: imagine one copy of the LSS centered at the origin, and then

position a second copy so that circle C on the second copy coincides with circle C' on the first copy. This construction gives us two neighboring lifts of the LSS in the universal cover, and thus the isometry taking one to the other is a covering transformation. In practice one writes the covering transformation as a matrix.

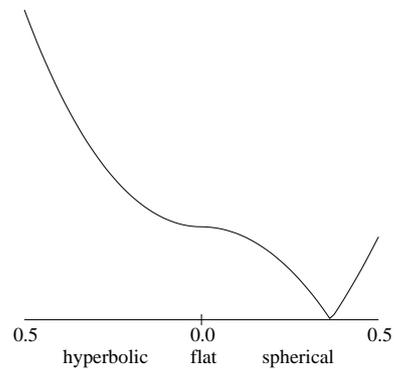
The question, then, is how do we know the geometry of space and the radius R_{LSS} ? The answer, in brief, is that we choose the values for which the resulting covering transformations form a discrete group. That is, for each choice of geometry and R_{LSS} we transform the set of matched circles to a set $\{g_i\}$ of covering transformations and compute all possible triple products $g_i g_j g_k$. If we have chosen the geometry and R_{LSS} correctly, many of the triple products will evaluate to the identity, because they represent relations in the group. Figure 1 shows how the error in the best-satisfied relation varies as a function of the geometry and R_{LSS} for perfect data. The introduction of realistic levels of noise raises the minimum only slightly, and also slightly broadens the trough because different (noisy) relations achieve their minima at slightly different values of R_{LSS} , but the overall picture remains the same. For safety, the implementation of this algorithm records the error in the five best relations, to guard against the possibility that a single relation might be approximately satisfied by chance.

The value of R_{LSS} selected in this way is only a tentative value. Step 5 of the algorithm will provide a more precise value.

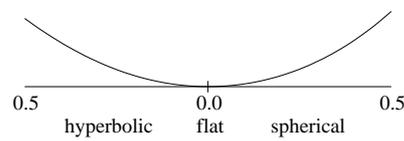
If the number of matched circles is very small, we can, at this point, augment it by adding in products of pairs of elements. Typically, though, this is not necessary and is not done.

Step 2. Recognize the discrete group.

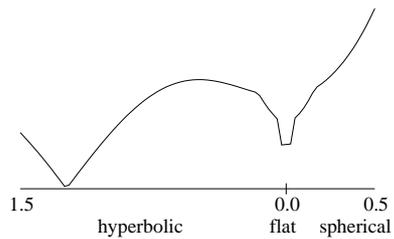
Given the geometry of space and a reasonable estimate for R_{LSS} , the next task is to recognize the structure of the group of covering transformations. As explained in Step 1, each pair of matched circles determines a covering transformation g_i and its inverse. We compute each triple product $g_i g_j g_k$ and compare it to the identity to decide whether it represents a group relation. There is no need to consider 2-element relations $g_i g_j = 1$ because the algorithm already stores each matrix along with its inverse, and there is no need to consider



(a)



(b)



(c)

Figure 1: The precision with which the group relations are satisfied varies with the choice of geometry and the assumed value for the radius R_{LSS} of the last scattering surface. These graphs plot the error in the best-satisfied relation as a function of R_{LSS} for (a) a spherical manifold (the Poincaré dodecahedral space), (b) a flat manifold (the 3-torus), and (c) a hyperbolic manifold (the manifold of volume 0.94...). The point at which the error is minimized tells us the correct geometry and provides a good estimate of the true value of R_{LSS} .

relations with 4 or more factors because a relation such as $g_i g_j g_k g_l = 1$ is a consequence of two simpler relations $g_i g_j h = 1$ and $g_k g_l h^{-1} = 1$.

At this point we count how many relations each matrix pair $\{g_i, g_i^{-1}\}$ occurs in. Matrix pairs that do not participate in any relations are discarded. In particular, this eliminates matrix pairs that came from falsely matched circles, thereby correcting Error B of Section 1. If a matrix pair occurs in only one relation, both the matrix pair and the relation are discarded, because such a relation would serve only to define that one matrix pair as a product of other group elements. Only matrix pairs that occur in two or more relations are retained, because only they serve to “crystallize” the group structure.

The removal of useless relations may produce a cascading effect. That is, a relation may be valid on a first pass through the list (meaning that each of its three factors occurs in at least one other relation), but invalid on a second pass through the list (meaning that one of its factors no longer occurs in other relations, because the other relations were removed during the first pass). Thus the removal process is applied repeatedly, until no further progress is possible. Each surviving matrix pair occurs in at least two relations, and the matrices and relations are passed on to Step 3 for refinement. With small initial data sets it can happen that no matrix pairs and no relations remain, in which case we resort to the augmentation technique mentioned in the last paragraph of Step 1.

Step 3. Refine the discrete group.

At this point we have a set of matrix pairs, each of which occurs in at least two relations. The relations define the structure of the group of covering transformations. Indeed, in the spherical and hyperbolic cases the relations determine unique values for the matrices, up to conjugacy. Here conjugacy corresponds to a change of coordinate system or, more physically, to a change in the location of the observer.

Our goal is to compute new values for the matrix pairs, satisfying the relations *exactly*. The given initial values for the matrices, which satisfy the relations approximately, serve as starting points. The changes required to move from the approximate matrices to the exact matrices are small, so we may work with a linear approximation.

The matrices are elements of the Lie group $Isom(S^3) = O(4)$, $Isom(H^3) = O(1, 3)$, or $Isom(E^3)$ according to whether the manifold is spherical, hyperbolic, or flat. In each case the Lie group is 6-dimensional, because we have three degrees of freedom in where to move the basepoint, and three more degrees of freedom in how to orient it once it gets there. Throughout the remainder of this section please think of the 4×4 matrices as points in a 6-dimensional Lie group.

To make a small improvement to a matrix M , we premultiply it by a matrix dM which is close to the identity:

$$(improved\ M) = (dM)(old\ M)$$

Because the matrix dM is close to the identity, we may work with a linear approximation. For example, a translation in the x -direction in hyperbolic 3-space is given exactly as

$$\begin{matrix} \cosh d & \sinh d & 0 & 0 \\ \sinh d & \cosh d & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

but for small values of d is adequately represented by its linear approximation

$$\begin{matrix} 1 & d & 0 & 0 \\ d & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

Thus the full sets of generators for $Isom(S^3)$, $Isom(E^3)$, and $Isom(H^3)$ are

	<i>spherical</i>	<i>flat</i>	<i>hyperbolic</i>
<i>x translation</i>	1 $-d$ 0 0	1 0 0 0	1 d 0 0
	d 1 0 0	d 1 0 0	d 1 0 0
	0 0 1 0	0 0 1 0	0 0 1 0
	0 0 0 1	0 0 0 1	0 0 0 1
<i>y translation</i>	1 0 $-d$ 0	1 0 0 0	1 0 d 0
	0 1 0 0	0 1 0 0	0 1 0 0
	d 0 1 0	d 0 1 0	d 0 1 0
	0 0 0 1	0 0 0 1	0 0 0 1
<i>z translation</i>	1 0 0 $-d$	1 0 0 0	1 0 0 d
	0 1 0 0	0 1 0 0	0 1 0 0
	0 0 1 0	0 0 1 0	0 0 1 0
	d 0 0 1	d 0 0 1	d 0 0 1
<i>xy rotation</i>	1 0 0 0	1 0 0 0	1 0 0 0
	0 1 $-d$ 0	0 1 $-d$ 0	0 1 $-d$ 0
	0 d 1 0	0 d 1 0	0 d 1 0
	0 0 0 1	0 0 0 1	0 0 0 1
<i>yz rotation</i>	1 0 0 0	1 0 0 0	1 0 0 0
	0 1 0 0	0 1 0 0	0 1 0 0
	0 0 1 $-d$	0 0 1 $-d$	0 0 1 $-d$
	0 0 d 1	0 0 d 1	0 0 d 1
<i>zx rotation</i>	1 0 0 0	1 0 0 0	1 0 0 0
	0 1 0 d	0 1 0 d	0 1 0 d
	0 0 1 0	0 0 1 0	0 0 1 0
	0 $-d$ 0 1	0 $-d$ 0 1	0 $-d$ 0 1

To a linear approximation, the six matrices in each column commute with one another. Matrix multiplication preserves the 1's on the main diagonal, and acts as scalar addition off the main diagonal. Thus matrix multiplication becomes (commutative!) vector addition in R^6 . Instead of visualizing the matrix dM as a 4×4 array, we will instead think of it as a vector

$$(d_0, d_1, d_2, d_3, d_4, d_5)$$

where the d_0, \dots, d_5 are the values of d in each of the six generating matrices of the appropriate column shown above. Note that if we read the values of d from below the main diagonal, we needn't worry about which geometry we're in.

For each matrix pair $\{g_i, g_i^{-1}\}$ we decide once and for all which member will be considered the *primary matrix* g_i and which will be considered the *inverse matrix* g_i^{-1} . The choice is arbitrary, but fixed. Applying a small change dM to the matrix pair means that g_i goes to $dM g_i$ while g_i^{-1} goes to $(dM g_i)^{-1} = g_i^{-1} dM^{-1}$.

Now consider how a product $g_i g_j g_k$ is affected by small changes to its factors. As shown in the following table, the effect of applying a change dM to a factor depends on the position of the factor in the product, and also on whether the factor is a primary or inverse matrix.

$$\begin{aligned} (dM g_i) g_j g_k &= dM g_i g_j g_k \\ g_i (dM g_j) g_k &= (g_i dM g_j^{-1}) g_i g_j g_k \\ g_i g_j (dM g_k) &= (g_i g_j dM g_j^{-1} g_i^{-1}) g_i g_j g_k \\ (g_i^{-1} dM^{-1}) g_j g_k &= (g_i^{-1} dM g_i) g_i^{-1} g_j g_k \\ g_i (g_j^{-1} dM^{-1}) g_k &= (g_i g_j^{-1} dM^{-1} g_j g_i^{-1}) g_i g_j^{-1} g_k \\ g_i g_j (g_k^{-1} dM^{-1}) &= (g_i g_j g_k^{-1} dM^{-1} g_k g_j^{-1} g_i^{-1}) g_i g_j g_k^{-1} \end{aligned}$$

In each case, premultiplying a factor by dM is equivalent to premultiplying the product by some dM' . The vector $(d'_0, d'_1, d'_2, d'_3, d'_4, d'_5)$ for dM' depends linearly on the vector $(d_0, d_1, d_2, d_3, d_4, d_5)$ for dM , so we may express the effect as a 6×6 matrix giving the former in terms of the latter. More generally, we may write a $6m \times 6n$ *derivative matrix* D that tells the net effect on all n relations of simultaneously varying all m matrix pairs. We then write a $6n$ -element row vector Δy giving the (small) deviation of each computed relation $g_i g_j g_k$ from the identity, and solve the equation $\Delta x D = -\Delta y$. The solution Δx tells us how to modify the matrix pairs $\{g_i, g_i^{-1}\}$ so that all relations are satisfied simultaneously, up to a linear approximation. Repeating this process three or four times typically gives a solution that is accurate to the precision of the underlying hardware. In other words, it gives a set of matrix pairs that satisfy the group relations

to full precision. This corrects Error C of Section 1.

Note: The number n of relations typically exceeds the number m of matrix pairs, so the equation $\Delta x D = -\Delta y$ is overdetermined. It is most efficiently solved using the Singular Value Decomposition.

Step 4. Refine the value of R_{LSS} .

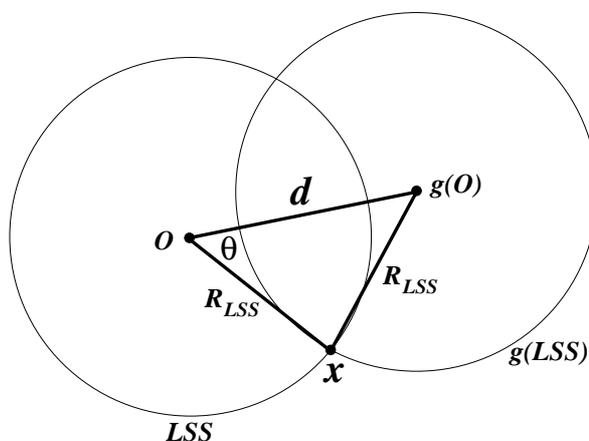


Figure 2: The Law of Cosines gives the radius R_{LSS} of the last scattering surface in terms of the distance d from the basepoint O to its translate $g(O)$, and the angle θ subtended by the corresponding matched circle.

Now that we know the matrix pairs to full precision, we may use them to compute a high precision value for the radius R_{LSS} of the last scattering surface. Each matrix g takes the origin O to one of its translates $g(O)$. The coordinates of $g(O)$ appear as the first column of the matrix g . Let x be an arbitrary point on the circle where LSS intersects $g(LSS)$ (Figure 2), and consider the isocetes triangle with vertices O , $g(O)$, and x . According to whether the geometry is spherical, flat, or hyperbolic, we have

Spherical Law of Cosines

$$\cos R_{LSS} = \cos R_{LSS} \cos d + \sin R_{LSS} \sin d \cos \theta$$

Definition of Cosine

$$\cos \theta = (d/2) / R_{LSS}$$

Hyperbolic Law of Cosines

$$\cosh R_{LSS} = \cosh R_{LSS} \cosh d - \sinh R_{LSS} \sinh d \cos \theta$$

where d is the distance from O to $g(O)$ and θ is the angular radius of the observed circles. It is easy to solve for R_{LSS} in terms of the observed value of θ and the value of d obtained from the refined matrix g .

Some matrices g will yield values of R_{LSS} that are too high, while others yield values that are too low. Averaging the computed values of R_{LSS} over all matrix pairs allows most of the error to cancel out, giving a final value that is accurate to about one part in a thousand for realistically noisy data.

Step 5. Fill out the group.

Even though the refined set of matrix pairs $\{g_i, g_i^{-1}\}$ is fully accurate numerically, it contains many gaps, partly because the original matched circle data had gaps (Error A in Section 1) and partly because in Step 2 we discarded matrix pairs that did not occur in at least two relations. To fill the gaps, consider all possible products of matrices in our set, and add those that (1) were previously missing, and (2) translate the basepoint a distance less than some fixed distance r_{cutoff} . Iterate this process until no new matrices are found. This corrects Error A of Section 1.

Step 6. Recognize the topology.

At this point we have corrected Errors A, B, and C of Section 1, so we may apply the algorithm of [9] to compute a Dirichlet domain for the universe. (A Dirichlet domain is a special type of fundamental domain, a polyhedron whose faces are identified in pairs to give the manifold.) The Dirichlet domain may be viewed using either the computer program Geomview [3] or a web applet available from

the author. From the Dirichlet domain it is straightforward to create a triangulation of the manifold. The triangulation may then be passed to the computer program SnapPea to compute its fundamental group and homology. If the manifold is hyperbolic, SnapPea will also compute its volume. If the manifold is spherical, its volume is $\text{vol}(S^3)/(\text{number of matrices})$.

REFERENCES

- [1] N. CORNISH, D. SPERGEL, AND G. STARKMAN, *Circles in the sky: finding topology with the microwave background radiation*, Classical and Quantum Gravity **15** (1998), 2657–2670.
- [2] N. CORNISH AND J. WEEKS, *Measuring the shape of the universe*, Notices of the American Mathematical Society **45** (1998), 1461–1469.
- [3] M. PHILLIPS ET AL., *Geomview: a viewing program for spherical, flat, and hyperbolic geometry*, available from <http://www.geomview.org>.
- [4] P. DE BERNARDIS ET AL., *A flat universe from high-resolution maps of the cosmic microwave background radiation*, Nature **404** (2000), 955, See also <http://www.physics.ucsb.edu/~boomerang/>.
- [5] S. HANANY ET AL., *MAXIMA-1: a measurement of the cosmic microwave background anisotropy on angular scales of 10 arcminutes to 5 degrees*, *astro-ph/0005123*, to appear in Astrophys. J. Lett.
- [6] J.P. LUMINET, G. STARKMAN, AND J. WEEKS, *Is space finite?*, Scientific American (1999).
- [7] B. RATRA, N. SUGIYAMA, A.J. BANDAY, AND K.M. GORSKI, *Astrophys. J.* **481** (1997), 22.
- [8] J. WEEKS, *SnapPea: a computer program for creating and studying hyperbolic 3-manifolds*, available from <http://www.northnet.org/weeks>.
- [9] J. WEEKS, *Reconstructing the global topology of the universe from the cosmic microwave background*, Classical and Quantum Gravity **15** (1998), 2599–2604.

Received February 1, 2001.