Geophys. J. Int. (1997) **129**, 347–354

Uniqueness of mainly dipolar magnetic fields recovered from directional data

G. Hulot,[1] A. Khokhlov[1,2] and J. L. Le Mouël[1]

[1] *Laboratoire de Géomagnétisme, URA 729 CNRS, Institut de Physique du Globe de Paris, 4 Place Jussieu, 75252 Paris, France. E-mail: ghulot@ipgp.jussieu.fr*

[2] *International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Warshavskoye shosse 79, kor. 2-113556 Moscow, Russia*

Accepted 1997 January 3. Received 1996 December 19; in original form 1996 March 29

SUMMARY

We show that a knowledge of either the signed or the unsigned direction of a potential field on a given smooth surface S, which separates the space into a volume containing the sources and a volume free of sources, sometimes gives enough information for the whole field to be recovered within the free volume, except for a constant multiplier (positive, for the signed case). We show that the best parameter to be considered on the surface S is the number n of loci where the field is known to be either zero (no direction) or normal to the surface. In the case of sources lying outside S ('external-sources' directional problem) we prove that the dimension of the space of solutions is no larger than $n-1$. This implies uniqueness for the external-sources directional problem when $n=2$. In the case of sources lying inside S ('internal-sources' directional problem), we distinguish fields with monopole sources (such as the gravitational field) from those without monopole sources (such as the magnetic field). For gravitational fields, we show that the dimension of the space of solutions cannot exceed n. We note that the only situation of interest is the one for which $n=1$, which implies in practice that the surface S is an isopotential and that the problem has a unique solution. For magnetic fields, we show that the dimension of the space of solutions cannot exceed $n-1$. It follows that the problem has a unique solution when $n=2$. This shows in particular that a geomagnetic field with only two poles (south and north magnetic poles) can be recovered, except for a constant multiplier (positive, for the signed case) from directional data gathered at the Earth's surface. Finally, we note that our results are not restricted to the 3-D space and can readily be extended to two dimensions and higher dimensions.

Key words: archaeomagnetism, geomagnetism, gravity.

1 INTRODUCTION

Recovering the Earth's internal magnetic field $\mathbf{B}$ from measurements that can only be carried out on a surface (such as the Earth's surface or slightly above) is possible in many ideal circumstances. Having all its sources within a bounded surface S, this field is divergence-free and curl-free outside S. It is therefore derived from a scalar potential $U(\mathbf{B}=-\nabla U)$ that satisfies the Laplace equation $\nabla^2 U=0$ and is completely defined if we have a complete set of measurements of either the whole field $\mathbf{B}(S)$, its horizontal component $\mathbf{B}_H(S)$ (assuming there are no magnetic monopoles), or even only its radial component $B_r(S)$ on the surface S. These well-known results are of obvious practical use as they ensure that the determination of $\mathbf{B}$ is then only limited by the accuracy and finite number of measurements and not by some fundamental non-uniqueness (see e.g. Langel 1987). Unfortunately in several other circumstances encountered in practice, such as when only declination information is available, such a fundamental non-uniqueness exists. A review of our current knowledge of this problem and the way it can sometimes be avoided by adding a possibly small but well-chosen set of measurements has recently been given by Lowes, De Santis & Duka (1995), to which the reader is referred. It appears that in two cases of great importance for geomagnetism, the nature and extent of the non-uniqueness problem is still not completely assessed, mainly because it arises in a rather counter-intuitive manner.

The first and best-known case is the one for which only the intensity $B(S)$ on the surface S is assumed to be known. Backus (1970) showed with the help of an *ad hoc* counter-example that, at least in one theoretical case, uniqueness of the recovered field (to within a global sign, of course) could not be guaranteed, but Backus (1968) also pointed out that there exist situations for which uniqueness can be guaranteed, such as when S is a sphere and the field is *a priori* known to be a finite sum of spherical harmonics, or when S is convex and the field has a non-zero monopole contribution (which is a case only relevant to gravitational fields). The exact circumstances leading to non-uniqueness, the extent of this non-uniqueness and the status of the Earth's magnetic field with respect to this problem therefore remain to be assessed. The only thing we know for sure is that the problem does lead to some practical non-negligible errors (the so-called 'Backus effect') when only a finite number of data is available (e.g. Stern & Bredekamp 1975).

The second case of interest arises when only the direction [inclination $I(S)$ and declination $D(S)$] on the surface S is assumed to be known. We will refer to this problem as the directional problem. Apart from the obvious non-uniqueness linked to the fact that any solution can be multiplied by a constant positive value and yet remain a solution, further, much more subtle non-uniqueness seems to be possible in at least a number of circumstances. The situation here is quite similar to that encountered in the previous case. Proctor & Gubbins (1990) displayed an *ad hoc* counter-example for which uniqueness of the recovered field could not be guaranteed (to within a positive constant factor, of course, as will also usually be implicitly assumed hereafter). However, in the same paper they also argued (although without a formal proof) that when the field is *a priori* known to be a finite sum of spherical harmonics, uniqueness would probably be ensured. Hence various circumstances seemingly exist under which uniqueness would, or would not, be ensured.

2 PREVIOUS RESULTS CONCERNING THE DIRECTIONAL PROBLEM

We now wish to consider the possibility of proving that the Earth's internal magnetic field belongs to a class of fields for which uniqueness of the solution can formally be established for the directional problem. Although to our knowledge this question has not yet received any definitive answer, it has already been addressed by several authors, mainly by Kono (1976) and Proctor & Gubbins (1990). Defined in precise terms, we will state the problem as follows.

Let S be a closed bounded surface satisfying the following condition (which amounts to requiring S to be smooth).

Condition 0: For each point P of S, there exists outside S a solid sphere that has P as a boundary point.

Then, let $\mathbf{M}$ be a continuous vector field defined on S, and G the space of functions U defined on and outside S, regular at infinity and assumed to satisfy the following conditions.

Condition 1: U is harmonic outside S, $\nabla^2 U = 0$;

Condition 2: The field $\mathbf{B} = -\nabla U$ is also defined and continuous on S;

Condition 3: There exists a function $g > 0$ such that $\mathbf{B} = g\mathbf{M}$ on S.

The question we then want to answer is simply: if the space G is assumed not empty, what kind of special additional assumption could possibly ensure that G reduces to a 1-D space (that is, to make sure that any function U satisfying conditions 1, 2 and 3 would be uniquely determined to within a constant positive factor, if it exists)?

Kono (1976) believed that no additional assumption was required and thought he had succeeded in proving it. This claim was not questioned for some time, probably because problems linked to some possible non-uniqueness never arose in practice. It was only 10 years later that it was felt by Bloxham (1985) and Gubbins (1986) that under certain (unspecified) circumstances the problem could get trickier. This, however, was not formalized until Proctor & Gubbins (1990) produced their counter-example. Since then, the general belief (e.g. Lowes *et al.* 1995) has been that, because of this counter-example, Kono (1976) appears to be in error. In fact, we have noted two successive loopholes in Kono's proof, which we briefly describe in Appendix A (see also Bloxham 1985).

The next serious attempt to understand the condition under which G reduces to a 1-D space was that by Proctor & Gubbins (1990). They produced the 3-D counter-example mentioned above, conjectured that uniqueness should be ensured if the field is a finite sum of spherical harmonics, but could not reach any additional analytical conclusion concerning the 3-D problem.

From a practical point of view, however, they showed how the uniqueness of an already known solution could actually be 'tested' numerically by searching the zero eigenvalues of a properly constructed matrix. This approach, however, has a number of limitations, linked to the fact that it is numerical and that it can only deal with models described by a finite number of parameters. In fact Proctor & Gubbins only tested fields that were finite sums of spherical harmonics and for which, as already mentioned, we expect theoretical uniqueness. Not surprisingly, their numerical tests always revealed one and only one 'true' (that is, within numerical accuracy) zero eigenvalue (that related to the possibility of multiplying the solution by a constant positive value). When they illustrated their method by 'checking' the non-uniqueness associated with a truncated version of their counter-example, they actually only checked the existence of an additional very small but non-zero (within numerical accuracy) eigenvalue. This revealed a tendency towards non-uniqueness of the tested truncated version of their counter-example but did not prove the non-uniqueness of the complete field. This means that the test proposed by Proctor & Gubbins (1990) can only suggest whether or not a given numerical model of the field behaves as a truncated version of a non-unique solution. It does not allow the direct testing of the theoretical solution of the problem. Eventually, and before turning to this formal aspect of the problem, it should be noted that the test of Proctor & Gubbins only amounts to testing the numerical stability of the linearized procedure used to invert the directional data (e.g. Barraclough & Malin 1971). The existence of stable solutions for historical data (e.g. Bloxham 1986) therefore ensured beforehand that these models would pass the test. Indeed, Proctor & Gubbins (1990) verified that historical models derived from sound directional data always passed the test.

The previous results nevertheless have two interesting consequences. First, they confirm that recovering Earth-like fields from directional data does not lead in practice to spurious effects akin to the 'Backus effect'. Second, they suggest that Earth-like fields do belong to a class of fields for which G reduces to a 1-D space. What kind of class could this be? Some hint can be found from the study of the equivalent 2-D case studied in some mathematical detail by Proctor & Gubbins (1990). S was then a circle, and an important conclusion was that if on S the field was of dipolar form (in some mathematical sense that need not be detailed here), uniqueness of the field recovered from directional data on S could be guaranteed. This conclusion was reached with the help of complex-variables methods that can only be used in two dimensions. The purpose of the present study is to show that this result can be extended to three dimensions (and in fact to any dimension $k > 1$) with the help of an entirely different and quite straightforward approach.

3 PUTTING A BOUND ON THE DIMENSION OF THE SPACE OF SOLUTIONS FOR SPECIAL CLASSES OF FIELDS

Let us first note the important property that the space of solutions G is an open cone (any non-zero positive linear combination of solutions is a solution). Cones are mathematical objects that are not always easy to handle, but they always lie in a linear space. In the present case, one such space is the space of all linear combinations of solutions in G, regardless of the sign of the coefficients involved in the combination. Elements of this space will not in general satisfy condition 3. They will, however, always satisfy the more general condition 3′, which is as follows.

Condition 3′: There exists a function g such that $\mathbf{B} = g\mathbf{M}$ on S.

This suggests that together with the directional problem defined by conditions 0, 1, 2 and 3, we should also consider the problem defined by conditions 0, 1, 2 and 3′. This second problem will be referred to as the 'unsigned' directional problem. It corresponds to the practical situation where the direction of the field is known everywhere on the surface S but not its sign.

Let us define H, the space of the solutions of the unsigned directional problem associated with the directional problem. This space is linear and clearly contains G. Hence we may claim the following.

Theorem 0: The space of solutions G of the directional problem is a cone entirely included in the linear space of solutions H of the unsigned directional problem.

It follows that if we can put some upper bound on the dimension of the space H, the same bound would apply to G. On the contrary, any lower bound applying to the dimension of G would also apply to the dimension of H. In this respect, we can then conclude from the previous section that neither the cone G, nor the space H, will in general be 1-D.

In order to reduce the dimension of H and hence G, we will now impose the following condition on the vector field $\mathbf{M}$.

Condition 4: The field $\mathbf{M}$ is never zero or normal to S, except on a discrete set of disconnected loci (each locus being made of arcwise connected points) $L = (L_1, \dots, L_n)$, where $\mathbf{M}$ is either zero or normal to S.

In the following, condition 4 will be assumed, and the corresponding cone and space of solutions for the two problems will be called G_n and H_n in order to avoid any confusion with the more general case. The interesting point, then, is that a number of simple theorems can be derived concerning H_n. As G_n is included in H_n, many straightforward conclusions can then be drawn about G_n. The logical path we will now follow is quite similar to the one introduced by several mathematicians in the past 30 years when dealing with the so-called 'slant-derivative' (or 'oblique-derivative') problem (e.g. Egorov & Kondratjev 1969; Winzell 1977, 1979; and especially Bicadze 1963).

Theorem 1: Each solution U of H_n (and hence of G_n) takes a constant value $U(L_i)$ on each locus L_i of L.

Proof: Since from condition 4 we know that on each locus L_i of L the field $\mathbf{M}$ is either zero or normal to S, it follows that the tangential component of both $\mathbf{M}$ and $\mathbf{B}$ (because of condition 3′) is zero. As a result, each arcwise locus L_i, lying on the surface S, is an isopotential $U(L_i)$ for each solution U of H_n or G_n.

Because of Theorem 1, we can now state that to each solution U of H_n (or G_n) there corresponds a set of n values $[U(L_1), \dots, U(L_n)]$. This in turn raises the question whether this set of values completely defines a solution U of H_n (or G_n). The answer is yes.

Lemma 1: If a function U different from a constant is harmonic in an open domain D, is continuous together with its first-order partial derivatives up to the boundary ∂D (satisfying condition 0) and assumes a maximum (minimum) at some point P lying on ∂D, then one obtains $\nabla U = \alpha\mathbf{n}$ for some $\alpha > 0$ ($\alpha < 0$), where $\mathbf{n}$ is the normal to ∂D pointing out of D.

Proof: This lemma is mentioned by Bicadze (1963) in a slightly different form, but Bicadze provided neither a proof nor a reference, and simply attributed the lemma to Zaremba. Therefore, we first provided a proof of our own, but Georges Backus pointed out that a proof could be found in Bers, John & Schechter (1954; theorem III, pp. 151–152), in a slightly different form and for the case where D is a bounded domain (this version of the lemma was first derived by Hopf 1931). There is no problem in deriving a similar proof for the case where D is an open domain.

Lemma 2: A solution U of H_n (or G_n) has all its extrema either at infinity or on L.

Proof: From the extremum principle (in its classical form, see e.g. Kellogg 1953) U has its extrema either at infinity or on S. However, from Lemma 1 any extremum at some point P on S is such that at this point $\nabla U = \alpha\mathbf{n}$ for some $\alpha \neq 0$. It then follows from condition 3′ that at P, $\mathbf{M}$ is either zero or normal to S. Thus P belongs to L, and U has all its extrema either at infinity or on L.

Lemma 3: A solution U of H_n (or G_n) that is zero everywhere on L is zero everywhere on and outside S.

Proof: Consider a solution of H_n that is zero everywhere on L. Since U is assumed regular at infinity, this means from Lemma 2 that all extrema of U are 0. Thus U is zero everywhere on and outside S.

Theorem 2: Two solutions U_1 and U_2 of H_n (or G_n) that share common values on L are identical.

Proof: Consider two solutions of H_n that share common values on L. Define U as the difference between the two solutions. Obviously U is a solution of H_n that is zero everywhere on L. It then follows from Lemma 3 that U is zero everywhere on and outside S, hence the two solutions are identical.

Theorem 2 shows that the set of n values $[U(L_1), \dots, U(L_n)]$ that a solution U of H_n (or G_n) takes on L completely defines U. We can also make more formal statements, as follows.

Lemma 4: If there exist n linearly independent solutions U_j in H_n (or in G_n), then the matrix $M_{ij} = U_j(i)$ formed of the n values each solution takes on the n loci L_i is invertible.

Proof: Assume M_{ij} is not invertible; there then exists a set of n values λ_j not all equal to zero and such that

$$\sum_{j=1}^{n} M_{ij}\lambda_j = \sum_{j=1}^{n} \lambda_j U_j(i) = 0 . \tag{1}$$

Then, $\Sigma_{j=1}^{n} \lambda_j U_j$ is a solution of H_n that takes zero values everywhere on L. From Lemma 3, this means that this linear combination must be identically zero. This contradicts the assumption that the U_j are independent. It follows that M_{ij} is invertible.

Theorem 3: There cannot be more than n linearly independent solutions in the space of solutions H_n (or in the cone of solution G_n).

Proof: Assume there are $n+1$ independent solutions in H_n. The first n of them then define a matrix $M_{ij} = U_j(i)$, which we know from Lemma 4 is invertible. It follows that there exists a set of n values λ_j, not all zeros, and such that

$$\sum_{j=1}^{n} \lambda_j U_j(i) = U_{n+1}(i).$$

This states that the two solutions $\Sigma \lambda_j U_j$ and U_{n+1} of H_n would share common values on L, and would therefore be identical (Theorem 2). This contradicts the assumption that the $(n+1)$ U_j be independent.

4 CONSEQUENCES FOR GRAVITATIONAL FIELDS

In the case $n=1$, we know from Theorem 3 that H_1 and G_1 are either empty or 1-D. This implies uniqueness of the solution (if it exists) for both (unsigned or not) directional problems, but to what practical situation does the assumption $n=1$, for which there is only one locus of poles on the surface S, correspond?

Theorem 4: Assume there is only one locus L_1 (i.e. $n=1$).

If $L_1 \neq S$, then there are no solutions in H_1 satisfying $\mathbf{B} \neq \mathbf{0}$ everywhere on $S - L_1$, and G_1 is empty.

If $L_1 = S$ and if condition 0 is extended so that for each point P of S there also exists *inside* S a solid sphere that has P as a boundary point, then H_1 and G_1 are both 1-D.

Proof: Assume $L_1 \neq S$. Consider a solution U of H_1 satisfying $\mathbf{B} \neq \mathbf{0}$ everywhere on $S - L_1$. U defines a surface function U_S on S. At any point P where U_S takes a (surface) maximum or minimum value on S, the surface derivatives must be zero. If P is not on L_1, then $\mathbf{B} \neq \mathbf{0}$ and $\mathbf{B}$ is normal to S. From condition 3′ this means that P must lie on L_1 anyway, hence U_S, and therefore, U, take a constant value U_0 on S. This in turn implies that $L_1 = S$, which contradicts our assumption, thus no such solution U of H_1 can exist. It also follows that G_1 must be empty (since solutions of G_1 are solutions of H_1 satisfying $\mathbf{B} \neq \mathbf{0}$ everywhere on $S - L_1$). Now assume $L_1 = S$ and extend condition 0 so that for each point P of S there also exists *inside* S a solid sphere that has P as a boundary point. We then know that S is an equipotential U_0 and that the solution of the corresponding Dirichlet problem exists (e.g. Kellogg 1953, p. 284). This solution is also a solution of our problem. H_1 and G_1 are then both 1-D.

It follows from Theorem 4 that the only physical situation corresponding to $n=1$ is one for which the surface S itself is an isopotential. Furthermore, using the Gauss theorem applied on S, we immediately see that this is a situation that would require a non-zero monopole contribution from the sources. It follows that it can be encountered only when one deals with gravitational fields (and not when one considers magnetic fields). Theorem 4, then, is simply the statement that knowing the shape of the geoid (and knowing that the masses lie inside it) completely defines the gravitational field to within a constant positive multiplier. Interestingly, we note that a situation for which S would slightly depart from the geoid would not lead to a similar conclusion. If, for instance, one considers the case of an ellipsoidal geoid and of a spherical surface S (sharing the same axis of symmetry), then there are three loci L_i (one at each geographical pole and the geographic equator), and the uniqueness of the field cannot be guaranteed. In a more realistic situation where S would be the surface of the Earth, which fluctuates about the geoid on very small length scales, and for which the number n of loci would be very large, the situation would be even worse. Obviously, using directional data measured at the Earth's surface does not seem to be the safest way to recover the gravitational field of the Earth. Fortunately, this is not the way people usually proceed.

5 CONSEQUENCES FOR MAGNETIC FIELDS

The cases directly relevant to magnetic fields are those for which n is larger than 1. Although it seems at first glance that we lose the uniqueness property we could claim for $n=1$ and the gravitational field, we now can take advantage of the fact that the magnetic field cannot have monopole sources. This is a strong constraint which implies the following.

Theorem 5: The solutions of H_n (and respectively G_n) with no monopole sources define a space (and respectively a cone) that cannot have more than $n-1$ linearly independent solutions.

Proof: Solutions of H_n (and respectively G_n) with no monopole sources obviously define a space H_n (and respectively a cone G_n'). Now assume there are n independent solutions U_j in H_n'. From Lemma 4 we know that for any given scalar $V>0$, a linear combination $U = \Sigma \lambda_j U_j$ such that $U(i) = V > 0$ on all loci L_i could be constructed. From Lemma 2, however, we see that this would imply that $0 \leq U \leq V$ everywhere on and outside S since U is zero at infinity. U would therefore remain positive even as we go towards infinite distances from S. This would not be possible if U had no monopole sources: at infinite distance, the lowest-order sources dominate, and without monopole sources this would lead to a potential of the form $U(r, \theta, \varphi) \propto r^{-(n+1)} \Sigma_m g_n^m Y_n^m(\theta, \varphi)$ (with $n>0$), which requires at least some negative values for U. Thus U must have some monopole sources. This contradicts the fact that U is a linear combination of solutions with no monopole sources.

The obvious consequence of Theorem 5 is that uniqueness is recovered for $n=2$ when one deals with magnetic fields. The exciting point now is that it corresponds to an Earth-like situation: if the field has exactly one North magnetic pole and one South magnetic pole, the field can be recovered to within a constant positive multiplier. This generalizes the 2-D result of Proctor & Gubbins (1990) to 3-D.

6 REMARK ABOUT EXTERNAL SOURCES

It is interesting to note that although we dealt with the directional problem only in the case of a field harmonic outside S with sources located inside S, similar conclusions can be reached when dealing with the directional problem for fields harmonic inside S and sources lying outside S. This can be seen either by using exactly the same reasoning as was done above, or in a more restricted way (when S is a sphere) by making use of Kelvin's transform (e.g. Kellogg 1953). In both cases, Theorems 0 to 3 can be reached in a straightforward manner.

Theorem 5 cannot be stated in exactly the same way and it is worth discussing this point in slightly more detail. If one assumes that S is a sphere, then to any potential field U_{out} with sources lying outside S there corresponds, via Kelvin's transform, a potential field U_{in} with sources lying inside S. The interesting point is that the (external) potential U_{in} of an internal monopole source then becomes a potential U_{out} uniform inside S. It follows that the requirement in Theorem 5 that the (external) field has no monopole (internal) sources, would translate into the requirement that the (internal) field has a zero average potential inside S. In other words this suggests that when one deals with external sources Theorem 5 should become as follows.

Theorem 5′: For a field defined inside S and sources lying outside S, the solutions of H_n (and respectively G_n) with zero average value define a space (and respectively a cone) that cannot have more than $n-1$ independent solutions.

Proof: A rigorous proof can in fact be derived in the general case (S not necessarily a sphere) by introducing the linear requirement that the average potential be zero in S and using it to reduce by one unit the maximum dimension of the space H_n. The details of the proof are then very similar to those given in the proof of Theorem 5, and we will only briefly outline them here. If one assumes that the space H'_n of solutions with zero average value can have n independent solutions, then a linear combination of them can be constructed that is strictly positive on S, and therefore in S. This contradicts the fact that it is a linear combination of solutions with zero average values.

The interesting point is that in physical situations we are only interested in recovering the *vectorial* (and physical) field **B**. The requirement that the potential be of zero average value in S does not in fact make a physical constraint and can always be imposed. It follows that in the case of external sources, there are always no more than $n-1$ independent vectorial (and physical) solutions **B**.

7 DISCUSSION AND CONCLUSIONS

We have shown that some upper bounds can be put on the dimension of the space and cone of solutions of the directional problems, provided that on the surface S where the direction of the field is given there only are a finite number n of loci where the field is known to be either zero or normal to the surface S. Although we have dealt with the 3-D problem, it is important to note that all results can readily be extended to the k-dimensional case, provided that k is at least 2 and that S is understood as a manifold satisfying a generalized condition 0. Note also that our results do not depend on the dimension K of the space. In the following we will focus on the consequences of our study for the 3-D case, which is directly relevant to geophysics.

In the event that one considers sources lying outside S and tries to recover the field inside S (external-sources directional problem), the dimensions of the space and cone of solutions are no larger than $n-1$ (Theorem 5′). Although we do not intend to discuss this case in any further detail, we note that this already implies that if $n=1$ (one locus), the external-sources directional problem cannot have a solution. Much more interesting is the conclusion that if $n=2$, the solution, *if it exists,* is completely defined to within a positive, unless one considers the unsigned problem, constant factor. This is a situation that would correspond to a mainly dipolar external geomagnetic field. Note, however, that no claim has been made in the present study about the *existence* of the solution (or more generally about the lower bound of the dimension of the space and cone of solutions of the two directional problems).

The internal-sources directional problem is much more relevant to the problems encountered in studies of the Earth's potential fields. The conclusions in this case depend on the nature of the field under consideration. If it is known to have monopole sources (such as in the case of the gravitational field), then the maximum dimension of the space and cone of solutions of the unsigned and signed directional problem is n. We noted that the only situation for which this bound is of immediate use is when the surface S happens to be an isopotential (e.g. the geoid). In this case $n=1$ and the solution *exists* and is known to within a (positive, if signed) constant factor. But we also noted that the usual practical situation is one for which n is large, and for which the bound we have found becomes useless. The situation here is exactly the reverse of that encountered in the intensity problem we mentioned briefly in the Introduction: whereas the presence of a monopole source efficiently warrants that the field can be recovered from the knowledge of its intensity at the surface S (if, in addition, S is convex; Backus 1968), such a monopole is essentially a problem when it comes to test the possibility of recovering the field from directional data.

Fortunately, the magnetic field has no monopole sources and for such types of field, the dimensions of the space and cone of solutions for the internal-sources directional problems have been shown to be no larger than $n-1$. This means that there are no solutions for $n=1$ (in fact, we noted that $n=1$ required a monopole source) and that the solution is unique, *if it exists,* to within a (positive, if signed) constant factor when $n=2$. It also means that if n is larger than 2, uniqueness can no longer be guaranteed. Of course, this does not mean that non-uniqueness will systematically arise if n is larger than 2, or that the space and cone of solutions are exactly of dimension $n-1$. Nonetheless, it is interesting to point out that this bound increases with n not just as a result of our inability to find a more stringent bound, but also as a result of some real possibility for the dimension of the space and cone of solutions to increase with n. The best way to illustrate this point is to note that our bound predicts the dimension of the cone of the symmetric solutions for the counter-example produced by Proctor & Gubbins (1990). Indeed, the axisymmetric directional boundary conditions they imposed on S (which was a sphere) implied four loci L_i: the two geographical poles (L_1 for North and L_4 for South), and two parallels symmetric about the equator and of intermediate latitude (L_2 in the northern hemisphere and L_3 in the southern hemisphere). Therefore, $n=4$, and the maximum dimension of the cone of solutions is 3. If we now want the solutions to lead to fields anti-symmetric about the equator (which Proctor & Gubbins assumed), this requires that $U(L_2)-U(L_1)=U(L_4)-U(L_3)$, which is one additional and independent linear relationship of the type we encountered when requiring that the field have no monopole sources. This will again reduce the maximum dimension of the cone of solutions by one unit (the rigorous proof of this statement is similar to that given for Theorem 5). If we further recall that Proctor & Gubbins did produce two independent anti-symmetrical solutions of their problem, we can then state that the cone of anti-symmetrical solutions is exactly of dimension 2. Of course, the space of solution of the corresponding unsigned directional problem is also of dimension 2.

This shows that although all our results have been derived in terms of upper bounds for the dimension of the space and cone of solutions, there are situations for which these bounds can be reached. Counting the loci L_i therefore appears to be the safest and easiest way of estimating this dimension. Taking into account symmetry properties can improve the prediction. The only reason why it may not always lead to an exact prediction is that this method does not test the internal consistency of the field of direction $\mathbf{M}$. Clearly, any more precise prediction would involve more sophisticated tests in order to take this aspect into account.

Finally, we return to the Earth's magnetic field and answer the question that motivated the present study. May we recover this field from pure directional data? The answer remains unknown if we wish to simultaneously consider the external and internal sources, but not if we are ready to ignore the weak external sources (which is what happens in practice when people deal with pure directional data). Indeed, in that case, and if we happen to know that there is only one South and one North magnetic pole, the answer becomes yes (to within a constant factor, positive if the direction is signed). How sure can we be that there is only one South and one North pole?

All models of the main geomagnetic field display only two poles at the Earth's surface. However, this property no longer holds when the models are continued down deep towards the core. This shows that geometrical attenuation is an efficient way of preventing the appearance of secondary poles (Appendix B illustrates this point in terms of a simple order-of-magnitude computation). It also shows that we would not be able to recover the field if observations were made just above the core. As pointed out by one of the referees, this is an intriguing and somewhat counter-intuitive conclusion. Intuition suggests that being nearer to the sources should be a good thing, not a problem, but intuition in this case is based on our habit of dealing with linear parameters of the field (such as any component, or the potential) defined on a sphere. Continuations up and down of such linear parameters are then straightforward, which means that knowing the parameter at some surface is formally equivalent to knowing the parameter at any other surface above the sources. Thus, being nearer to the sources is better in that case only to the extent that errors on the observations would then be smaller relative to the signal. In the present case the situation is completely different because directions are non-linear parameters, and non-linearity is likely to prevent a one-to-one correspondence to be found between directions defined on two different surfaces, one near and one far from the sources. This seems to be the case at present. We can claim that the field is completely recovered only when we are in a situation where the field is mainly dipolar. Starting from such a situation, continuations up and down are then possible on a one-to-one basis. However, if we start from a surface standing nearer to the sources, so that several secondary poles are to be found, upward and downward continuation no longer seem possible on a one-to-one basis. It should, however, be emphasized that our study is based on bounds. Some of the ambiguity could possibly be removed by future refined studies. In any case, the good news is that we stand far from the core.

The geomagnetic field of internal origin also includes crustal sources of smaller length scales. These shallow sources can be neglected when one considers the field of very large scale, but make the main contribution to the degrees of the field that are larger than 13 (Counil, Cohen & Achache 1991; Jackson 1994). The spatial behaviour of this field is very different from that of the main field. Instead of decreasing as an exponential function of degree n, each degree above 13 tends to contribute a more or less constant value to the total magnetic field. It follows that the condition for additional poles to arise is much more likely to be reached for some critical degree. Estimating this critical degree from the spatial spectrum of the crustal field would be somewhat hazardous and we will not try to do it here. Rather, we will rely on a number of observations that have been made in the immediate vicinity of the 'true' magnetic poles.

Searching for the North magnetic pole, Sir James Clark Ross recorded the remarkable value of 89°59′ several times on June 1st, 1831 (Barraclough & Malin 1981). This extraordinary observation is unique. Ross only obtained an inclination of 88°40′ when he went searching for the South magnetic pole (Mayaud 1953). We now know that this is because of the external field, which makes the apparent pole (defined as being a place where inclination is 90°) describe an ellipse of the order of a couple of tens of kilometres wide *every day* (Dawson & Newitt 1982). In addition, local anomalies, possibly reaching a thousand nanoteslas, are responsible for contradictory local declination measurements (Mayaud 1953). In fact, present estimates of the location of the magnetic poles always rely on some processing of a local survey, precisely in order to avoid this problem (e.g. Mayaud 1953; Dawson & Newitt 1982). From local observations, it therefore seems that several secondary poles could coexist within a region of typical lateral dimension of about 100 km.

The presence of two and only two poles would guarantee theoretical uniqueness for the directional problem. The fact that several secondary poles may exist is unfortunate. However, we should recall that in practice, and for the early data for which no intensity value is available (archaeomagnetic and pre-1840 historical data), measurement errors in declination and inclination can be fairly large. In addition, these measurements have sometimes been made 1000 km apart and only very occasionally near the actual poles. In such a situation, and given the fact that we are essentially interested in the large-scale main geomagnetic field, it seems reasonable to address the data as if there were only two magnetic poles at the Earth's surface. It follows that any archaeomagnetic or early historical geomagnetic field could be recovered from pure directional data to within a constant (positive, if the direction is signed) multiplier, provided it is mainly dipolar (for example made of a strong dipole component and displaying a spatial spectrum comparable to that of the present field). It must be kept in mind, however, that uniqueness is no longer guaranteed whenever the field is not mainly dipolar, as is possibly the case during reversals.

ACKNOWLEDGMENTS

We thank Georges Backus for enthusiastically suggesting that we also address the unsigned directional problem, not just the signed directional problem as we had done in an earlier version of the paper. This made it possible to simplify a number of proofs and generalize our earlier results. We also thank Andrew Jackson for his interesting remark about the puzzling consequences of our results with respect to upward and downward continuation of the field. This work was partly supported by INTAS grant no. 94-3950. This is DBT contribution no. 67 and IPGP contribution no. 1453.

REFERENCES

Backus, G.E., 1968. Applications of a non-linear boundary value problem for Laplace's equation to gravity and geomagnetic intensity surveys, *Q. J. Mech. appl. Math.,* **21,** 195–221.

Backus, G.E., 1970. Non-uniqueness of the external geomagnetic field determined by surface intensity measurements, *J. geophys. Res.,* **75,** 6339–6341.

Barraclough, D.R. & Malin, S.R.C., 1971. Synthesis of International Geomagnetic Reference Field values, *Inst. geol. Sciences Rept,* **71/1.**

Barraclough, D.R. & Malin, S.R.C., 1981. 150 years of the North magnetic pole, *Nature,* **291,** 377.

Bers, L., John, F. & Schechter, M., 1964. *Partial Differential Equations,* Wiley, New York, NY.

Bicadze, A.V., 1963. A homogeneous problem for the slant derivative for harmonic functions in three-dimensional regions, *Soviet Math. Doklady,* **4,** 156–159.

Bloxham, J., 1985. Geomagnetic secular variation, *PhD thesis,* Cambridge University, UK.

Bloxham, J., 1986. Models of the magnetic field at the core–mantle boundary for 1715, 1777, and 1842, *J. geophys. Res.,* **91,** 13 954–13 966.

Counil, J.L., Cohen, Y. & Achache, J., 1991. The global continent–ocean magnetisation contrast, *Earth planet. Sci. Lett.,* **103,** 354–364.

Dawson, E. & Newitt, L.R., 1982. The magnetic poles of the Earth, *J. Geomagn. Geoelectr.,* **34,** 225–240.

Egorov, Y. & Kondratjev, V., 1969. The oblique derivative problem, *Math. USSR Sbornik,* **7,** 139–169.

Gilbarg, D. & Trudinger, N., 1986. *Elliptic Differential Equations of Second Order,* Springer-Verlag, Berlin.

Gubbins, D., 1986. Global models of the magnetic field in historical times: augmenting declination observations with archeo- and paleomagnetic data, *J. Geomagn. Geoelectr.,* **38,** 715–720.

Hopf, E., 1931. Über den functionalen, insbesondere den alanytischen charakter, der Lösungen elliptischer differentialgleichungen zweiter ordnung, *Math. Z.,* **34,** 194–233.

Jackson, A., 1994. Statistical treatment of crustal magnetisation, *Geophys. J. Int.,* **119,** 991–998.

Kellogg, O.D., 1953. *Foundations of Potential Theory,* Dover, New York, NY.

Kono, M., 1976. Uniqueness problems in the spherical harmonic analysis of the geomagnetic field direction data, *J. Geomagn. Geoelectr.,* **28,** 11–29.

Langel, R.A., 1987. The Main Field, in *Geomagnetism,* Vol. 1, pp. 249–512, ed. Jacobs, J.A., Academic Press, London.

Langel, R.A. & Estes, R.H., 1982. A geomagnetic field spectrum, *Geophys. Res. Lett.,* **9,** 250–253.

Lowes, F.J., De Santis, A. & Duka, B., 1995. A discussion of the uniqueness of a Laplacian potential when given only partial field information on a sphere, *Geophys. J. Int.,* **121,** 579–584.

Mayaud, P.N., 1953. Rapports scientifiques des expéditions polaires françaises SIV 2: Le pôle magnétique sud en 1952 et les déplacements comparés des pôles Nord et Sud de 1842 à 1952, *Ann. Geophys.,* **9,** 266–276.

Proctor, M.R.E. & Gubbins, D., 1990. Analysis of geomagnetic directional data, *Geophys. J. Int.,* **100,** 69–77.

Stern, D.P. & Bredekamp, J.H., 1975. Error enhancement in geomagnetic models derived from scalar data, *J. geophys. Res.,* **80,** 1776–1782.

Winzell, B., 1977. The oblique derivative problem, 1, *Math. Ann,* **229,** 267–278.

Winzell, B., 1979. The oblique derivative problem, 2, *Ark. Mat.,* **17,** 107–122.

APPENDIX A: LOOPHOLES IN THE PROOF OF KONO (1976)

As mentioned in the main text, we noted two successive loopholes in Kono's proof, which we will now briefly describe.

The first arises when Kono considers two functions g (positive) and W, assumed to be analytical on S and sharing common isovalues on this same surface. The claim of Kono at this point is that g can then be written in the form of a single-valued function of W on S. This seems incorrect. Indeed, considering g a positive analytical function with at least one isovalue line g_0 on S, and defining $W=(g-g_0)^2$ on S, then W is also analytical on S, shares common isovalues with g, yet g is not a single-valued function of W. Being more specific, the error in the proof apparently lies in the fact that Kono makes the additional assumption in his eq. (4) that the derivatives of g with respect to W exist, where actually they may not exist (as when W and $g-g_0$ are zero in our counter-example).

The second difficulty arises a little later when Kono claims that $P(W)$ being a given function of W, if a general function Q of W and of two curvilinear coordinates u and v (taken along the iso-W surfaces) satisfies the conditions $dP/dW+\partial Q/\partial W=\partial Q/\partial u=\partial Q/\partial v=0$ on S (his eq. 9), then $dP/dW=\partial Q/\partial W=0$ on S. To prove this Kono used a dubious reasoning, stating that Q on S has to be a function of W and of at least one additional parameter, while $\partial Q/\partial W=-dP/dW$ can only be a function of W only. From this he concluded that $\partial Q/\partial W=-dP/dW=cte$ on S. This seems incorrect. From the boundary conditions assumed for Q on S, we explicitly have the conclusion that on S, Q is a function of W only. Therefore, the only conclusion we can reach is that $P(W)+Q(W)=cte$ on S. It follows that Kono's proof appears to be incomplete.

APPENDIX B: GEOMETRICAL ATTENUATION AND SECONDARY POLES

Consider a mainly dipolar axisymmetric field with some degree n axisymmetric multipolar contribution, $U(\theta)=U_0[\cos\theta+\alpha_n P_n(\cos\theta)]$. A pole (or a line of poles) will arise if and only if $U'(\theta)=U_0[-\sin\theta+\alpha_n\partial P_n(\cos\theta)/\partial\theta]=0$. This equation has two trivial solutions corresponding to the two poles of the dipole. The question of interest being the possibility for the field to have additional poles away from these poles, we may divide the previous equations by $\sin\theta$, which then become $dP_n(u)/du=-1/\alpha_n$, and seek the α_n for which this new equation has at least one solution. We know that the Legendre polynomials $P_n(u)$ are well-behaved degree n polynomials, with a norm of $(n+1/2)^{-1/2}$. Its derivative would not exceed a typical value of order $n^{1/2}$. It follows that for new poles to arise, the minimum relative strength of the multipolar field of degree n with respect to the dipolar field would typically have to be of order $\alpha_n \propto n^{-1/2}$. An additional similar computation can be carried out with the help of a pure sectorial non-dipole field $Y_n^n(\theta,\varphi)$ in place of $P_n(u)$. This non-dipole term being normalized following the Schmidt normalization in use in geomagnetism, a relative amplitude of the order of $\alpha_n^n \propto (2^n n!/(n\sqrt{2}\sqrt{(2n)!})) \propto n^{-3/4}$ would then be needed. Clearly, in both cases (corresponding to two extremes), the relative amplitude of the non-dipole field required for new poles to arise is of the order of a weakly decreasing power law of degree n. Because of this and because the degree-two field at the Earth's surface is already an order of magnitude smaller than the dipole field, while the amplitude of the higher degrees of the non-dipole main field decrease as an exponential function of n (e.g. Langel & Estes 1982), the condition necessary for additional poles to arise is never met in practice.