Office Note 456

NORMALIZATION OF THE DIFFUSIVE FILTERS THAT REPRESENT THE INHOMOGENEOUS COVARIANCE OPERATORS OF VARIATIONAL ASSIMILATION, USING ASYMPTOTIC EXPANSIONS AND TECHNIQUES OF NON-EUCLIDEAN GEOMETRY; PART I: ANALYTIC SOLUTIONS FOR SYMMETRICAL CONFIGURATIONS AND THE VALIDATION OF PRACTICAL ALGORITHMS

R. James Purser*
Science Applications International Corp., Beltsville, Maryland

June 30, 2008

* email: jim.purser@noaa.gov

Abstract

A versatile way of synthesizing the spatial covariance linear operators of a variational assimilation scheme is by using basic building blocks that are either of pure Gaussian form or, more generally, conform to the distributions that result from a finite-time diffusive process whose effective diffusivity is non-trivially tensorial and spatially varying. Such a synthesis is exemplified by the so-called 'recursive filter' approach to assimilation and by the explicitly diffusive syntheses proposed by Derber and Rosati and generalized by Weaver and Courtier. An outstanding problem associated with the practical implementation of these methods in the important cases where there is smooth spatial variability of the diffusive tensor is that of anticipating the local amplitude and derivatives of the result of diffusing a unit impulse. It is on the basis of such an estimate that the diffusion operator or filter must be symmetrically modulated by an amplitude profile in order to ensure that the background covariance is being applied in the assimilation with the amplitude (i.e., the variance) intended.

In the spatially homogeneous case, there is no difficulty because the result of a diffusive process acting over a half-unit of effective time leads to a pure Gaussian with second moment tensor equal to the diffusivity and an amplitude calculable precisely from the standard formula for such a Gaussian. In the inhomogeneous case, the amplitude is altered from the Gaussian value by an adjustment factor that depends in a complicated way on the local variations of the diffusivity tensor. This note and its sequel describe an approach, based on a reformulation of the variable-diffusivity problem into an equivalent problem of uniform diffusivity, but in a non-Euclidean geometry, which has the advantage of forcing the amplitude adjustment factor to depend only upon the particular combinations of higher derivatives of the induced metric that define 'curvature' and its derivatives. The proposed approach leads to the systematic construction of successive asymptotic approximations for the adjustment factor in terms of the relevant 'curvature' diagnostics of the local diffusivity variations.

In this Part I, the problem is discussed in the restricted case where the geometry of the transformed representation happens to possess certain symmetries. In the special case where the implied curvature is uniform (spheres and hyperbolic spaces), analytic methods can be brought to bear and lead to benchmark asymptotic series for both the two- and three-dimensional cases. A more general approach, based on a systematic iteration to recover successive 'slabs' of the asymptotic parameters, is introduced and exhibited, for the two-dimensional case, in the only slightly less restrictive case of a geometry with variable, yet still axi-symmetric, distribution of curvature. It is demonstrated that the proposed procedure reproduces the analytic coefficients that correspond to the uniform curvature cases but that additional terms involving derivatives of curvature emerge in the more general geometry.

Treatment of the diffusion problem in the most general smooth manifolds of two and three dimensions requires the full resources provided by the tensorial machinery of Riemannian geometry. This topic, and the demonstration that the corresponding versions of essentially the same iterative algorithm apply also to general manifolds, are deferred to the sequel, where they are discussed thoroughly.

1. Introduction

The recursive filter method for covariance synthesis has been shown by Purser et al. (2003) to be capable of generating quasi-Gaussian shapes with arbitrary degrees of second-moment anisotropy quantified by the local rank-two symmetric 'aspect tensor'. The process by which the covariance contributions are generated is essentially one of simulated diffusion and, in this respect, the method is closely related to the method proposed by Derber and Rosati (1989) for oceanic horizontally-isotropic covariance synthesis and its recent extension to fully anisotropic covariance synthesis proposed by Weaver and Courtier (2001). In the methods of Derber and Rosati and of Weaver and Courtier, an explicit simulation of the diffusive process is carried out, usually with several hundreds, or even thousands, of steps of 'pseudo-time'; the recursive filter method of Purser et al. employs an approximate factorization of the diffusive process into a relatively small sequence of carefully chosen one-dimensional filters of implicit type, permitting an efficient simulation of finite-duration diffusion provided the grid has neither coordinate singularities nor irregular bounderies. When the correlations are homogeneous (same shape and size) in space, regardless of whether they are isotropic or not, and regardless of whether the intended covariance amplitude ('variance') is constant or not, the correlation shapes are exactly Gaussian and the intended normalization of their amplitudes, and the implied amplitudes of the associated covariances, poses no problem. But a serious difficulty occurs when the shape or size of the correlation function, as encoded by the diffusivity or aspect tensor used in the generation process, changes from place to place for, in this case, the impulse-response functions obtained as solutions to the diffusion equation can diverge significantly from the classical Gaussian shapes and the naive application of the Gaussian amplitude formula can produce unacceptably large errors. Weaver and Courtier proposed an empirical 'Monte-Carlo' technique for estimating this amplitude that requires a very large number of solutions to the diffusion process to be performed on white-noise initial fields produced from a random-number generator. This method works with any arbitrary degree of anisotropy and can be made as accurate as desired by supplying a sufficiently large set of independent trials from which the amplitude estimates are obtained by a process of averaging over all trials but, like all Monte-Carlo methods, the accuracy is improved by each factor of two only by increasing the number of trials by a factor of four. In operations, it is usually not feasible to apply this method when the pattern of inhomogeneity changes from day to day, owing to the cost involved.

When the inhomogeneity is of a kind where spatial variations of the aspect tensor components are restricted to being smooth and gradual over the intrinsic scales of the local covariances themselves, we might expect there to be asymptotic methods, derived from local diagnostics of these spatial variations, that adequately approximate the adjustment factors we need to apply to the classical Gaussian amplitudes to recover the actual local amplitudes. If so, and if the estimation formulas are not too complicated, such estimates might be obtainable at a fraction of the cost of the Monte-Carlo methods and within the budget allowable for operational implementation at each assimilation period. In this note and its sequel (Purser 2008) we investigate the prospect of constructing an asymptotic expansion for the amplitude based on what is known as the 'parametrix expansion' for the diffusion problem reformulated as a process of *isotropic* diffusion in a non-Euclidean space whose metric assumes the role formerly taken by physical-space diffusivity tensor. A 'parametrix' is a real scalar function of spatial displacement and

pseudo-time which, at least in a small neighborhood, multiplies the idealized Gaussian solution imposed upon local radial geodesic coordinates to yield the corresponding true solution to the diffusion process that properly incorporates the effects of the space's intrinsic curvature. The idea of factoring the true solution in this way appears to derive from the work of Levi (1907) although the term, 'parametrix', was coined by Hilbert (1912). In practice, only the leading coefficients of the power series expansion of the parametrix about the origin are sought and used to make approximate corrections to the diffusion solution's central amplitude and (optionally) its derivatives – the so-called 'parametrix expansion method'.

The reformulated diffusion problem, while formally different from the diffusion equation we have hitherto modelled the covariance generation procedure on, achieves the same practical objective of producing smooth quasi-Gaussian covariance shapes. But it has two important advantages: first, it formally eliminates any dependence of the diffusion model on the physical-space metric terms, which makes it immaterial which vertical or horizontal coordinates happen to be used for the presentation of the assimilation data; second, it standardizes the diffusivity problem so that the asymptotic series depends only upon diagnostics derived from the trans-formed non-Euclidean metric that are intrinsic and covariant properties of the implied curvature of that space. This second property allows a relatively simple asymptotic formula to be con-structed at each given 'degree' of expansion, for which the equivalent degree of approximation for the untransformed diffusion problem expressed in the usual physical-space would involve an impractical multitude of exceedingly complicated terms. Also, as we shall discuss in more detail in the sequel, the reformulated problem is one that has been studied in detail by math-ematicians in various different and unrelated fields in contexts unrelated to our own narrow objective. [†]

We begin by reviewing the case of homogeneous diffusion. The solution of the homogeneous diffusion equation acting at time t on a distribution that begins at initial time $t = 0$ as a unit impulse is easily verified to have a simple Gaussian solution at each subsequent time. We shall denote the diffusivity, D, and write the equation:

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x} D \frac{\partial P}{\partial x}. \tag{1.1}$$

Then its solution,

$$P(t, x) = \frac{1}{(2\pi)^{\frac{1}{2}}(2Dt)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{4Dt}\right), \tag{1.2}$$

shows that, at time $t = \frac{1}{2}$, the Gaussian solution has the (centered) second moment, or 'aspect tensor' equal to the diffusivity itself. The amplitude is then simply,

$$P(1/2, 0) = \frac{1}{(2\pi D)^{\frac{1}{2}}}. \tag{1.3}$$

In several (N) dimensions, principal components of the diffusivity can be used to construct coordinates which can be used to separate the independent orthogonal components of the

---

[†] I am indebted to Dr. Dezso Devenyi of NOAA/ESRL for bringing some of these works to my attention.

diffusion of a non-trivially tensorial, but constant, $\boldsymbol{D}$. Then it becomes clear (e.g., Hogg and Craig 1978, Chap. 12, for a similar problem in the context of statistical distributions) that the amplitude after a half unit of time also has a simple form,

$$P(1/2, \boldsymbol{0}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{D}|^{\frac{1}{2}}} \tag{1.4}$$

where $|\boldsymbol{D}|$ denotes the determinant of $\boldsymbol{D}$.

In the more challenging circumstances where the diffusion is spatially inhomogeneous, there is no closed-form expression for the amplitude of the resulting distribution. In the case of variable diffusivity, we continue to identify the aspect tensor with twice the product of the diffusivity tensor and the duration of its action; therefore, as a matter of numerical convenience, we shall henceforth mostly assume this 'standard' duration to be one half, so that the diffusivity and aspect tensors are numerically equivalent. The Gaussian amplitude formula (1.4) is no longer able to provide the amplitude of the result of applying the diffusion equation *exactly*. The disparity may be expressed in terms of a correction factor we shall refer to as the 'amplitude quotient', A, by which the Gaussian amplitude must be multiplied in order to match the true amplitude. The amplitude quotient is therefore just the estimation of the parametrix at zero spatial displacement. While the principal aim of this note and its sequel is to provide a systematic approach to formulating the asymptotic formulae by which such an amplitude quotient can be approximated, a possible by-product of the proposed procedure, which we shall not develop explicitly, is a corresponding method for estimating the amplitudes of the covariances of the vector derivatives of the primary variables. This might provide a valuable capability when the primary variables are streamfunction or velocity potential; the gradient fields, combining to form the wind, are then more closely associated with the measurements than are the primary analysis variables themselves.

Assuming, as we always shall, that the diffusivity varies smoothly, then one way to formulate an approximation to the adjustment factor is to express it as a power series expansion in a 'diagnostics vector', whose components are $\boldsymbol{D}$ together with local spatial derivatives of $\boldsymbol{D}$ up to a given degree of truncation. This is a valid approach, and is approximately similar to the one we propose here. However, the diagnostics vector, as defined above, always contains demonstrably more components than are relevant to the specific task of estimating the amplitude. By redefining a shorter diagnostics vector in a way that excludes the extraneous information the computational cost of evaluating the amplitude adjustment factor in the space of this diagnostics vector can be greatly reduced.

The method by which we winnow the diagnostics vector down to a smaller size exploits certain invariances that emerge from a consideration of the appearance of the inhomogeneous diffusion equation when the coordinates are smoothly transformed. The tensor analysis of non-Euclidean 'Riemannian' spaces (e.g., Kreyszig 1991) provides the necessary algebraic machinery for carrying out these manipulations. Conceptually, the idea is to interpret the given diffusivity tensor as a metric tensor of the analysis domain's space, now regarded as intrinsically curved. Then the effective diffusivity consistent with this distorted-metric interpretation of space is rendered both isotropic (invariant to in situ rotation) and homogeneous (invariant to arbitrary spatial displacements). However, in general, the transformed diffusion process now includes a varying effective 'capacitance' of the space, which subtly alters the form of the diffusion

4

equation. The capacitance is somewhat analogous to a density or specific heat capacity in diffusive heat flow. The derivatives of this capacitance significantly affect the amplitude quotient and, when present, must therefore be included in the diagnostics vector.

In some respects, the new formulation will lead to a significant simplification in the construction of the diagnostics vector for this problem. First, since all the components of the new diffusivity tensor correspond to the identity matrix in any locally Cartesian coordinate system, these components are redundant as far as the diagnostics vector is concerned. Second, when we relate the derivatives of the new diffusivity at a point O to an important special class of such coordinates known as (Riemann) 'normal coordinates' whose axes are orthogonal in the new metric and whose straight radials through O all correspond to O-intersecting geodesics, then even the first derivatives of the new diffusivity can be shown to vanish. Thus, by implicitly adopting normal coordinates around each point to express spatial derivatives, all essential features of the transformed diffusivity distribution relevant to the amplitude adjustment factor involve only derivatives of the metric of degree two or higher, and derivatives of the capacitance. The information in the second derivatives of the metric, as expressed in normal coordinates, is equivalent to the information contained in the 'Riemann' (or, sometimes, 'Riemann-Christoffel') curvature tensor. Correspondingly, the information associated with all derivatives of the metric up to degree, $m + 2$, is essentially equivalent to that contained in derivatives, up to degree m, of the curvature tensor. These facts, together with an exploitation of known symmetries of the curvature tensor in two and three dimensions, ensure that the size of the diagnostics vector that encapsulates derivative information of the metric up to degree $m + 2$ can be kept to a minimum. Moreover, if we use our discretionary freedom in formulating the original diffusivity model to ensure that the capacitance term in the transformed problem is constant, then the diagnostics vector consists *only* of these metric terms and the procedure for computing asymptotic approximations is significantly simplified.

Rather than plunging into the full complexity of the general problem in N dimensions, our approach here will be to introduce in the next section the form of the diffusion equation as it appears transformed into a non-Euclidean geometry, then to illustrate in section 3 the amplitude adjustment puzzle in a family of specific idealized cases where the transformed representation of the geometry implies a locally uniform and isotropic curvature. When the curvature is positive, the prototypical geometry is that of the 'N-sphere' (the 2-sphere being the familiar spherical surface; note that the 'body' of such a sphere, being three-dimensional in its familiar manifestation, is of geometrical irrelevance here). However, it is worth noting that, for each dimensionality, N, there also exists a topologically distinct isotropic geometry, the 'projective space', whose curvature is identical to that of the N-sphere but whose measure (area, volume, etc.) is exactly half that of the N-sphere; this ensures that the finite time solution of the diffused-impulse problem is *different* in the two geometries, even though their local curvature diagnostics are identical. This circumstance implicitly informs us that the amplitude adjustment expansion is of the asymptotic kind whose convergence at finite time t cannot be guaranteed in general and which, if it *is* convergent, cannot be expected to converge exactly to the true amplitude quotient.

Section 4 considers the simplest 2D generalization beyond the purely spherical case; namely, the case of a manifold radially symmetric about a special pivot point. We use this 'axisymmetric' geometry as the vehicle to introduce and illustrate the iterative solution algorithm.

Further refinements needed to accommodate the full set of geometrical terms do not change the essential features of this asymptotic procedure, only its algebraic details.

In section 5 we test and compare the methods we have developed in two dimensions for degrees of asympototic expansion, zero, one and two, using for the 'control' the method which does not use the new non-Euclidean geometrical approach but just the standard Gaussian amplitude appropriate to the local value of the aspect tensor under the assumption that the metric is Euclidean. The tests are conducted, firstly, for a family of distributions of the aspect tensor that might be considered appropiate for the case of an intense symmetric large-scale vortex, such as an idealized hurricane, in which the variations in the aspect tensor parameters are of a regular and systematic kind. Then a second series of tests is carried out for a family of smooth but random idealized aspect tensors synthesized over a doubly-periodic region where we are able to exert some control over the amplitude and scales of the variability of the aspect tensor field we are synthesizing. The insights provided by these experiments are discussed in the concluding section 6.

2. DIFFUSION IN NON-EUCLIDEAN SPACE

The usual form of the diffusion equation relates to the case in which an area or volume integral of diffused substance is conserved. Where the diffusivity is a positive-definite symmetric, but generally not isotropic, spatially varying tensor, this equation expresses the rate of change of the concentration of the substance as the negative-divergence of its flux, and the flux as the diffusivity operator applied to the negative-gradient of the substance. Applying the standard tensor notation we therefore have:

$$\frac{\partial P}{\partial t} = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \sqrt{g} D^{ij} \frac{\partial P}{\partial x^j}, \tag{2.1}$$

where g is the determinant of the covariant metric tensor, $g_{\bullet\bullet}$, and $\sqrt{g}$ therefore the 'measure' per unit coordinate volume. We emphasize again that 'time' parameter, t, is *not* to be interpreted as a physical time in this equation. If we now interpret diffusivity, $D^{\bullet\bullet}$, instead of the physical distance metric, $g^{\bullet\bullet}$, as the *effective* metric in terms of which the diffusion is isotropic and of unit magnitude everywhere, and write D as the determinant of the covariant diffusivity, $D_{\bullet\bullet}$, then we can only generally reconcile the transformed (2.1) with a diffusion equation of slightly more general form. The more general form involves a spatially varying 'capacitance', $J = \sqrt{g/D}$ for the diffused substance:

$$\frac{\partial P}{\partial t} = \frac{1}{J\sqrt{D}} \frac{\partial}{\partial x^i} J \sqrt{D} D^{ij} \frac{\partial P}{\partial x^j}. \tag{2.2}$$

so that the conserved quantity is the product JP, integrated with respect to the effective new area or volume measure, $\sqrt{D} \prod_i dx^i$. A mental model in two dimensions is the case of diffusion of heat within a conducting thin shell of uniform composition but variable gauge, where the capacitance is now the relative thickness; in three dimensions, the analogy of a specific heat capacity per unit volume, or an effective 'density', might serve to explain the role played by the capacitance.

6

By our transformation we have gained a simplification in the diffusion tensor (it reduces to the identity) but entailed a complication in the appearance of the capacitance. However, our objective is not to model actual physical diffusion, but only to mimic its process as a means of producing properly posed covariances. The question of existence or nonexistence of a capacitance term is therefore not of practical relevance. Reviewed in this light, practical objectives are better met by exercising the freedom to choose the form of diffusion in which the distribution of capacitance in the original (non-uniform diffusivity) physical-space realization of the diffusion problem is exactly apportioned so as to render it uniform once transformed to the final (uniform diffusivity) realization. Exercising this choice removes the last vestige (the inconvenient J term) of the 'physical' metric from the final diffusion equation, (2.2). The resulting diffusion equation is an elegant prescription for generating the covariance (or this quasi-Gaussian piece of the covariance) having the uniquely special property that the solution of this equation no longer depends upon the accidental choice of the physical coordinates in which the intended tensorial spread, or 'aspect tensor' (Purser et al. 2003) happened to be expressed. But, without any further involvement of the purely physical metric terms in the new statement of the diffusion problem which this note and its sequel are exclusively concerned with, there remains no good reason for us unnecessarily to duplicate here the functionally identical tensorial notations that correspond to the 'physical metric' quantities which we have referred to as $g^{\bullet\bullet}$, $g_{\bullet\bullet}$, $\sqrt{g}$, etc., on the one hand, and 'diffusivity metric' quantities we have been calling, $D_{\bullet\bullet}$, etc. Thus, henceforth it is to be tacitly understood that the metric quantities referred to in the conventional (g-variables) metric notation are invariably the diffusivity metric terms. We therefore end up stating the diffusion problem, in which $g^{\bullet\bullet}$ is now identified as the original contravariant diffusivity, simply as:

$$\frac{\partial P}{\partial t} = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \sqrt{g} g^{ij} \frac{\partial P}{\partial x^j}. \tag{2.3}$$

The elliptic operator on the right-hand side of (2.3) is the familiar Laplace (or 'Laplace-Beltrami') operator, and the impulse-response solutions to this simplest form of Riemann-space diffusion equation are referred to as 'heat kernels' (Rosenberg 1997).

Given an infinitely differentiable (C_∞) diffusivity, $g^{\bullet\bullet}$ playing the role of the metric, and any point on the Riemannian manifold, a local normal coordinate system of the special kind alluded to in the Introduction can be set up having this chosen point as the origin. At this origin, and in these new coordinates, the transformed covariant and contravariant metric tensors reduce to the identity matrix. Also, by virtue of the vanishing of the metric's covariant derivatives, the first order partial derivatives of both these metric tensors vanish at this origin. Expressing the covariant metric tensor as a Taylor series then, within a small enough neigborhood of the origin, the metric tensor is expressible to arbitrary accuracy using a Taylor series whose 'constant' terms form the identity and whose first-order terms vanish, leaving the first nontrivial terms in the expansion to be quadratic in the normal coordinates. Intuitively, it follows that the short-time solution to the heat-kernel implied by this neighborhood distribution should physically be a slight perturbation to the Gaussian heat kernel solution of Euclidean geometry, provided the respective normal and Euclidean coordinates of the two problems are brought into formal coincidence for the purposes of this comparison. The vindication of this intuition can be

proved rigorously (Rosenberg 1997) through a description of the Riemannian geometry solution that involves a smooth function multiplying the Euclidean solution.

We have seen that, for unit isotropic diffusivity in Euclidean N-space the central amplitude for the solution to the diffused-impulse evolves according to $P(t,0) = (4\pi t)^{-N/2}$. In the next section we examine the central amplitude of the diffused impulse in non-Euclidean spaces of N dimensions with uniform and isotropic curvature through an application of Mercer's theorem (Mercer 1909; which states that a positive-definite symmetric integral kernel can be represented as the convergent summation of products of a complete set of orthogonal eigenfunctions). The geometrical simplification allows us to exploit properties of the analytically-known eigenfunctions of the Laplacian, and their eigenvalues, for the modes of zonal symmetry about the central point at which the initial impulse is imposed. The resulting modal analysis of the solution leads both to exact solutions, and to asymptotic expansions of the amplitude quotient (actual central amplitude divided by the corresponding Euclidean space amplitude). The comparison between true and approximate solutions carried out to different degrees yields several notable insights into what can be expected from the expansions that pertain to more general geometries.

3. Amplitude quotients for diffusion in the isotropic spaces of constant curvature

(a) Positive curvature

The surface harmonics on the unit N-sphere with 'zonal' symmetry form a discrete spectrum. They are the Gegenbauer polynomials, $C_n^{(\alpha)}(\cos r)$, in $z = \cos r$ (cosine of the geodesic radial distance) where $\alpha = (N-1)/2$ (Abramowitz and Stegun 1970, henceforth denoted 'AS', Chap. 22). For the ordinary 2-sphere, $N = 2$, these are equivalent to the Legendre polynomials. In the general case, the indexing n runs through all nonnegative integers, and the eigenvalue corresponding on the N-sphere to the application of the negative-Laplacian, is:

$$\lambda = n^2 + (N-1)n = n^2 + 2\alpha n. \tag{3.1}$$

(There is actually a duplication of the modes' indices, since (3.1) is a quadratic in n, but, as we discuss below, it is important that the modes are not inadvertently counted twice in computations where the assumption of their mutual orthogonality is critical.)

The conventional standardization as defined in AS is such that the polar value:

$$C_n^{(\alpha)}(1) = \binom{n+N-2}{n}. \tag{3.2}$$

For Legendre harmonics, $N = 2$, this convention already conveniently confers a unit amplitude to all harmonics at the pole, $C_n(1) = 1$, and it is desirable to extend this convenience to *all* N-spheres. Therefore, we define the rescaled Gegenbauer polynomial,

$$\hat{C}_n^{(\alpha)}(z) = \binom{n+N-2}{n}^{-1} C_n^{(\alpha)}(z). \tag{3.3}$$

Pairs of distinct harmonics are orthogonal with respect to integration over the domain. If $S_{(N-1)}$ denotes the integrated measure over a unit $(N-1)$-sphere, which we derive amongst

the identities discussed in Appendix A, then, since the shape of a hypersurface of constant r is always an $(N-1)$-sphere of effective radius, $(1-z^2)^{1/2} = \sin r$, the orthonormality of the zonal harmonics takes the form:

$$\int_{-\pi/2}^{\pi/2} S_{(N-1)}(\sin r)^{N-1} \hat{C}_n(\cos r)\hat{C}_m(\cos r)\, dr = \delta_{nm} h_n^{(N)}. \tag{3.4}$$

The normalization constants, h_n, need to be modified from those given in AS (Chap. 22, p. 774) in order to include both the latitude integration term, $S_{(N-1)}$ and the square of the rescaling factor, that is:

$$h_n^{(N)} = \frac{S_{N-1}\,\pi\,2^{(2-N)}(n+N-2)!}{n!\left(n+\frac{N}{2}-\frac{1}{2}\right)\left[\left(\frac{N-3}{2}\right)!\right]^2}\left[\frac{n!(N-2)!}{(n+N-2)!}\right]^2.$$

After substitution for the constant, $S_{(N-1)}$ and application of the half-integer factorial identities, which are all discussed in Appendix A, we recover a slightly simpler expression:

$$h_n^{(N)} = \frac{2^N \pi^{(N/2)} n! \left(\frac{N-2}{2}\right)!}{(2n+N-1)(n+N-2)!}. \tag{3.5}$$

Thus, for the 2-sphere,

$$h_n = \frac{2\pi}{n+\frac{1}{2}}, \tag{3.6}$$

while for the 3-sphere,

$$h_n = \frac{2\pi^2}{(n+1)^2}. \tag{3.7}$$

If, for time t and a sine-latitude, z, a zonally-symmetric function $f(z)$ has the expansion,

$$f(t,z) = \sum_{n=0}^{\infty} c_n(t)\hat{C}_n(z), \tag{3.8}$$

then

$$\int f\hat{C}_n = h_n c_n(t), \tag{3.9}$$

where the integral symbol is implicitly taken to mean with respect to measure over the entire N-sphere (as is done in explicit detail in (3.4)). Suppose initially $(t=0)$ we take f to be a unit delta function impulse at the pole. Then the integral simply evaluates the other factor of the integrand, and hence:

$$\hat{C}_n(1) = 1 = h_n c_n(0). \tag{3.10}$$

Now consider the evolution of this impulse under the influence of unit-diffusivity isotropic diffusion in spectral terms. Recalling the interpretation of the negative-Laplacian as the spectral eigenvalue, we find that:

$$c_n(t) = c_n(0)\exp\left(-t(n^2 + (N-1)n)\right), \tag{3.11}$$

9

and hence, the evaluation of the reconstructed function at time t is:

$$f(t, z) = \sum_{n=0}^{\infty} \frac{\hat{C}(z)}{h_n} \exp\left(-t(n^2 + (N-1)n)\right), \tag{3.12}$$

and, in particular, its polar value is:

$$f(t, 1) = \sum_{n=0}^{\infty} \frac{1}{h_n} \exp\left(-t(n^2 + (N-1)n)\right). \tag{3.13}$$

By comparison, the standard amplitude of the Gaussian resulting from the same duration of diffusion on the Euclidean N-space is,

$$\hat{f}(t, 1) = \frac{1}{(4\pi t)^{N/2}}. \tag{3.14}$$

In comparing these amplitudes, it is most convenient to adopt a parameter rescaling the 'time',

$$K = 2t, \tag{3.15}$$

since this makes the 'spread' (centered second moment in each cartesian direction) equal to K in the Euclidean case, and the amplitude quotient:

$$A(K) = f(K/2, 1)/\hat{f}(K/2, 1), \tag{3.16}$$

then gives us the cleanest comparison of the distinctions between results on the curved and on the Euclidean spaces. We can readily see that, for an N-sphere of *arbitrary* curvature, the factor $A(K)$ can alternatively be interpreted as providing the adjustment factor corresponding to diffusion for the 'standard' duration of $1/2$ (unit Euclidean spread) when the sectional or Gaussian curvature of the N-sphere is just that parameter K.

In 2D, we see that:

$$A(K) = \sum_{n=0}^{\infty} K\left(n + \frac{1}{2}\right) \exp\left(-\frac{1}{2}K(n^2 + n)\right). \tag{3.17}$$

This is a form that suggests an application of the one-sided Euler-Maclaurin (EM) summation formula, since the summand for *continuous* positive n looks like a smooth function on the scale of the unit 'grid' upon which the summation occurs, when $K \to 0$. Informally, the EM formula applied to a case like this states that:

$$I = \sum_{n=0}^{\infty} F(n) \approx I^{(k')} = \int_0^{\infty} F(x)\, dx + \frac{1}{2}F(0) - \sum_{k=1}^{k'} \frac{B_{2k}}{(2k)!} F_{(0)}^{(2k-1)}, \tag{3.18}$$

where B_{2k} are the even-index Bernoulli numbers (AS, p.810), and $F_{(0)}^{(m)}$ in the last term refers to the mth derivative of F at $x = 0$. The first few terms of $B_{2k}/(2k)!$ for $k = 1$, 2, etc., are: $1/12$, $-1/720$, $1/30240$, $-1/1209600$. Application to the 2D case is made fairly straightforward

10

by the fact that the summand of (3.17) is so obviously a simple differential of the negative of the exponential term. It is relatively easy to mechanize the evaluation of the successive terms, which gives us the divergent series:

$$A(K) \approx 1 + K/6 + K^2/60 + K^3/630 + K^4/5040 + \ldots \tag{3.19}$$

The 3D case is complicated by the presence of fractional powers of K in the original formula and in all the expansion terms that involve the Bernoulli numbers in the EM method. Another serious complication is the fact that the summand is not now such a simple function to integrate. We find now that:

$$A(K) = \frac{2^{\frac{1}{2}} K^{\frac{3}{2}}}{\pi^{\frac{1}{2}}} \sum_{n=0}^{\infty} (n+1)^2 \exp\left(-\frac{1}{2} K(n^2 + 2n)\right). \tag{3.20}$$

This time, let us set

$$F(x) = (x+1)^2 E(x), \tag{3.21}$$

where,

$$E(x) = \exp(-\frac{1}{2} K(x^2 + 2x)), \tag{3.22}$$

and apply the EM formula to this $F(x)$.

In order to perform the integral, use the substitution,

$$z = \frac{K}{2}(x+1)^2, \tag{3.23}$$

hence,

$$dx = \frac{dz}{(2Kz)^{\frac{1}{2}}}, \tag{3.24}$$

to find

$$
\begin{aligned}
J &= \int_0^{\infty} F(x)\, dx \\
&= \frac{2^{\frac{1}{2}}}{K^{\frac{3}{2}}} \exp\left(\frac{K}{2}\right) \Gamma\left(\frac{3}{2}, \frac{K}{2}\right).
\end{aligned}
\tag{3.25}
$$

As discussed in AS (Chap. 6, p. 262), the incomplete Gamma function, Γ, while not itself amenable to well-behaved power expansion directly, can be written in terms of a modified function γ^*, that *is* perfectly well-behaved in this sense:

$$\Gamma\left(\frac{3}{2}, \frac{K}{2}\right) = \frac{\pi^{\frac{1}{2}}}{2}\left(1 - \left(\frac{K}{2}\right)^{\frac{3}{2}} \gamma^*\left(\frac{3}{2}, \frac{K}{2}\right)\right), \tag{3.26}$$

with

$$\gamma^*\left(\frac{3}{2}, \frac{K}{2}\right) = \exp(-\frac{K}{2}) \sum_{n=0}^{\infty} \frac{\left(\frac{K}{2}\right)^n}{\left(\frac{3}{2} + n\right)!}, \tag{3.27}$$

11

(summation here is convergent), and hence,

$$J = \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \frac{\exp\left(\frac{K}{2}\right)}{K^{3/2}} - \frac{\pi^{\frac{1}{2}}}{4} \sum_{n=0}^{\infty} \frac{\left(\frac{K}{2}\right)^n}{\left(\frac{3}{2}+n\right)!}. \tag{3.28}$$

The terms we collect into the expansion approximating $A(K)$ are now of two distinct types: those in whole integer powers of K, which come entirely from the expansion of:

$$A_1(K) = \exp\left(\frac{K}{2}\right); \tag{3.29}$$

and those in half-odd-integer powers of K, which come from the remaining term of J together with all the other terms left over from the EM summation:

$$A_2(K) = \frac{2^{\frac{1}{2}} K^{3/2}}{\pi^{\frac{1}{2}}} \left[\frac{1}{2}F(0) - \sum_{k=1}^{k'} \frac{B_{2k}}{(2k)!} F_{(0)}^{(2k-1)}\right] - \left(\frac{K}{2}\right)^{3/2} \sum_{n=0}^{\infty} \frac{\left(\frac{K}{2}\right)^n}{\left(\frac{3}{2}+n\right)!}. \tag{3.30}$$

From general considerations of the smoothness of the modulating function from which the non-Euclidean solution is obtained as the modulation of the corresponding Euclidean solution, the asymptotic expansion for $A(K)$ for small K should be of whole-integer powers. We can therefore immediately conclude (and an explicit examination of the first few terms will corroborate it) that the half-odd-integer terms all cancel out!

We are left with a paradoxical result that the asymptotic expansion for $A(K)$ in 3D is, term by term, identical with the *convergent* Taylor expansion for the exponential function,

$$A(K) \approx \hat{A}(K) = \exp(K/2), \tag{3.31}$$

although it is quite clear from the behavior of $A(K)$ for large K that these functions A and $\hat{A}$ cannot be identical. Numerical models of the diffusive process along a meridian of the 3-sphere confirm the consistency of the first few terms, viz:

$$A(K) \approx 1 + K/2 + K^2/8 + K^3/48 + \dots \tag{3.32}$$

The right halves of Figs. 1 and 2 provide some graphical comparisons between the exact calculation of the amplitude quotient, $A(K)$, and asymptotic approximations to it, $\hat{A}(K)$, for positive K out to $+5$. Naturally, the discrete spectra of the positive curvature cases, from which a summation results, can no longer apply when the geometries are of the negative curvature 'hyperbolic' types. The next subsection describes how a modification of the modal procedure can be made to obtain integral representations of the amplitude quotients in these cases.

(*b*) *Negative curvature*

In the hyperbolic spaces the same general principles apply except we end up with a continuous spectrum to integrate over. It is conceptually helpful to approach this continuous case by
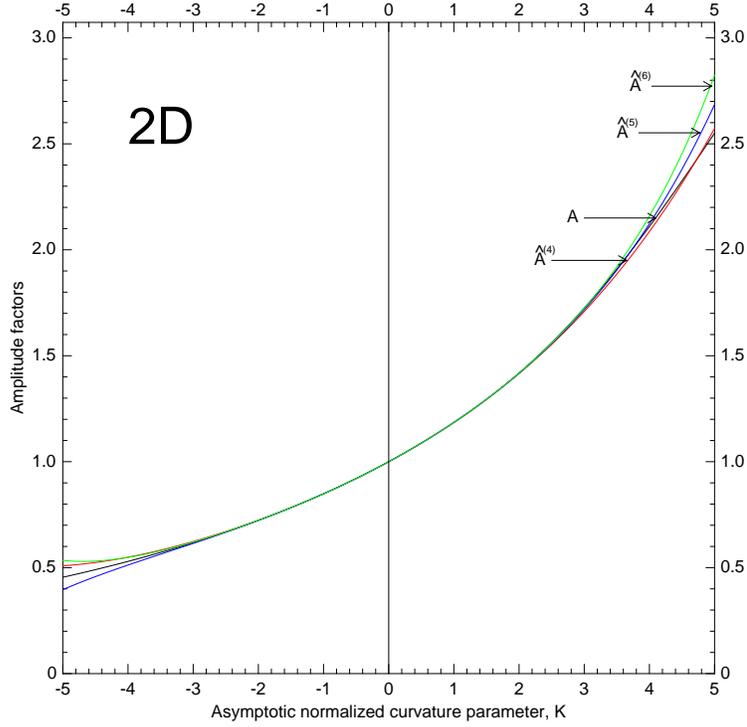
Figure 1.   Comparison between the true 2D amplitude adjustment factor $A(K)$ and some of the the asymptotic approximations to it, $\hat{A}^{(p)}(K)$, where p denotes the degree to which the divergent expansion is truncated and summed. The parameter K can be interpreted as the Gaussian curvature when the diffusivity is unity, and the duration of the diffusive process is one half; this ensures that, in the planar case (with which comparisons are made) the second-moment 'spread' of the diffused impulse is one unit.

evolving the spectrum associated with a limiting progression of finite geodesic-radius domains of radius $R \to \infty$, with Neumann conditions on the modes at R becoming a condition of boundedness on the zonal integration of the squared-amplitude. The eigenfunctions of the Laplacian on the hyperbolic N-space at a radial distance $r < R$ are still essentially the same family of Gegenbauer functions, $\hat{C}_n^{(\alpha)}(z)$, with $\alpha = (N-1)/2$, as in the corresponding positive-curvature geometries. However, we are now interested in the argument range, $z \geq 1$, since, in hyperbolic geometry, this argument relates to geodesic radius r according to $z = \cosh r$ (instead of $\cos r$ for spherical geometries). Also, the negative-Laplacian, whose hyperbolic geometry radial part is:

$$-\nabla^2 \equiv -\frac{1}{(\sinh r)^{(N-1)}}\frac{\partial}{\partial r}(\sinh r)^{(N-1)}\frac{\partial}{\partial r}, \qquad (3.33)$$

has non-negative eigenvalues, λ, that relate to the index n in a formula like (3.1), but now with the opposite sign:

$$\begin{aligned} \lambda &= -[n^2 + (N-1)n] \\ &= -[n^2 + 2\alpha n]. \end{aligned} \qquad (3.34)$$

As with positive curvature geometry, we must recognize that, since (3.1) and (3.34) are quadratic in n the eigenmode of a given λ is indexed redundantly by the two equally valid

13
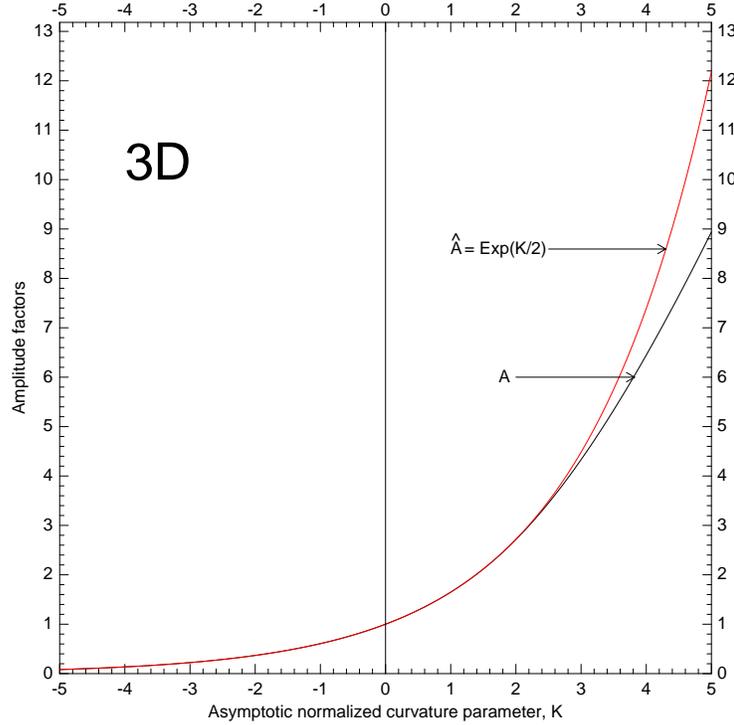
Figure 2. Comparison between the true amplitude adjustment factor $A(K)$ and the asymptotic approximation to it, $\hat{A}(K)$, carried out to an infinite number of terms for the 3D case. This is possible because the asymptotic expansion is convergent in this case. The function to which it converges is $\exp(K/2)$.

choices of possible n and we must avoid double-counting in order to maintain modal orthogonality. Clearly, our new spectra will involve complex-valued indices, n, in general, and the modes themselves must now be referred to as Gegenbauer *functions*, since they are no longer *polynomials*, when n is complex. Inspection of the form of (3.33) at $r \gg 1$ reveals that the zonally-integrated 'energy norm' function,

$$|\hat{C}_n(\cosh r)|^2 (\sinh r)^{(N-1)}, \tag{3.35}$$

grows exponentially with r for those *non-oscillatory* modes whose values of $\lambda < \alpha^2$ allow the index n to remain real. These modes would therefore be excluded from consideration as $R \to \infty$ in any reasonable choice of limiting condition.

We can therefore adopt a re-parameterization,

$$n = -\alpha + ik, \tag{3.36}$$

with real-valued k and, to avoid double-counting, assume at this point that $k > 0$. We shall still have $\lambda = 0$ as a special isolated eigenvalue associated with the trivial (constant-valued) eigenmode, but a vanishingly small proportion of the initial impulse projects onto it in the limit $R \to \infty$, so it can also be excluded from practical consideration. We are left with the continuum whose eigenvalues are bounded away from zero:

$$\lambda = \alpha^2 + k^2. \tag{3.37}$$

14

At $r \gg 1$ the new parameter k is approximately the wavenumber, but of waves that are exponentially decaying in amplitude (although tending asymptotically to uniform period-averaged energy-norm function in the sense of (3.35)). For a large R at which the Neumann condition is applied, the discrete values of k therefore tend to occur at a regular spacing,

$$\Delta k \approx \frac{\pi}{R}, \tag{3.38}$$

or more precisely,

$$\lim_{R \to \infty} (k_{j+1} - k_j)R = \pi, \tag{3.39}$$

where j is the index counting the total number of zeroes of the eigenmode.

A standard identity (AS, Chap. 15, p. 561) expresses the complex-indexed Gegenbauer function more conveniently as a hypergeometric function:

$$\hat{C}_n^{(\alpha)}(z) = \binom{n+N-2}{n}^{-1} C_n^{(\alpha)}(z) \; = \; F\left(-n, n+2\alpha; \alpha + \frac{1}{2}; \frac{1-z}{2}\right)$$

$$= \; F\left(\alpha - ik, \alpha + ik; \frac{N}{2}; \frac{1-z}{2}\right). \tag{3.40}$$

The large-z behavior of each solution is then revealed by one of the so-called 'linear transformation formulas' (AS, Chap 15, p 559) that transform the Gegenbauer function's limiting real argument, $z \to \infty$, to zero arguments of a 'conjugate' pair of complex hypergeometric functions. We can then use the smooth differentiability of the hypergeometric function in the neighborhood of where $F(a, b; c; 0) = 1$ to argue about the large-z behavior.

The transformation we need may be written:

$$\hat{C}_n^{(\alpha)}(z) = F^{(+)}(z) + F^{(-)}(z), \tag{3.41}$$

with the two terms themselves being eigen-solutions to our radial wave equation (albeit complex ones) that are mutually orthogonal in the Hermitian sense. These new solutions share the form:

$$F^{(\pm)}(z) = M^{\pm}(z)G^{\pm}F\left(\alpha \mp ik, \frac{1}{2} \mp ik; 1 \mp 2ik; \frac{2}{1+z}\right), \tag{3.42}$$

where, as stated, the third factors (the 'Fs') in each product tends to unity in the limit of $z \to \infty$. The middle factors, $G^{\pm}$, are the constants:

$$G^{\pm} = \frac{\Gamma\left(\frac{N}{2}\right)\Gamma(\pm 2ik)}{\Gamma(\alpha \pm ik)\Gamma(\frac{1}{2} \pm ik)}, \tag{3.43}$$

whose squared absolute magnitudes can be simplified by applying the identities (A.7), (A.8a) and (A.8b) of Appendix A. Each first factor of (3.42) is the function of radius that defines the way the mode's phase and amplitude in the asymptotic limit of interest:

$$M^{\pm}(z) = \left(\frac{1+z}{2}\right)^{-\alpha \pm ik}, \tag{3.44}$$

15

which, using the large-z approximation, $z = \cosh r \approx \exp(r)/2$, becomes approximately:

$$M^\pm(z) \approx 2^{2\alpha} \exp(-\alpha r) \exp[\pm ik(r + \ln 4)], \tag{3.45}$$

confirming the role of k as the effective wavenumber and α as the asymptotic radial decay-rate of amplitude. If we define $h_{k,R}^{(\alpha)}$ for each $k > 0$ to be the integral of the square of the function $\hat{C}_{(-\alpha+ik)}^{(\alpha)}$, or equivalently, twice the squared-absolute magnitude of $F^{(+)}$, over the domain out to a boundary at $r = R$, and renormalized by R itself:

$$h_{k,R}^{(\alpha)} = \frac{1}{R} \int_0^R S_{(2\alpha)} (\sinh r)^{2\alpha} 2 \left|F^{(+)}\right|^2 dr, \tag{3.46}$$

and make the usual large-r exponential substitutions for $\sinh r$ and $\cosh r$, we find that

$$
\begin{aligned}
h_k^{(\alpha)} &= \lim_{R \to \infty} h_{k,R}^{(\alpha)} \\
&= \lim_{R \to \infty} \frac{1}{R} \int_0^\infty S_{(2\alpha)} \left(\frac{\exp(r)}{2}\right)^{2\alpha} 2 \left(2^{4\alpha} \exp(-2\alpha r)\right) |G^+|^2 dr \\
&= 2^{2\alpha+1} S_{(2\alpha)} |G^+|^2.
\end{aligned}
\tag{3.47}
$$

In $N = 2$ dimensions, $\alpha = 1/2$, $S_{(1)} = 2\pi$, and the $|G^+|^2$ term evaluates to

$$
\begin{aligned}
|G^+|^2 &= \frac{|\Gamma(2ik)|^2}{\Gamma(\frac{1}{2} + ik)|^4} \\
&= \frac{1}{4\pi k \tanh(\pi k)}.
\end{aligned}
\tag{3.48}
$$

Hence,

$$h_k = \frac{2}{k \tanh(\pi k)}. \tag{3.49}$$

By direct analogy with (3.13), and including the factor, $1/\pi$, from the asymptotic density of wavenumbers, we obtain the corresponding formula but with the summation taken to the limit of an equivalent integral:

$$
\begin{aligned}
f(t, 1) &= \int_0^\infty \frac{1}{\pi h_k} \exp(-t/4) \exp(-tk^2) dk \\
&= \exp(-t/4) \frac{1}{2\pi} \int_0^\infty \exp(-tk^2) k \tanh(\pi k) dk,
\end{aligned}
\tag{3.50}
$$

or, integrating by parts,

$$f(t, 1) = \frac{1}{4t} \exp(-t/4) \int_0^\infty \exp(-tk^2) \frac{1}{[\cosh(\pi k)]^2} dk. \tag{3.51}$$

Now we must multiply by $\hat{f}(t, 1)$, and substitute,

$$K = -2t, \tag{3.52}$$

16

to express the amplitude quotient for negative curvature parameter, K, as the final integral representation:

$$A(K) = \exp(K/8) \int_0^\infty \exp\left(\frac{Kk^2}{2}\right) \frac{\pi}{[\cosh(\pi k)]^2} dk. \tag{3.53}$$

This integral represention is used to furnish the solution on the left half of the Fig. 1. But by formal differentiation of it at $K = 0$, and by evaluating the resulting successive k^{2p}-moments of the $\cosh^{-2}$ term, we acquire another way of generating the coefficients of the asymptotic series, (3.19).

In $N = 3$ dimensions, $\alpha = 1$, $S_{(2)} = 4\pi$, and $|G^+|^2$ evaluates to:

$$
\begin{aligned}
|G^+|^2 &= \frac{\left|\Gamma\left(\frac{3}{2}\right)\right|^2 |\Gamma(2ik)|^2}{|\Gamma(1+ik)|^2 \left|\Gamma\left(\frac{1}{2}+ik\right)\right|^2}, \\
&= \frac{1}{16k^2}.
\end{aligned}
\tag{3.54}
$$

Hence,

$$h_k = \frac{2\pi}{k^2}. \tag{3.55}$$

The 3D integral form for $f(t, 1)$ now becomes

$$
\begin{aligned}
f(t, 1) &= \int_0^\infty \exp(-t) \exp(-tk^2) \frac{k^2}{2\pi^2} dk, \\
&= \exp(-t)(4\pi t)^{-3/2},
\end{aligned}
\tag{3.56}
$$

and hence, we find, after multiplying by $\hat{f}(t, 1) = (4\pi t)^{3/2}$ and substituting $K = -2t$, that the true amplitude quotient depicted on the left half of Fig. 2 for $K < 0$ is *exactly*:

$$A(K) = \exp(K/2), \tag{3.57}$$

which the asymptotic series converges to.

Examination of the procedure we have set out here as it applies to spaces of higher dimensions shows that, for even dimensionality, N, the asymptotic expansion is of the nonconvergent type with the true amplitude quotient's integral representation involving the 'Gaussian' integral with hyperbolic functions; for odd dimensions, N, the integral representation reduces to simple polynomials to which the asymptotic expansion converges.

4. Diffusion on a smooth 2D manifold with axial symmetry

(a) *Construction of the parametrix expansion for this problem*

The solutions we have obtained in the last section are very special, but serve as valuable benchmarks for the next level of generalizations to which we want to proceed. This section describes the simplest next level of refinement for the 2D case, where gradients of curvature are present, but only in the very restricted manner that a constraint of axial symmetry permits.

While we do obtain a new and significant coefficient of the general expansion (the coefficient giving the principal effect of *second derivatives* of curvature, the first derivatives being unimportant at this level of expansion, as Part II will demonstrate), the main motivation is to present, in its simplest possible form, the structure of the more general iterative algorithm by which the theoretical influence of *any* combination of the possible multinomial contributions of curvature, combined with derivatives of curvature, can be revealed in principle.

The planar solution of the diffusion equation in radial coordinate r:

$$\frac{\partial P_0}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial P_0}{\partial r},\tag{4.1}$$

is:

$$P_0 = \frac{1}{4\pi t}\exp\left(-\frac{r^2}{4t}\right),\tag{4.2}$$

as we readily verify from the individual contributing terms:

$$\frac{\partial P_0}{\partial t} = \left(\frac{r^2}{4t^2} - \frac{1}{t}\right)P_0,\tag{4.3a}$$

$$\frac{\partial P_0}{\partial r} = -\frac{r}{2t}P_0,\tag{4.3b}$$

$$\frac{\partial^2 P_0}{\partial r^2} = \left(\frac{r^2}{4t^2} - \frac{1}{2t}\right)P_0.\tag{4.3c}$$

When the manifold has r-dependent Gaussian curvature, the diffusion equation generalizes to:

$$\frac{\partial P}{\partial t} = \frac{1}{g^{\frac{1}{2}}}\frac{\partial}{\partial r}g^{\frac{1}{2}}\frac{\partial P}{\partial r}.\tag{4.4}$$

The function, g, is the metric in the definition of infinitesimal transverse distance,

$$ds^2 = dr^2 + gd\theta^2,\tag{4.5}$$

which we expand explicitly out to fourth degree in r:

$$g = Hr^2 = \left(1 - \frac{Kr^2}{3} + Nr^4\right)r^2 + \mathcal{O}(r^6),\tag{4.6}$$

where K is the Gaussian curvature at the origin, N is the next term, involving a combination of K^2 and the second derivative of Gaussian curvature at the origin. The role of g in the diffusion equation context is to supply the following factor required when (4.4) is expanded:

$$\frac{\partial \ln(g^{\frac{1}{2}})}{\partial r} = \frac{1}{r} - \frac{Kr}{3} + \left(2N - \frac{K^2}{9}\right)r^3 + \mathcal{O}(r^5).\tag{4.7}$$

It will be algebraically slightly less complicated in the development below to adopt the alternative parameterization,

$$\frac{\partial \ln(g^{\frac{1}{2}})}{\partial r} = \frac{1}{r}(1 + 2Dr^2 + 2Er^4) + \mathcal{O}(r^4),\tag{4.8}$$

18

and to note that, in the important special case of a *uniform* Gaussian curvature,

$$g^{\frac{1}{2}}_s = \frac{\sin(r\,K^{\frac{1}{2}})}{K^{\frac{1}{2}}}$$

$$= r\left(1 - \frac{Kr^2}{6} + \frac{K^2 r^4}{120}\right) + \mathcal{O}(r^7), \tag{4.9}$$

which implies that the corresponding 'spherical' N is:

$$N_s = \frac{2K^2}{45}, \tag{4.10}$$

while D and E become:

$$D_s = -\frac{K}{6}, \tag{4.11a}$$

$$E_s = -\frac{K^2}{90}. \tag{4.11b}$$

Our approach is to assume the solution P can be factored into the part, P_0, together with a smooth multiplicative perturbation function, $T^+(t, r)$, where,

$$T^+(t, r) = 1 + T(t, r) \tag{4.12}$$

and where $T(t, r)$ can be expanded asymptotically into non-negative powers of t and r^2:

$$T(t, r) \approx \sum_{s,p \geq 0} T_{s,p} t^s r^p, \tag{4.13}$$

with $T_{0,0} = 0$, and $T_{s,p} = 0$ for p odd. The diffusion equation now becomes:

$$\frac{\partial P_0}{\partial t} T^+ + P_0 \frac{\partial T}{\partial t} = \left(\frac{\partial \ln(g^{\frac{1}{2}})}{\partial r}\right)\left(\frac{\partial P_0}{\partial r} T^+ + P_0 \frac{\partial T}{\partial r}\right)$$

$$+ \frac{\partial^2 P_0}{\partial r^2} T^+ + 2\frac{\partial P_0}{\partial r}\frac{\partial T}{\partial r} + P_0 \frac{\partial^2 T}{\partial r^2}. \tag{4.14}$$

Expand the metric term involving $\ln(g^{\frac{1}{2}})$ and divide by P_0 to obtain:

$$\frac{1}{P_0}\frac{\partial P_0}{\partial t} T^+ + \frac{\partial T}{\partial t} = \frac{1}{r}(1 + 2Dr^2 + 2Er^4)\left(\frac{1}{P_0}\frac{\partial P_0}{\partial r} T^+ + \frac{\partial T}{\partial r}\right)$$

$$+ \frac{1}{P_0}\frac{\partial^2 P_0}{\partial r^2} T^+ + \frac{2}{P_0}\frac{\partial P_0}{\partial r}\frac{\partial T}{\partial r} + \frac{\partial T}{\partial r^2}. \tag{4.15}$$

Separate the '1' term of the metric factor from the $(2Dr^2 + 2Er^4)$ term, cancel those terms involving T^+ that result from evaluating the derivatives of P_0 that multiply these T^+ factors, and expand the remaining T^+ factor as $(1 + T)$:

$$0 = -\frac{\partial T}{\partial t} + (Dr^2 + Er^4)\frac{2}{rP_0}\frac{\partial P_0}{\partial r}(1 + T)$$

$$+ \frac{1}{r}\frac{\partial T}{\partial r} + (Dr^2 + Er^4)\frac{2}{r}\frac{\partial T}{\partial r} + \frac{2}{P_0}\frac{\partial P_0}{\partial r}\frac{\partial T}{\partial r} + \frac{\partial^2 T}{\partial r^2}. \tag{4.16}$$

19

Make the substitutions for the derivatives of P_0, multiply through by $-t$, and group the terms to obtain the zero sum of the following three terms:

$$0 = \mathcal{T} + \mathcal{G} + \mathcal{Q}, \tag{4.17}$$

where:

$$\mathcal{T} = t\left(\frac{\partial T}{\partial t} - \frac{1}{r}\frac{\partial T}{\partial r} - \frac{\partial^2 T}{\partial r^2}\right) + r\frac{\partial T}{\partial r}, \tag{4.18a}$$

$$\mathcal{G} = (Dr^2 + Er^4), \tag{4.18b}$$

$$\mathcal{Q} = \mathcal{G}\left(T - \frac{2t}{r}\frac{\partial T}{\partial r}\right). \tag{4.18c}$$

If we introduce what we shall refer to as a 'slab index', $h = 2s + p$, for each term involving the powers $t^s r^p$, then we notice that each of the terms of $\mathcal{T}$ involves an operator whose effect on a contribution, $T_{s,p}$, is to leave its slab index unchanged (although in two of the four terms, the individual indices, s and p, of the result are different from those of the operand). We also see that the operator $\boldsymbol{U}$ that acts upon T to produce $\mathcal{T}$ is a linear one, and independent of the metric terms. We shall write the vector of the expansion coefficients of T at a given particular slab index h as $\boldsymbol{T}_{[h]}$, but it will be convenient to express the concatenation of the vectors $\boldsymbol{T}_{[h']}$ up to and including h as $\boldsymbol{T}^{(h)}$. Convenient notations for the corresponding vectors of the coefficients of the results, $\mathcal{T}$, at, and up to h, are $\hat{\boldsymbol{T}}_{[h]}$ and $\hat{\boldsymbol{T}}^{(h)}$, respectively. The restriction of the operator $\boldsymbol{U}$ to a particular h will also be denoted $\boldsymbol{U}_{[h]}$. Within a vector (or matrix) of given h, we shall list the components in order of decreasing time exponent, s, and increasing space index, p. For example,

$$\boldsymbol{T}_{[4]} \equiv [T_{2,0}, T_{1,2}, T_{0,4}]. \tag{4.19}$$

Each matrix, $\boldsymbol{U}_{[h]}$, for even $h \geq 2$ is upper triangular and nonsingular. For examples, since $T_{[2]} = T_{1,0}t + T_{0,2}r^2$, we find:

$$\mathcal{T}_{[2]} = (T_{1,0} - 4T_{0,2})t + 2T_{0,2}r^2, \tag{4.20}$$

and hence,

$$\begin{bmatrix} \hat{T}_{1,0} \\ \hat{T}_{0,2} \end{bmatrix} = \begin{bmatrix} 1, & -4 \\ 0, & 2 \end{bmatrix} \begin{bmatrix} T_{1,0} \\ T_{0,2} \end{bmatrix}. \tag{4.21}$$

The matrix in (4.21) is obviously $\boldsymbol{U}_{[2]}$. Similarly, we find that the matrix $\boldsymbol{U}_{[4]}$ is given by:

$$\boldsymbol{U}_{[4]} = \begin{bmatrix} 2, & -4, & 0 \\ 0, & 3, & -16 \\ 0, & 0, & 4 \end{bmatrix}. \tag{4.22}$$

The terms of $\mathcal{G}$ are purely geometrical, time invariant, and do not involve T. Expanded in powers,

$$\mathcal{G} = \sum_{p \geq 2} \hat{\boldsymbol{G}}_{0,p} r^p, \tag{4.23}$$

20

with p even, so the vector of coefficients, $\hat{\boldsymbol{G}}_{s,p}$, at a given slab index, h, are only nonzero for the parts that correspond to $s = 0$ (and are entirely absent when h is odd). The vectors, $\hat{\boldsymbol{G}}_{[2]}$ and $\hat{\boldsymbol{G}}_{[4]}$ are given according to our defined conventions respectively as:

$$\hat{\boldsymbol{G}}_{[2]} = [0, D], \tag{4.24a}$$

$$\hat{\boldsymbol{G}}_{[4]} = [0, 0, E]. \tag{4.24b}$$

The third term, (4.18c), of (4.17) is separately linear in T and in the metric forcing, $\mathcal{G}$. The presence of the $\mathcal{G}$ factor, whose smallest slab index is 2, ensures that the portion of the coefficients vector, $\hat{\boldsymbol{Q}}_{(h)}$, of $\mathcal{Q}$ up to and including slab index h can only involve $\boldsymbol{T}^{(h-2)}$, the components of the solution of slab index $h - 2$ or less. We are therefore able to accumulate a series of the solution coefficients of T by iteratively solving:

$$\boldsymbol{T}_{[h]} = -\boldsymbol{U}_{[h]}^{-1} \left(\hat{\boldsymbol{G}}_{[h]} + \hat{\boldsymbol{Q}}_{[h]} \right), \tag{4.25}$$

for $h = 2, 4, \ldots$, with $\hat{\boldsymbol{Q}}_{[2]} = \boldsymbol{0}$ in the first step of the procedure.

We proceed by noting that, at slab index $h = 2$, (4.25) gives:

$$\begin{bmatrix} T_{1,0} \\ T_{0,2} \end{bmatrix} = - \begin{bmatrix} 1, & 2 \\ 0, & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ D \end{bmatrix} \tag{4.26}$$

and hence,

$$\boldsymbol{T}_{[2]}^{T} = \left[-2D, \ -\frac{1}{2}D \right], \tag{4.27}$$

or

$$T^{(2)}(t, r) = -2Dt - \frac{1}{2}Dr^2. \tag{4.28}$$

In step two, we begin by evaluating

$$\mathcal{Q}^{(4)} = -\frac{1}{2}D^2 r^4, \tag{4.29}$$

or,

$$\hat{\boldsymbol{Q}}_{[4]} = \left[0, \ 0, \ -\frac{1}{2}D^2 \right]. \tag{4.30}$$

Since,

$$\hat{\boldsymbol{G}}_{[4]} = [0, \ 0, \ E], \tag{4.31}$$

we achieve the next level of approximation:

$$\begin{bmatrix} T_{2,0} \\ T_{1,2} \\ T_{0,4} \end{bmatrix} = - \begin{bmatrix} \frac{1}{2}, & \frac{2}{3}, & \frac{8}{3} \\ 0, & \frac{1}{3}, & \frac{4}{3} \\ 0, & 0, & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2}(2E - D^2) \end{bmatrix}, \tag{4.32}$$

21

and hence,

$$\boldsymbol{T}_{[4]}^T = \left[\frac{4}{3}(D^2 - 2E),\ \frac{2}{3}(D^2 - 2E),\ \frac{1}{8}(D^2 - 2E) \right]. \tag{4.33}$$

Up to terms with $h = 4$, the modulating function T becomes

$$T^{(4)}(t, r) = -2Dt - \frac{1}{2}Dr^2 + (D^2 - 2E)\left(\frac{4}{3}t^2 + \frac{2}{3}tr^2 + \frac{1}{8}r^4 \right). \tag{4.34}$$

If we are interested only in the evolution of the amplitude quotient, A, at the center, at the standard diffusion time, $t = \frac{1}{2}$, then we evaluate the expansion for $1 + T$ at this time with $r = 0$:

$$A^{(2)} = 1 - D + \frac{1}{3}(D^2 - 2E). \tag{4.35}$$

Thus for the case of constant Gaussian curvature,

$$
\begin{aligned}
A^{(2)} &= 1 + \frac{K}{6} + \frac{1}{3}\left(\frac{K^2}{36} + \frac{K^2}{45} \right) \\
&= 1 + \frac{K}{6} + \frac{K^2}{60}.
\end{aligned}
\tag{4.36}
$$

In the more general case, it is inconvenient to employ the N of (4.6) to quantify the principal non-sphericity; for applications, the (covariant) derivatives of the Gaussian curvature are the practical diagnostics. The next subsection therefore analyzes the differential geometry of the non-spherical, but axially symmetric surface in the vicinity of the axis and obtains the connecting relationships between variations of curvature and variations of the metric.

(*b*) *Curvature and metric variations near the axis of an axisymmetric surface*

In the embedding of an axially-symmetric 2-manifold into 3D Cartesians, let $Z(X)$ be the altitude of the graph of the embedded surface at a normal distance X from the axis.

The 'spherical image' maps a point on the manifold to the corresponding point on the unit sphere at which the corresponding tangent planes are parallel. The colatitude, ϕ, on the sphere is therefore:

$$\phi = \arctan\left(-\frac{dZ}{dX} \right), \tag{4.37}$$

and the inscribed solid angle (area of the spherical cap poleward of the image point) is

$$\sigma = 2\pi(1 - \cos\phi). \tag{4.38}$$

But

$$\frac{d\phi}{dX} = \frac{-\frac{d^2 Z}{dX^2}}{\left[1 + \left(\frac{dZ}{dX} \right)^2 \right]}, \tag{4.39}$$

so,

$$\frac{d\sigma}{dX} = -2\pi \frac{\sin\phi}{\left[1 + \left(\frac{dZ}{dX} \right)^2 \right]} \frac{d^2 Z}{dX^2}. \tag{4.40}$$

Let r be the integrated radial distance along a curve within the surface. Then, by Pythagoras,

$$\frac{dr}{dX} = \left[1 + \left(\frac{dZ}{dX}\right)^2\right]^{\frac{1}{2}}, \tag{4.41}$$

so the increment in the inscribed area of the manifold bounded by a circle at normal axial distance X is:

$$\frac{da}{dX} = 2\pi X \left[1 + \left(\frac{dZ}{dX}\right)^2\right]^{\frac{1}{2}}, \tag{4.42}$$

and since the definition of Gaussian curvature, κ, is the quantity, $d\sigma/da$, the Jacobian of the spherical image, we obtain, after substituting,

$$\sin \phi = \frac{\tan \phi}{(1 + \tan^2 \phi)^{\frac{1}{2}}}, \tag{4.43}$$

$$\kappa = \frac{d\sigma}{da} = \frac{\left(\frac{1}{X}\frac{dZ}{dX}\right)\left(\frac{d^2Z}{dX^2}\right)}{\left[1 + \left(\frac{dZ}{dX}\right)^2\right]^2}. \tag{4.44}$$

We want to be able to relate the variations of κ to the radial variations in the transverse metric factor,

$$H \equiv \frac{X^2}{r^2}. \tag{4.45}$$

We proceed by expanding Z and its derivatives, and hence dr/dX and its integral, r, and κ as Taylor expansions in X. By an inversion of the series for r, we obtain X, and hence H, expanded in r, which allows the relationship between the radial profiles of H and κ to be found. We shall only pursue this method out to the first non-spherical term in the expansions.

Let,

$$Z = Z_2 X^2 + Z_4 X^4 + \mathcal{O}(X^6), \tag{4.46a}$$

$$\frac{dZ}{dX} = 2Z_2 X + 4Z_4 X^3 + \mathcal{O}(X^5), \tag{4.46b}$$

$$\frac{d^2Z}{dX^2} = 2Z_2 + 12Z_4 X^2 + \mathcal{O}(X^4), \tag{4.46c}$$

Hence, by applications of the binomial theorem:

$$\left(\frac{dZ}{dX}\right)^2 = 4Z_2^2 X^2 + 16Z_2 Z_4 X^4 + \mathcal{O}(X^6), \tag{4.47a}$$

$$\left(\frac{dr}{dX}\right)^2 = 1 + 4Z_2^2 X^2 + 16Z_2 Z_4 X^4 + \mathcal{O}(X^6), \tag{4.47b}$$

$$\frac{dr}{dX} = 1 + 2Z_2^2 X^2 + (-2Z_2^4 + 8Z_2 Z_4)X^4 + \mathcal{O}(X^6). \tag{4.47c}$$

23

Obtain r by integrating:

$$r = X + \frac{2Z_2^2}{3}X^3 + \frac{(-2Z_2^4 + 8Z_2Z_4)}{5}X^5 + \mathcal{O}(X^7). \tag{4.48}$$

The Gaussian curvature profile is given by:

$$\begin{aligned} \kappa &= \frac{(2Z_2 + 4Z_4X^2)(2Z_2 + 12Z_4X^2)}{(1 + 4Z_2^2X^2)^2} + \mathcal{O}(X^4) \\ &= 4Z_2^2 + 32(-Z_2^4 + Z_2Z_4)X^2 + \mathcal{O}(X^4). \end{aligned} \tag{4.49}$$

But we need to express X as a function of r. This can be done in the general case, and up to any arbitrary degree of approximation, by iterative inversion of the series expansions. Expressing $X^{(p)}$ as the terms accurate up to degree r^p and sequentially substituting:

$$X^{(p+2)} = r - \frac{2Z_2^2}{3}(X^{(p)})^3 - \frac{(-2Z_2^4 + 8Z_2Z_4)}{5}(X^{(p)})^5 + \dots \tag{4.50}$$

up to the necessary terms on the right, we get the new expression of the left. Thus we obtain the following sequence of approximations:

$$\begin{aligned} X^{(1)} &= r, & \text{(4.51a)} \\ X^{(3)} &= r - \frac{2Z_2^2}{3}r^3, & \text{(4.51b)} \\ X^{(5)} &= r - \frac{2Z_2^2}{3}r^3 + \frac{(26Z_2^4 - 24Z_2Z_4)}{15}r^5, & \text{(4.51c)} \end{aligned}$$

and so on. Expressing κ as a Taylor series in r:

$$\begin{aligned} \kappa &= K + K_2r^2 + \mathcal{O}(r^4), & \text{(4.52a)} \\ &= 4Z_2^2 + 32(-Z_2^4 + Z_2Z_4)r^2 + \mathcal{O}(r^4). & \text{(4.52b)} \end{aligned}$$

Then squaring, and dividing the expression for X by r^2 gives us:

$$H = 1 - \frac{4Z_2^2}{3}r^2 + \frac{16}{45}(11Z_2^4 - 9Z_2Z_4)r^4 + \mathcal{O}(r^6), \tag{4.53}$$

which can therefore be written in terms of the curvature coefficients:

$$H = 1 - \frac{K}{3}r^2 + \left(\frac{2}{45}K^2 - \frac{K_2}{10}\right)r^4 + \mathcal{O}(r^6). \tag{4.54}$$

In terms of the original parameterization of metric variation, we can now express,

$$N = \frac{2}{45}K^2 - \frac{K_2}{10}, \tag{4.55}$$

so

$$E = \frac{-K^2}{90} - \frac{K_2}{10}. \tag{4.56}$$

Thus, in terms of K and K_2, we obtain the first two approximations to the amplitude quotient:

$$A^{(1)} = 1 + \frac{K}{6}, \tag{4.57a}$$

$$A^{(2)} = 1 + \frac{K}{6} + \left(\frac{K^2}{60} + \frac{K_2}{15}\right). \tag{4.57b}$$

(c) *Remarks*

The two terms of the approximation of the amplitude quotient for this axi-symmetric geometry confirm the coefficients for the spherically symmetric special case obtained by the previously discussed method. But the new second order term includes the second derivatives of Gaussian curvature and, significantly, shows that this contribution enters the formula *linearly*. The only linear second derivative diagnostics of a scalar that are invariant with respect to the arbitrary choice of rotation of the coordinate frame are constants times the Laplacian. The linearity of the second derivative in the second term of the approximation can be shown also to hold in the *non*-axi-symmetric case. Moreover, it can be shown that, in the more general geometry, the first derivatives are still absent from this second term in the expansion. It is therefore deduced that, in the unrestricted 2D geometry, we can express the second order expansion:

$$A^{(2)} = 1 + \frac{K}{6} + \frac{1}{60}(K^2 + \nabla^2 \kappa), \tag{4.58}$$

or equivalently,

$$A^{(2)} = 1 + \frac{\kappa}{6} + \frac{1}{60}(\kappa^2 + \nabla^2 \kappa), \tag{4.59}$$

which is consistent with (4.57b) in the special axi-symmetric case.

5. EXPERIMENTAL VALIDATION AND PRACTICAL REFINEMENTS

We have derived the simplest asymptotic approximations for the amplitude quotients in two and three dimensions by examining the solutions to the diffusion equation in highly idealized geometries. In this section we begin to investigate some practical amplitude normalization procedures, at least for the two-dimensional case for which we have the complete asymptotic approximation up to second order. It is important to remember that, although the polynomial asymptotic approximation for the amplitude quotient up to a given order is unique when it is expressed in the intrinsic curvature variables that are obtained from the parametrix expansion, at a fixed order of approximation, there are infinitely many approximating functions of these same curvature diagnostics that are exactly consistent with the polynomial expansion to this fixed order. It is also important in practice to ensure not only that this basic consistency is maintained, but that, for relatively large excursions of the curvature diagnostics where the asymptotic approximation is not even expected to be accurate, the approximating function we adopt does not 'blow-up' or produce obviously absurd magnitudes for the re-normalized filter amplitudes.

To illustrate the point, consider the first-order asymptotic expansion for the amplitude quotient which, in two dimensions, is just a first-degree polynomial of Gaussian curvature. For

small values of this curvature, this approximation to the quotient is close to unity and a satisfactory normalization of the filter is obtained by compounding the usual Gaussian amplitude formula by a further division by this polynomial. However, it is clear that for a sufficiently large (negative) Gaussian curvature, we will be thereby dividing by a *negative* amount, or even by zero, so the first-order approximation by itself leads to an unsatisfactory practical method. It is true that, at the second order of expansion, the part of the polynomial in (real) Gaussian curvature remains positive, but the problem is not avoided since the part of the polynomial in the *Laplacian* of Gaussian curvature enters the approximation at this order of accuracy at first degree, leading to the same dilemma as before.

The problem of the approximation to the amplitude quotient changing sign, and therefore being completely in error, is sometimes alleviated by proceeding to higher orders of approximation, but, as is well-known in the general theory of asymptotic series (e.g., Erdelyi 1955), there usually comes a point in such a series where the additional terms in the summation make the approximation worse, not better. In approximations to a single isolated variable it is customary to stop at this point and to accept the approximation and its associated error, imperfect as it is. However, the task we are faced with does not even allow us the option of stopping at one point of the asymptotic expansion in one geographical region, and at another more appropriate point in a region where the series can be carried further – such a strategy would lead to unacceptable discontinuities in the filter amplitude between the distinct regions. Instead, for a given order of truncation of the asymptotic series, we need to seek a function of the curvature diagnostics which is exactly consistent with the asymptotic approximation to this order, which is always positive (since we wish to divide by it during normalization) and which does not go wild even when the excursions of the curvature diagnostic on which it depends are large.

Maintaining positivity is not difficult; one approach is to express the approximate amplitude quotient as the exponential of another function, since the exponential is positive for all arguments. This other function must look like the logarithm of the amplitude quotient at small values of the curvature diagnostics, and, in the simplest cases, can be taken as the Taylor polynomial (in the same curvature diagnostic independent variables) to the appropriate degree.

We take as our control experiments the Gaussian normalization of the Euclidean space version of the diffusion problem, where the diffusivity is the aspect tensor and the coordinates are cartesian. This has been the conventional method of normalizing the covariance filter and will be referred to in tables as method 'A_0'. All the other experiments use the non-Euclidean geometry formulation in which the aspect tensor is interpretted as the metric and the diffusivity within this implied curved space is the identity tensor. The simplest form of this approach is to carry out the diffusion, and to use the trivial Gaussian normalization formula without any further correction, a method we refer to as 'B_0'. From an initial unit impulse, the Gaussian normalization of this method requires the division of the final diffused state by $(2\pi)^{N/2}$ for N dimensions. We shall refer to the formulas for further normalization by the basic polynomial taken directly from the first-order and second-order parametrix expansions as the methods 'B_1' and 'B_2'. If we denote the filter amplitudes resulting from an initial unit impulse following each of these normalization methods by the corresponding lower case letters, then the first- and second-order parametrix polynomial renormalizations, b_1 and b_2, relate to the amplitude

estimate, b_0, obtained from the above Gaussian amplitude normalization:

$$b_1 \;=\; b_0/(1 + \kappa/6), \tag{5.1a}$$

$$b_2 \;=\; b_0/(1 + \kappa/6 + \kappa^2/60 + \nabla^2\kappa/60). \tag{5.1b}$$

[These equations are the direct interpretation of the formulas (4.57a) and (4.59).] Denoting the first- and second-order methods that employ the exponential function, methods 'C_1' and 'C_2', the amplitudes they prescribe when applied as multiplicative corrections to the Gaussian amplitude are:

$$c_1 \;=\; b_0/\exp(\kappa/6), \tag{5.2a}$$

$$c_2 \;=\; b_0/\exp(\kappa/6 + \kappa^2/360 + \nabla^2\kappa/60). \tag{5.2b}$$

One potentially undesirable feature of the exponential function in this role of a normalizing divisor is that it can still produce quite large resulting amplitudes for large negative values of its argument. There are approximations to the exponential function that achieve the same advantages of positivity while avoiding the possibility of the normalization producing such exceptionally large amplitudes. In order to maintain the validity of the exponential form (5.2b) of the second-order asymptotic expansion we require that any such 'pseudo-exponential function' approximates the true exponential up to at least second-degree terms in its Taylor expansions about zero. If this pseudo-exponential, $\widetilde{\exp}$ is defined in terms of a correspondingly faithful 'pseudo-sinh function' that approximates the sinh function:

$$\widetilde{\exp}(x) = \widetilde{\sinh}(x) + \sqrt{1 + \widetilde{\sinh}^2(x)}, \tag{5.3}$$

then the symmetry, $\widetilde{\exp}(-x) = 1/\widetilde{\exp}(x)$ is preserved. The obvious simplest choice is then to let

$$\widetilde{\sinh}(x) = x \tag{5.4}$$

(a next simplest approximation, using the third-order Taylor expansion of sinh, is then available if higher-orders of the parametrix expansion are ever needed). Using the simplest option we define:

$$\widetilde{\exp}(x) = x + \sqrt{1 + x^2}. \tag{5.5}$$

We denote normalization methods that use this function instead of the true exponential, 'D', and define their first- and second-order forms:

$$d_1 \;=\; b_0/\widetilde{\exp}(\kappa/6) \tag{5.6a}$$

$$d_2 \;=\; b_0/\widetilde{\exp}(\kappa/6 + \kappa^2/360 + \nabla^2\kappa/60). \tag{5.6b}$$

We test these simple normalization methods, and some extensions of them, in the experiments we describe below.

(a) *Experiments representing an idealized 'vortex'*

In the first set of experiments we attempt to idealize the influence a symmetric vortex might be expected to have on the aspect tensor distribution, with a pronounced suppression of correlation in the radial direction compared to the tangential direction in the vicinity of the vortex.
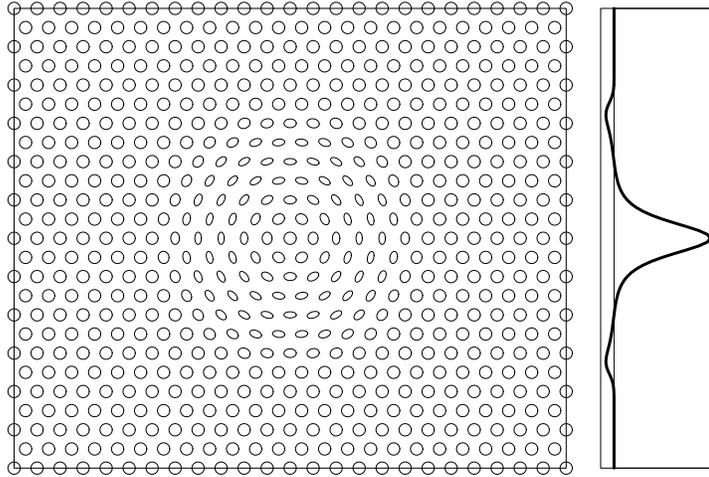
Figure 3. The main panel shows the distribution of the aspect tensor in for the idealized 'vortex' model characterized by the covariance scale parameter, $X = 4$ and the vortex intensity parameter, $Z = 40$. The side panel shows the normalized profile, along the vertical mid-line of the left panel, of the implied Gaussian curvature when the aspect tensor is interpreted as the Riemannian metric.

In order to minimize the purely numerical effects in the simulation of the diffusive processes, the domain is taken to be doubly periodic (thereby avoiding boundary ambiguities) with a resolution of 144×120 grid points in the two cartesian directions spanning each period. The 'vortex' is modeled by constructing a bell-shaped 'altitude' deflection in a third coordinate direction of the imagined two-dimensional surface that remains undeflected at the rectangular border of the repeating region, and then taking as the aspect tensor the orthogonal projection, back into the original plane, of the metric that measures the correct three-dimensional infinitesimal distances embedded within the deflected surface. Thus, the projected distances remain 'true' in the direction tangent to the vortex flow in this metric. The bell-shaped deflection profile is almost Gaussian, being modulated to ensure that it goes to zero smoothly for a finite radius corresponding to the smaller half-width of the repeating region.

The amplitude Z (in grid-space units) of the deflection in the imaginary third dimension offers a useful control on the deviation of the implied geometry from Euclidean, and can be liken to an 'intensity' of the vortex, while a uniform scaling (X) of the implied metric provides a way of controlling the effective size of the vortex and therefore the abruptness of *changes* in the implied Gaussian curvature (but at the same time affecting the magnitude of that curvature, of course).

The Gaussian curvature in this case is just half the Ricci scalar:

$$\kappa = \frac{1}{2}R = \frac{1}{2}g^{ij}R_{ij}, \tag{5.7}$$

where the Ricci tensor is obtained:

$$R_{ij} = \frac{1}{2}\left((\ln g)_{,k}\Gamma^k_{ij} - (\ln g)_{,ij}\right) + \Gamma^k_{ij,k} - \Gamma^k_{il}\Gamma^l_{jk}, \tag{5.8}$$

28

and where Γ^k_{ij} is the Christoffel symbol (of the 'second kind') defined in terms of the metric by:

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}\left(g_{li,j} + g_{lj,i} - g_{ij,l}\right).$$ (5.9)

In all the above expresions, the comma subscripts are used to denote partial differentiation by the indicated coordinate component. A more extensive derivation of the tensor analysis of curved manifolds, including the above expressions, can be found in Part II.

An example of the metric field is illustrated in Fig. 3 by the array of small ellipses that approximately show loci of constant effective distance from their centers as defined by the metric given. In this case the amplitude Z of the imagined surface deflection is $Z = 40$. The slight shrinking of these ellipses in the radial direction of the large central vortex can be seen clearly. The side panel shows the shape of the profile of Gaussian curvature, κ, along the transect that passes vertically through the center of the main panel. The positive curvature at the center is balanced by the shallow 'moat' of negative curvature around the vortex periphery.

EUCLIDEAN   NONEUCLIDEAN
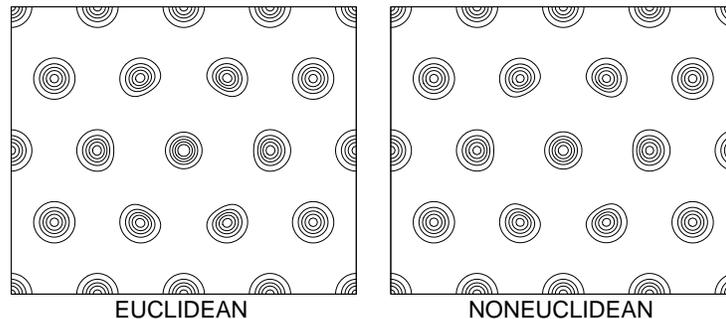
Figure 4. One of 36 arrays of diffused kernels for the 'vortex' aspect tensor distribution prescibed by the parameters $X = 4$ and $Z = 40$. The left panel shows the result of applying the diffusion with effective diffusivity tensor equal to the aspect tensor and with the Euclidean metric; the right panel shows the almost identical result obtained by modifying the diffusion problem to one in which the diffusivity is assumed uniform and isotropic, but the space is Riemannian with a metric equal to the aspect tensor.

The numerical differentiations required to obtain the metric, the Christoffel symbol components, and the curvature, are all centered, are of eighth-order accuracy, and use 64-bit precision. Coefficients of centered differencing and mid-point interpolations at various orders of accuracy can be found in the tables of Purser and Leslie (1988). The numerical truncation and round-off errors are thereby kept very small. The integration of the diffusion equation uses the classical fourth-order Runge-Kutta method (Johnson and Riess 1982) applied over 200 cycles of its four-step sequence. The statistics on errors of the various methods of amplitude estimation are accumulated over an array of 16 adequately separated impulses giving rise to the kernels such as those shown in Fig. 4. The vortex parameters generating this field are $X = 4$ and $Z = 40$. In order to be absolutely sure that a large enough statistical sampling of the domain is achieved, the diffusion integration is repeated 36 times with a different lateral translation of this array, so that the effective sampling array is also a well-spaced triangular one, but with a linear resolution six times that shown in the two panels of Fig. 4, making a total of 576 distinct sampling kernels. Although the overlaps among the 16 kernels of each realisation are small, their effect is further minimized by choosing a pattern of the signs of the initial impulses that

ensures that each kernel is surrounded by three 'positive' and three 'negative' signed neighbors, so that any residual overlap effects tend to cancel to first order at any given kernel center, where amplitudes are measured.

The scale and intensity parameters for Fig. 4 are $[X, Z] = [4, 40]$ as before. For both for the Euclidean and non-Euclidean outcomes of the diffusion problem. The shapes are visually identical from one panel to the other. For the Euclidean control, A_0, the use of the Gaussian formula for the estimation of amplitude gives errors whose root-mean-square (rms) over the entire set of 576 kernel locations is 0.01145, or a little exceeding one percent. Merely by transforming the problem to the non-Euclidean geometry and applying the same 'order-zero' approximation for the amplitude (experiment B_0), the rms error reduces to just 0.00272, which already suggests that the non-Euclidean framework is intrinsically advantageous. The first-order parametrix polynomial normalization B_1 improves this score to an rms error of 0.00026, which is further reduced to a mere 0.00010 by the application of the second order method, B_2. These statistics are listed in the top part of the first column of Table 1. With only the relatively weak effects of spatial inhomogeneity present in this experiment, all these rms errors are very small (as are the maximum and minimum error, which we do not list), but the results at least serve to reassure us that the asymptotic methods are behaving correctly.

EUCLIDEAN   NONEUCLIDEAN
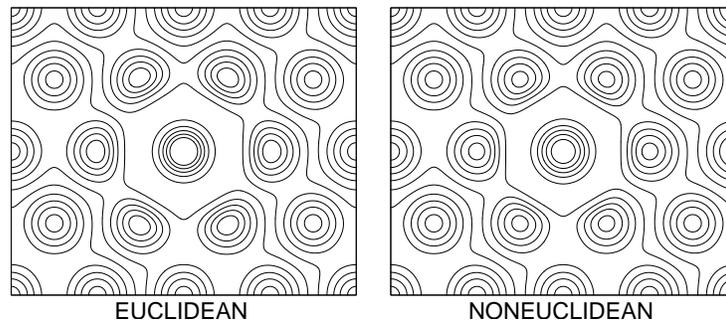
Figure 5.   Contours of one of the filtered arrays of test-kernels in the vortex geometry characterized by the parameters $[X, Z] = [8, 40]$.

Without changing the vortex intensity ($Z = 40$) but scaling the metric up by a factor of two ($X = 8$) to allow the diffusion process to spread the kernels to twice their linear size (or alternatively, shrinking the effective size of the vortex by a factor of two), we get the rms results displayed in the second column of the table, and an example of the pattern of diffused kernels shown in Fig. 5. The broader kernels are more distorted by the changes in the metric in this example. (Contours, plotted at only the odd integer multiples of 0.1, now reveal the parity pattern $[\pm 1]$ of the kernels.) While the tabulated rms statistics show a progressive improvement through the sequence of methods A_0, B_0 and B_1, we no longer see an improvement in accuracy in going to higher order of expansion, B_2. Also, for these cases and the last example, we see no significant change, and certainly not an improvement, when we use the C or D methods.

With a more intense vortex, $Z = 80$, even for the small-scale kernels ($X = 4$), the table shows no significant gain in going to a second-order expansion (B_2), although the first-order correction is still beneficial. At the larger kernel scales, $X = 8$, the second-order correction, B_2,

TABLE 1. Amplitude errors for quasi-diffusive filters normalized by various methods from idealized simulations of the aspect tensor distribution typical of the vicinity of an intense vortex. The standard method, using Euclidean space simulated diffusion and the Gaussian amplitude formula, is the control (A_0); also using the Gaussian amplitude formula, but with the diffusion problem reformulated in a space in which the aspect tensor is interpretted as a Riemannian metric, is denoted B_0. Normalization directly by the polynomials prescribed by the parametrix expansion methods at orders 1 and 2 are given by the results labeled B_1 and B_2. The other methods are all variations on the use of the parametrix expansion, as described in the text. Two parameters: kernel scale, X; and 'vortex intensity', Z; quantify the aspect tensor distribution. The tabulated results are RMS errors (compared with a true amplitude of unity) in the filter kernels normalized by the different methods and for a total of 576 essentially independent kernel locations evenly distributed over the entire domain. For the non-Euclidean interpretations of the geometry with aspect tensor serving as effective metric, the periodic domain of these experiments has an effective non-dimensional area, which is also tabulated. Also given are the minimum and maximum Gaussian curvature and minimum and maximum Laplacian of this curvature.

$[X, Z]$:	[4, 40]	[8, 40]	[4, 80]	[8, 80]	[4, 160]	[8, 160]
A_0	0.01145	0.03735	0.02446	0.06843	0.08015	0.12854
B_0	0.00272	0.00982	0.00785	0.01818	0.06136	0.03267
B_1	0.00026	0.00244	0.00559	0.01002	0.06074	0.02588
B_2	0.00010	0.00286	0.00578	0.40366	0.13822	0.04713
C_1	0.00030	0.00295	0.00576	0.01442	0.06151	0.03815
C_2	0.00011	0.00340	0.00602	0.16565	0.10526	$***$
D_1	0.00030	0.00292	0.00575	0.01335	0.06130	0.03127
D_2	0.00011	0.00340	0.00602	0.12069	0.09689	2.81527
E_1	0.00029	0.00226	0.00552	0.00366	0.06012	0.01429
E_2	0.00012	0.00108	0.00528	0.00283	0.06009	0.01413
F_1	0.00027	0.00168	0.00538	0.00610	0.06029	0.02589
F_2	0.00012	0.00127	0.00533	0.00544	0.06027	0.02571
G_2	0.00011	0.00184	0.00546	0.00322	0.06012	0.01378
H_2	0.00012	0.00085	0.00526	0.00268	0.06009	0.01487
area:	1225	306	1480	370	2047	512
$\kappa_{\min}$:	-0.021	-0.082	-0.030	-0.122	-0.040	-0.161
$\kappa_{\max}$:	0.244	0.975	0.975	3.90	3.76	15.0
$\nabla^2\kappa_{\min}$:	-0.515	-8.24	-7.690	-123.	-92.5	-1480.
$\nabla^2\kappa_{\max}$:	0.027	0.434	0.412	6.59	8.62	138.

is positively damaging and, although the damage is somewhat mitigated by the exponential and pseudo-exponential normalizations C_2 and D_2, the recovery is not sufficient to make even these methods improve even the zeroth-order non-Euclidean methods.

If we double the intensity parameter again, $Z = 160$, the metric distribution and associated curvature cross-section are as depicted in Fig. 6 and we can be sure that the diffusion kernels, even at the smaller scale, $X = 4$ will become severely distorted away from the Gaussian ideal. Fig. 7 confirms this expectation and shows the first visible difference between the Euclidean and non-Euclidean results (in the subtly different shapes of the central kernel, although both central
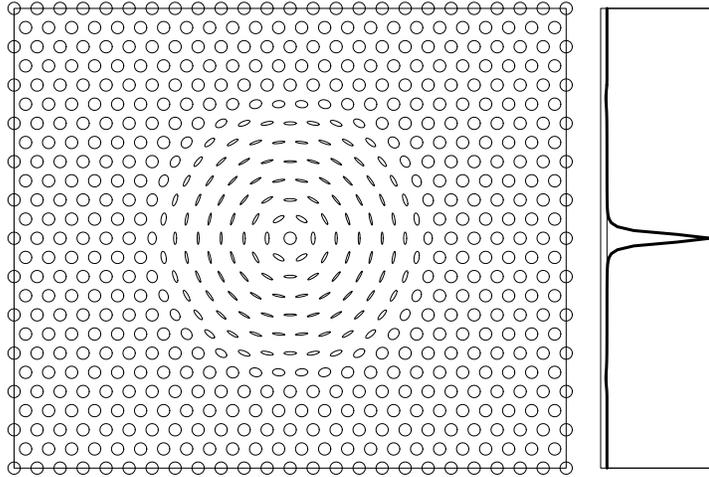
Figure 6. Metric distribution for the 'vortex' example with scale parameter $X = 4$ and intensity parameter $Z = 160$. As before, the associated curvature profile through the central vertical midline is shown (normalized) in the right side-panel.

EUCLIDEAN   NONEUCLIDEAN
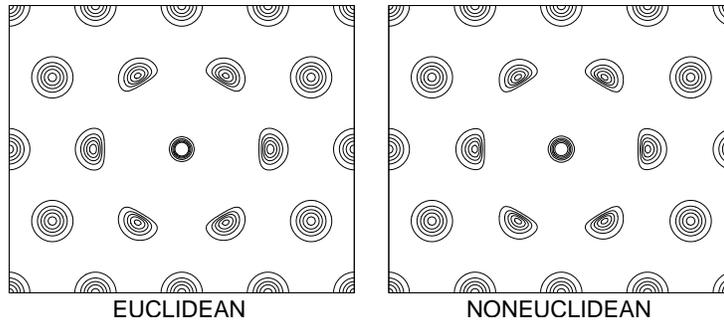
Figure 7. One of the 36 arrays of smoothed kernel used to test the normalization methods in the case of the 'vortex' model with same scale and intensity parameter, $[X, Z] = [4, 160]$, as in Fig. 6. The kernels within the vortex become obviously highly non-Gaussian now that the change in aspect tensor occurs over distances even smaller than the instrinsic size of the kernels themselves.

kernels become very plateau-like in this aspect tensor distribution in which the radial component of diffusion is strongly inhibited). The tabulated results continue to show the advantage of the non-Euclidean method and the first-order parametrix correction, but the basic second-order 'correction' in method B_2 cancels these gains. For the larger kernels obtained with $X = 8$, the errors can become very large. The maximum errors exceed 100 per cent for some kernels normalized by in the Euclidean method A_0 and in the second-order parametrix method B_2, but are not as bad for the other B methods. The rms errors for the B_2 method are actually much smaller in this geometry than they were in the case with the same $Z = 160$ but with $X = 4$; a result which we attribute to the divisor for the normalization factor for a few of the kernels in the earlier B_2 experiment being, by chance, close enough to zero to give absurdly large and erroneous normalized amplitudes while, for $X = 8$, this unlucky circumstance happens not to occur. However, for $[X, Z] = [8, 160]$ we find that the C_2 method, far from alleviating difficulties, suffers a debilitating failure which can be traced to the large negative value of the Laplacian

of the curvature. Recall that this term appears as part of the argument of the exponential function, causing the amplitude divisor to become exceedingly small in this case. This defect is at least partially cured by the use of the pseudo-exponential function, although the results, D_2, are then only marginally better in the rms sense than those of the method B_2.

While these preliminary experiments have confirmed the benefits of the asymptotic approach for relatively gentle variations of the aspect tensor, they already show the dangers inherent in trying to continue the asymptotic method too far when the absolute magnitudes of the curvature diagnostics begin to get large. They also show that the use of the pseudo-exponential (quadratic polynomial) function to guarantee positivity of the normalizing divisor is far better than using the exponential function itself. But we see that, if we are to enjoy the benefits of the second-order expansion, additional safeguards against deteriation of results at large curvature diagnostics must be included.

(b) *Incorporating saturation functions of diagnostics*

Since it is the larger absolute values of the curvature, κ, or of its Laplacian, $\nabla^2\kappa$, that result in the deteriation of the asymptotic normalization, we can avoid the more egregious errors by replacing these diagnostics by a suitable function that approximates them to first and second order at small values, but saturates to appropriate finite limits, or that even return towards zero after finite values of the argument. The first class of 'saturation function' would be exemplified by something like the hyperbolic tangent function, whose asymptotic values are ± 1 but which can be scaled to make these asymptotes any finite magnitude whatsoever. Various other functions, such as the arc-tangent, and the symmetrized 'error function', share the same qualitative form, but perhaps the easiest to work with numerically is the function, with a scaling, or 'saturation parameter', s, which we define:

$$\mathrm{SAT}_1(s, x) = \frac{x}{[1 + (x/s)^2]^{1/2}} \qquad (5.10)$$

Only the minor change of an alternative exponent of the denominator in this prescription produces a second kind of saturation function which, having saturated at a finite argument, x, thereafter drops back towards zero at very large $|x|$:

$$\mathrm{SAT}_2(s, x) = \frac{x}{[1 + (x/s)^2]} \qquad (5.11)$$

Both of these functions are illustrated for $s = 1$ and $s = 2$ in Fig. 8 for x in the range $[-5, 5]$.

The methods 'E' employ the first kind of saturation function of κ and of $\nabla^2\kappa$ in combination with the pseudo-exponential function we used in experiments D. As tabulated in table 1 the rms errors for E_1 and E_2 come from the amplitude estimates e_1 and e_2 defined:

$$e_1 = b_0/\widetilde{\exp}[\mathrm{SAT}_1(s_\kappa, \kappa)/6] \qquad (5.12a)$$

$$e_2 = b_0/\widetilde{\exp}[\mathrm{SAT}_1(s_\kappa, \kappa)/6 + (\mathrm{SAT}_1(s_\kappa, \kappa))^2/360 + \mathrm{SAT}_1(s_L, \nabla^2\kappa)/60], \qquad (5.12b)$$

where both saturation parameters, s_κ and s_L are arbitrarily set to 2. Experiments 'F' are defined similarly except that SAT_2 replaces SAT_1 in the definitions above. The saturation parameters in the tabulated results for experiments F are also both arbitrarily set to 2 here. The
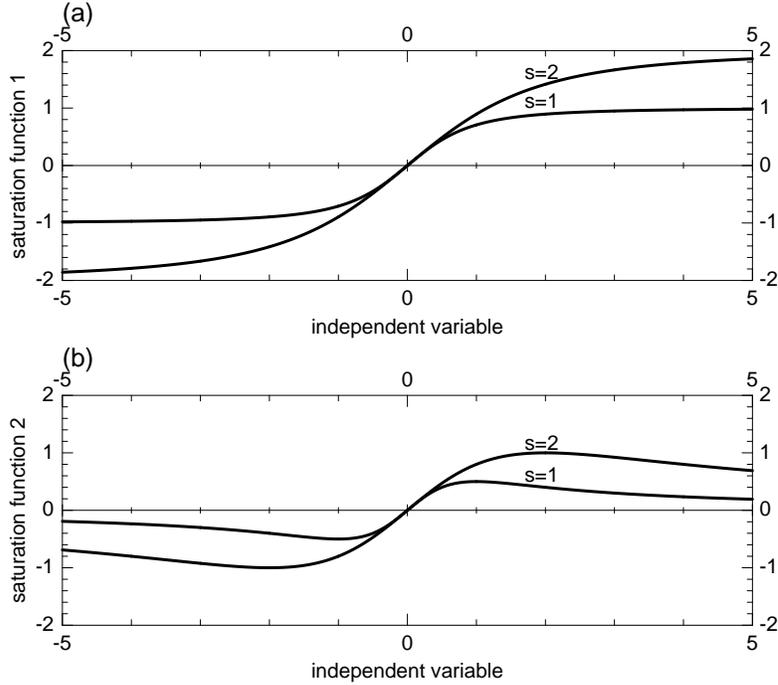
Figure 8. The two saturation functions, SAT_1 and SAT_2 for values $s = 1$ and $s = 2$ of their scaling parameters (the 'saturation parameters'). Other generalizations are clearly possible by choosing the denominator's exponent to be other than the $1/2$ or 1 used to produce the curves of panels (a) and (b).

rms scores do not show any significant degradation of any of the scores and, for the larger values of Z, there are some encouraging and quite dramatic improvements. Although the results with the E methods seem better than those of the F method, it is premature to judge between them when neither set of schemes have had their saturation parameters tuned. On the assumption that the diagnostic, $\nabla^2 \kappa$, is more likely to be worthless for the more abruptly adapting aspect tensors fields, we also introduce a hybrid second-order parametrix-based scheme, 'G_2', in which the influence $\nabla^2 \kappa$ is more thoroughly suppressed through its conditioning via the SAT_2 function, while the less repressive SAT_1 function continues to limit the influence of any large κ diagnostic:

$$g_2 = b_0 / \widetilde{\exp}[\mathrm{SAT}_1(s_\kappa, \kappa)/6 + (\mathrm{SAT}_1(s_\kappa, \kappa))^2/360 + \mathrm{SAT}_2(s_L, \nabla^2 \kappa)/60]. \qquad (5.13)$$

With the same saturation parameter values $(= 2)$, this hybrid scheme shows no particular advantage according to the rms results listed for it in table 1 but, again, without tuning the parameters, it is inappropriate to prejudice the final decision.

The Laplacian of the Gaussian curvature, which forms one of the diagnostics of the parameterix expansion at second order, may be regarded as the sum of the eigenvalues of the Hessian tensor of the Gaussian curvature. If we take the additional step of evaluating this Hessian and its eigenvalues, then it is reasonable to consider applying a saturation function to these separate contributions before including the result in the pseudo-exponential correction function.

The covariant definition of this Hessian is given by:

$$\mathcal{H}_{ij}(\kappa) = \left(\frac{\partial^2 \kappa}{\partial x^i \partial x^j} - \Gamma^k_{ij} \frac{\partial \kappa}{\partial x^k} \right). \tag{5.14}$$

If we write $\eta_{(1)}(\kappa)$ and $\eta_{(2)}(\kappa)$ as the two eigenvalues of the Hessian of κ, and $\boldsymbol{V}_{(1)}$, $\boldsymbol{V}_{(2)}$ the corresponding pair of contravariant eigenvectors, then

$$g^{ij} \mathcal{H}_{jk} \boldsymbol{V}^k_{(\alpha)} = \boldsymbol{V}^i_{(\alpha)} \eta_{(\alpha)}, \quad \alpha = 1, 2. \tag{5.15}$$

A convenient numerical approach to solving this eigen-problem is to use an appropriate square-root of g^{ij} to transform the non-standard eigen-problem into a standard one with symmetric matrix. For example, by a Cholesky decomposition of the metric:

$$\sum_{\nu=1}^{2} L^i_\nu L^j_\nu = g^{ij} \tag{5.16}$$

and a new pair of 'local Cartesian' eigenvectors, $\boldsymbol{U}_{(\alpha)}$, with

$$\boldsymbol{V}_{(\alpha)} = \boldsymbol{L} \boldsymbol{U}_{(\alpha)}, \quad \alpha = 1, 2 \tag{5.17}$$

enables the eigenvalues to be found from the standard symmetric eigen-problem:

$$\boldsymbol{L}^T \mathcal{H} \boldsymbol{L} \boldsymbol{U}_{(\alpha)} = \boldsymbol{U}_{(\alpha)} \eta_{(\alpha)}. \tag{5.18}$$

The experiments in which the amplitude estimates are like e_2, except with the final term replaced with the sum of the pair of terms, $\mathrm{SAT}_1(s_H, \eta_{(\alpha)})/60$, is called '$H_2$', where the new saturation parameter, s_H, is taken to be smaller than the corresponding s_L of experiment E_2 to allow for the expectation that the eigenvalue components are typically smaller, on average, than the Laplacian to which they add up. In table 1, the value of $s_H = \frac{2}{3} s_L = 4/3$ was arbitrarily taken. The method of splitting the components of the Laplacian in this way certainly does no serious harm compared to experiments E_2 and in some cases appears to further improve the estimates, but, again, the special symmetries of this series of tests make it unwise to draw premature conclusions.

(*c*) *Experiments representing smooth random distributions of the aspect tensor*

The idealized vortex model hardly represents the typical variety of aspect tensor distributions likely to be encountered in practice. We therefore seek greater variety through the construction of smooth but *random* distributions of the aspect tensor, still on the same periodic domain and with the same triangular-array distributions of test kernels repeatedly displaced. The aspect tensor must be positive-definite, which can be ensured by making it the exponential of a spatially smoothed random symmetric matrix (this devise works in any number of dimensions). The smooth matrices from which the metric tensors are obtained by exponentiation are themselves obtained by filtering Gaussian white-noise matrices with a low-pass spatial filter, which is easily done in the spectral domain via the application in both cartesian directions of the fast Fourier transform. The spectral shape of this smoothing filter is Gaussian, and it has

a certain characteristic scale parameter which we shall call S, with small values corresponding to small spatial scales of coherence. The variability of the random matrices is a parameter we call V, with larger values of V corresponding to generally larger excursions of the aspect tensor (size and anisotropy) away from some standard size and isotropic shape. (Very roughly, the reciprocal of the new S parameter has an effect analogous to the X parameter of the 'vortex' model, while the new variance parameter, V, has an effect analogous to the intensity parameter, Z, of the vortex, as far as the distortions of kernels is concerned.)

Examples of such randomly constructed geometries are illustrated, for relatively smooth and low-intensity variations, $[S, V] = [0.32, 0.3]$, in Fig. 9 and, at the higher variance or intensity, $[S, V] = [0.32, 0.6]$, in Fig. 10. In both cases the same random number sequence was employed, identifed by a 'seed' (equal to '1' in the cases shown), but the excursions from uniform isotropy of the aspect tensor of the second of these figures are visibly larger, even though most of the patterns are qualitatively very similar. Using the same random number sequence, but a smaller scale parameter now of only $S = 0.08$, we obtain the very choppy distributions of Fig. 11 for $V = 0.3$ and of Fig. 12 for $V = 0.6$. To give an idea of the effect of these geometries on the non-Euclidean representation of the diffused kernels, Figs. 13 – 16 show the shapes of the kernels resulting from one of the 36 arrays of impulses used in the amplitude tests. The difficulties of normalization will clearly occur in the more choppy geometries exemplied by Fig. 15 and especially Fig. 16, where the distortions to the kernels are clearly occurring at a scale significantly smaller than the kernels themselves. Therefore, our efforts to tune the saturation parameters of the functions SAT_1 and SAT_2 will focus mainly on the realizations of geometries similar to these.

Our tests will also involve the intermediate scales, $S = 0.16$ (paired with the same two variance parameters, V). Also, we shall repeat each trial with alternative random number sequences, identified by their 'seeds', '2' and '3', in order to avoid distortions of our assessments that might come from accidentally atypical results of a single set of realizations.
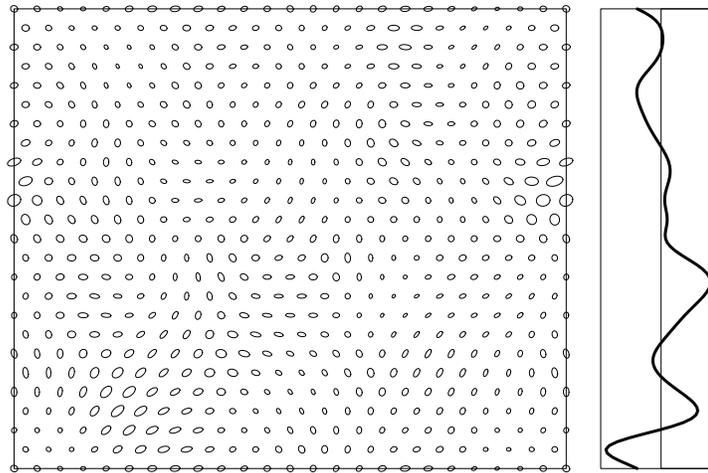
Figure 9.   Metric (or aspect tensor) distribution, together with the implied Gaussian curvature profile of a central transect, for a random non-Euclidean geometry constructed with the coherence scale parameter, $S = 0.32$, and the metric's variance parameter, $V = 0.3$.

36
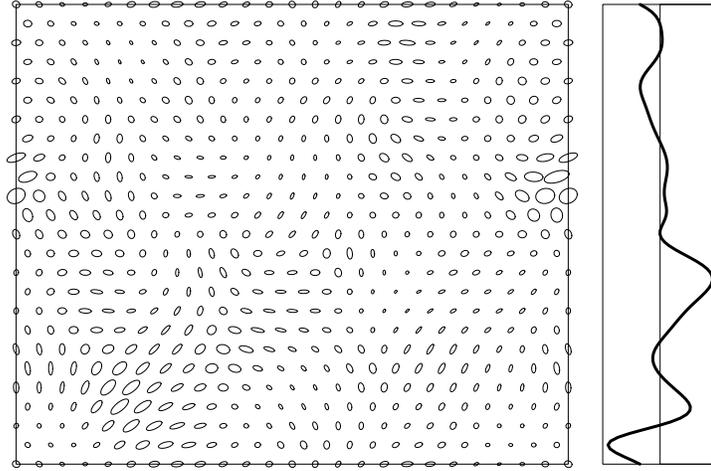
Figure 10. As in Fig. 9, and based upon the same initial random number 'white-noise' field, but filtered to produce a geometry with the characteristic parameters $S = 0.32$ and $V = 0.6$. The magnitudes of the vagaries of the aspect tensors away from isotropic uniformity are discernably larger than in Fig. 9, making the normalization problem more difficult.

Since both saturation functions reduce to x in the limit as their respective parameter, $s \to \infty$, and we can set these saturation parameters at will, it is hardly necessary to continue to dignify methods D_1 and D_2 with separate entries in any tables. Also, we have seen that methods B_1, B_2, C_1 and C_2 all exhibit unsatisfactory behavior for even the uncomplicated 'vortex' geometries, so we also winnow out these unreliable methods from further consideration. But we retain the Euclidean method, A_0, and the zeroth-order non-Euclidean method, B_0, as benchmarks by which we assess the improvements obtained by the more sophisticated methods, E_1, E_2, F_1, F_2, G_2 and H_2. Also, for these latter methods of guaranteed first- and second-order consistency with the parametrix expansion, we wish to find which choice of saturation parameter s_κ in the case of E_1 and F_1, and which parameter pair, $[s_\kappa, s_L]$, in the case of E_2, F_2 and G_2, yield the best expected rms statistics for random geometries characterized by different S and V.

The random geometries are first created. The diffusion integrations of all the kernel arrays are then performed, and the true amplitudes of each is gathered into a tabulation that is stored. Then, for each random geometry, the different normalization methods are tested, with the saturation parameters s_κ being permitted to take all half-integer values up to a suitable maximum, and parameters s_L being permitted (independently) to take all integer values up to another maximum, both maxima being chosen large enough for the data to contain the best parameters (smallest rms error) for all random geometries synthesized. From this discrete selection of the saturation parameters, the tabulated rms values in the upper sections of tables 2–4 are the best (smallest) possible, and chosen independently for each tabulated rms. The three tables are equivalent except that each corresponds to a different choice, 1, 2 and 3, respectively, of the 'seed' employed to initiate the random number white-noise sequence from which each random geometry is synthesized. Nevertheless, there is an excellent degree of consistency in the results revealed by the three tables.

First, we immediately see that the saturation function SAT_1 used for the methods E is more effective than SAT_2 used in experiments F. The hybrid saturation method G_2 is also not
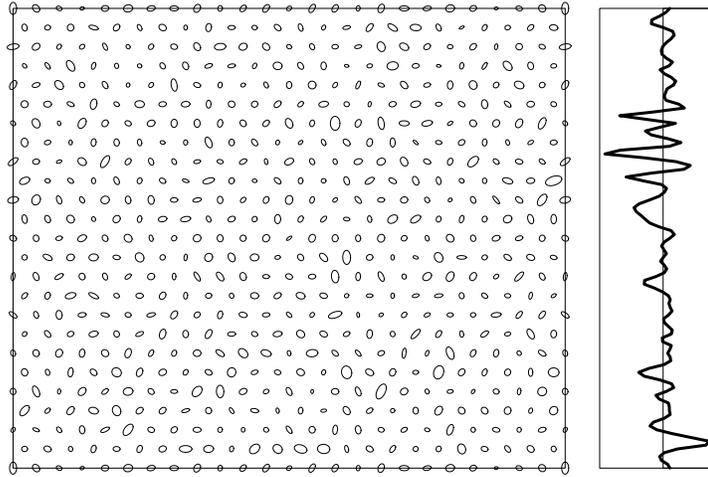
Figure 11.   As in Fig. 9 but with the parameters $S = 0.08$ and $V = 0.3$, showing a much 'choppier' distribution of the metric than in that otherwise comparable figure.
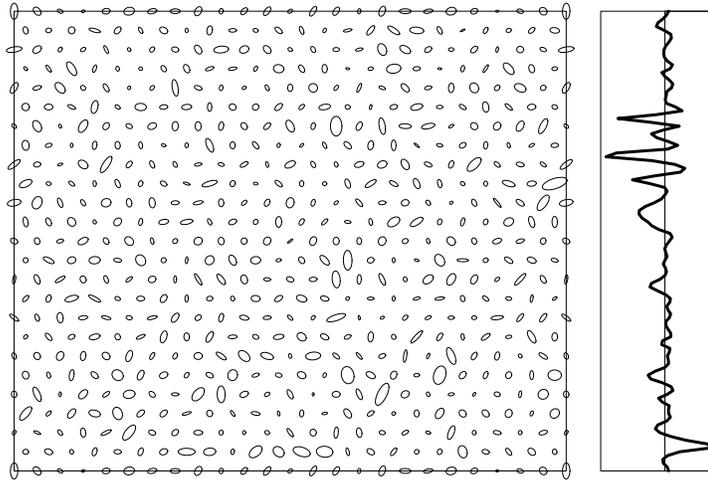
Figure 12.   As in Fig. 9 but with the parameters $S = 0.08$ and $V = 0.6$. Not only are the magnitudes of the excursions of the aspect tensors away from isotropy relatively large, but they now occur over a much smaller characteristic distance than they do in Figs. 9 and 10.

as good as E_2. Detailed inspection of the full array of rms errors at all the different values of the saturation parameters tested (which were too many to be tabulated here) showed no particularly sharp minimum to the rms errors for any of the choices of geometry parameters $[S, V]$, but yet a very definite and consistent grouping of the 'best' values to identifiable regions of the saturation-parameter space for each of these geometries.

Given the clear indications of the superiority, at both first- and second-order, of the E methods and the second order H_2 method, compared to the F and G methods, and given that the critical errors occur with the most difficult geometries when S is small and V is large, we have looked at the detailed results from the geometries associated with the right-most columns of these tables to identify the best overall choice of the saturation parameters. For the first-order

NONEUCLIDEAN
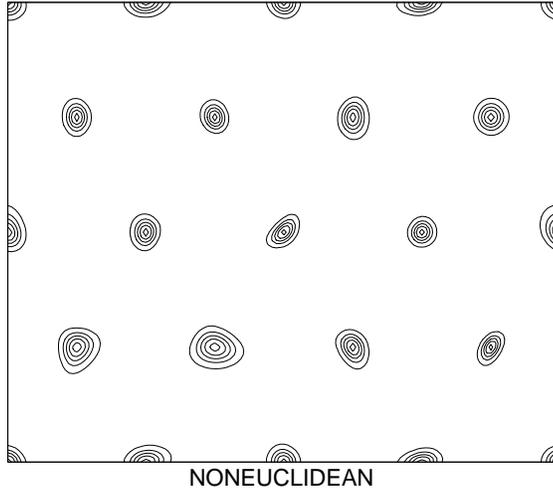
Figure 13. One of the 36 test arrays of smoothed kernel when diffusion is performed in the geometry with the metric distribution shown in Fig. 9 and with the parameters $S = 0.32$ and $V = 0.3$.

NONEUCLIDEAN
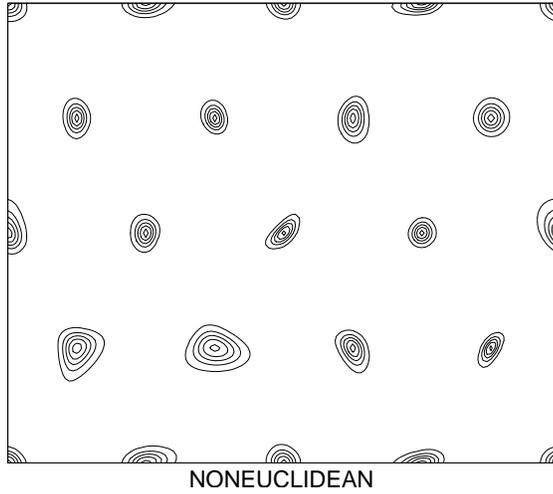
Figure 14. As in Fig. 13 but with the parameters $S = .32$ and $V = 0.6$ that correspond to the metric field shown in Fig. 10.

scheme with its single saturation parameter, s_κ, all three geometrical realizations for the case $[S, V] = [0.08, 0.6]$ suggested that this parameter should be chosen to be about $s_\kappa = 1$. When the second-order terms are included, the best value of this parameter appeared to be slightly larger; the parameter pair, $[s_\kappa, s_L] = [1.5, 3.0]$, was found to be never far from the 'best' for all the geometries corresponding to $S = 0.08$ and $X = 0.16$ and also all geometries corresponding to $V = 0.3$ or $V = 0.6$. It therefore remains for us to check that fixing the saturation parameters to these approximate optima does not unduly degrade the already very excellent rms scores for the remaining categories of random geometries we have synthesized. The entries for these 'optimal' E methods, which we denote E_1^* and E_2^* at first and second order respectively, are listed in the three tables, together with the corresponding scheme H_2, whose parameter pair

NONEUCLIDEAN
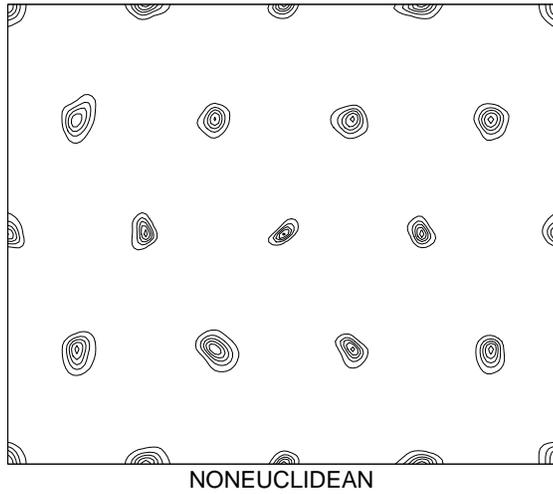
Figure 15. As in Fig. 13 but with the parameters $S = .08$ and $V = 0.3$ that correspond to the metric field shown in Fig. 11.

NONEUCLIDEAN
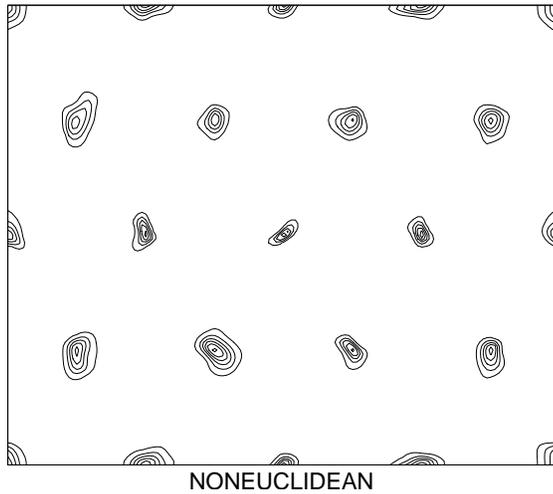
Figure 16. As in Fig. 13 but with the parameters $S = .08$ and $V = 0.6$ that correspond to the metric field shown in Fig. 12.

was fixed at $[s_\kappa, s_H] = [1.5, 2.0]$. It is reassuring to see that essentially no serious damage is done by adopting these choices. There is no clear-cut suggestion that either one of E_2 or H_2 is consistently the better method in these two-dimensional tests; both appear to perform well and exhibit the desired robustness.

(d) *Remarks*

Clearly, there are many other refinements we could consider to attempt to further improve the implementation of the asymptotic methods that we have developed based on the first- and second-order parametrix expansions derived in earlier sections of this note. Also, having overcome the initial difficulties of achieving consistently positive improvements with the second-

TABLE 2. TABLE OF RMS ERRORS OF THE NORMALIZED AMPLITUDES OF DIFFUSION EQUA-
TION FILTERS IN SMOOTH RANDOM GEOMETRIES WHOSE STATISTICAL CHARACTERISTICS ARE
GOVERNED BY A COHERENCE SCALE PARAMETER, S AND A VARIANCE PARAMETER, V. THE
METHODS E, F G AND H ARE DEFINED AS IN TABLE 1 EXCEPT THAT THE SATURATION
PARAMETERS, s_κ (FOR THE E_1 AND F_1 SCHEMES) AND s_L (FOR THE E_2, F_2 AND G_2 SCHEMES)
AND s_H (FOR SCHEME H_2), ARE NOW APPROXIMATELY OPTIMIZED FOR EACH PARTICULAR
DATASET. THE PRESENT TABLE IS FOR JUST ONE SET OF THE THREE RANDOM REALIZATIONS OF
THE GEOMETRIES WITH THE PRESCRIBED S AND V PARAMETERS, THE RANDOM REPEATABLE
NUMBER SEQUENCE BEING INITIATIALIZED BY THE SAME SEED, '1' IN THESE EXAMPLES. THE
FOLLOWING TWO TABLES PROVIDE THE EQUIVALENT DATA FOR THE REALIZATIONS STARTED
FROM SEEDS '2' AND '3', AND SHOW THAT THE QUALITATIVE RESULTS ARE CONSISTENT.
AMONG THE REMAINING CHOICES OF ASYMPTOTICALLY CORRECTED SCHEMES, IT IS THE E
SCHEMES WHICH ARE CLEARLY SUPERIOR. THE CRITICAL TESTS, WHERE THE OPTIMIZED
ERRORS ARE LARGEST, ARE THOSE AT THE SHORTEST COHERENCE SCALES S AND THE HIGHEST
METRIC VARIANCE, V, CORRESPONDING THE RIGHT-MOST COLUMN. THE 'BEST' PARAMETERS
FOR E_1 AND E_2 FROM THIS COLUMN AND THE CORRESPONDING COLUMNS FOR THE OTHER
TWO REALIZATIONS, ARE ROUNDED TO THE NEAREST HALF-INTEGER FOR s_κ, AND WHOLE
INTEGER FOR s_L, AND THEN HELD CONSTANT IN SCHEMES WE CALL E_1^* (FOR WHICH THE
BEST PARAMETER IS TAKEN TO BE $s_\kappa = 1$), E_2^* (FOR WHICH THE BEST PAIR JOINTLY ARE
$[s_\kappa, s_L] = [1.5, 3.0]$) AND H_2^* (FOR WHICH THE BEST PAIR OF PARAMETERS ARE TAKEN TO BE
$[s_\kappa, s_H] = [1.5, 2.0]$). THEY ARE THEN RE-TESTED FOR ALL THE SAME REALIZATIONS, WITH THE
RMS RESULTS THAT ARE SET OUT IN THE INDICATED ROWS OF THESE TABLES. THE EFFECTIVE
AREA OF THE DOMAIN UNDER THE GIVEN RANDOMIZED METRIC, AND THE LIMITS OF THE
CURVATURE AND ITS LAPLACIAN, ARE GIVEN IN THE BOTTOM SECTION OF THE TABLE.

$[S, V]$:	$[0.32, 0.3]$	$[0.16, 0.3]$	$[0.08, 0.3]$	$[0.32, 0.6]$	$[0.16, 0.6]$	$[0.08, 0.6]$
A_0	0.01444	0.04088	0.13788	0.03289	0.07667	0.23847
B_0	0.00915	0.02418	0.06197	0.02095	0.04306	0.09319
E_1	0.00397	0.00747	0.03296	0.01311	0.02102	0.06229
E_2	0.00394	0.00581	0.02382	0.01295	0.01760	0.05584
F_1	0.00396	0.00756	0.03556	0.01310	0.02188	0.06950
F_2	0.00394	0.00589	0.02900	0.01295	0.01847	0.06402
G_2	0.00394	0.00593	0.02745	0.01295	0.01812	0.05834
H_2	0.00394	0.00565	0.02564	0.01294	0.01721	0.05903
E_1^*	0.00397	0.00747	0.03296	0.01311	0.02102	0.06229
E_2^*	0.00394	0.00614	0.02398	0.01314	0.01794	0.05618
H_2^*	0.00394	0.00594	0.02610	0.01309	0.01803	0.05954
area:	4900	4874	4855	5394	5423	5421
$\kappa_{\min}$:	-0.470	-1.643	-9.392	-1.279	-4.829	-26.536
$\kappa_{\max}$:	0.140	0.843	5.337	0.318	2.743	15.878
$\nabla^2 \kappa_{\min}$:	-0.230	-15.243	-191.719	-1.379	-125.270	-1447.082
$\nabla^2 \kappa_{\max}$:	0.662	16.917	304.452	4.110	177.380	2938.975

order terms, we might be emboldened to proceed with the much more difficult algebraic task of
securing the third- and higher-order terms of the general expansion. However, the second-order
methods would appear to be quite adequate for all the foreseeable practical applications that
relate to the amplitude control of covariances for operational data assimilation, so we call a
halt at this point.

TABLE 3. Results equivalent to those of table 2 but for realizations of the geometry derived from the random number sequence initialized by a value of the 'seed' equal to 2

$[S, V]$:	$[0.32, 0.3]$	$[0.16, 0.3]$	$[0.08, 0.3]$	$[0.32, 0.6]$	$[0.16, 0.6]$	$[0.08, 0.6]$
A_0	0.01502	0.04120	0.12291	0.03515	0.07897	0.20992
B_0	0.00815	0.02294	0.06091	0.01772	0.04037	0.09204
E_1	0.00410	0.00591	0.03249	0.01082	0.01700	0.05972
E_2	0.00408	0.00414	0.02248	0.01066	0.01300	0.05011
F_1	0.00411	0.00610	0.03434	0.01081	0.01788	0.06526
F_2	0.00408	0.00427	0.02747	0.01066	0.01353	0.05933
G_2	0.00408	0.00425	0.02607	0.01066	0.01328	0.05536
H_2	0.00408	0.00400	0.02148	0.01067	0.01236	0.05161
E_1^*	0.00412	0.00591	0.03268	0.01082	0.01700	0.05972
E_2^*	0.00408	0.00484	0.02266	0.01079	0.01420	0.05167
H_2^*	0.00408	0.00463	0.02349	0.01072	0.01381	0.05469
area:	4737	4750	4788	5237	5250	5328
$\kappa_{\min}$:	-0.365	-1.691	-7.727	-1.203	-4.330	-22.044
$\kappa_{\max}$:	0.116	0.628	3.199	0.246	1.239	8.259
$\nabla^2 \kappa_{\min}$:	-0.372	-3.480	-103.060	-3.501	-26.911	-781.197
$\nabla^2 \kappa_{\max}$:	0.984	9.186	229.124	10.267	69.875	1593.366

6. Conclusion

By assuming that the quasi-diffusive filtering operations employed to generate covariances of an optimal data assimilation can be reformulated from the present model of variable diffusivity in an Euclidean geometry to an alternative model of uniform isotropic diffusivity in a Riemannian geometry, we have established the first steps toward the development of a general asymptotic procedure, known as the 'parametrix expansion' method, for estimating the amplitude quotient by which the actual diffused amplitude of each smoothing kernel differs from the amplitude calculated on the basis of a purely local Gaussian idealization of the kernel.

In applications, there is no need for numerical differentiation to be performed in the local normal coordinates; the analysis grid itself provides the framework in which to perform all numerical differencing. The quantities evaluated by this means are the appropriate tensor representation of the curvature, together with a sufficient number of its covariant derivatives. Being tensorial quantities, these can be transformed easily to the local normal frame representation. However, even this step can be avoided since the amplitude quotient's expansion coefficients ultimately reduce to scalar invariants of these tensorial diagnostics.

In two dimensions (2D), intrinsic curvature is expressible as a single scalar 'Gaussian curvature' field, and the diagnostics vector can be constructed from this evaluation and its successive (covariant) derivatives in appropriate local normal coordinates. Rotating the local coordinates to make the eigenvectors of the Hessian of the Gaussian curvature lie along the two normal coordinate axes does lead to a modest reduction in the complexity of the calculations needed to derive the first few expansion coefficients; more rewarding symmetries are available in the case of three dimensions (3D). However, the high level of algebraic complexity involved beyond

TABLE 4. RESULTS EQUIVALENT TO THOSE OF TABLE 2 BUT FOR REALIZATIONS OF THE GEOMETRY DERIVED FROM THE RANDOM NUMBER SEQUENCE INITIALIZED BY A VALUE OF THE 'SEED' EQUAL TO 3

$[S, V]$:	[0.32, 0.3]	[0.16, 0.3]	[0.08, 0.3]	[0.32, 0.6]	[0.16, 0.6]	[0.08, 0.6]
A_0	0.01401	0.04376	0.13562	0.03020	0.08422	0.23768
B_0	0.01006	0.02487	0.06555	0.01968	0.04484	0.09820
E_1	0.00423	0.00618	0.02980	0.00756	0.01693	0.05796
E_2	0.00420	0.00436	0.02049	0.00728	0.01273	0.04804
F_1	0.00423	0.00627	0.03407	0.00757	0.01759	0.06746
F_2	0.00420	0.00439	0.02551	0.00728	0.01354	0.05942
G_2	0.00420	0.00439	0.02361	0.00728	0.01332	0.05233
H_2	0.00420	0.00426	0.02147	0.00729	0.01229	0.04922
E_1^*	0.00424	0.00618	0.02980	0.00756	0.01756	0.05796
E_2^*	0.00421	0.00525	0.02072	0.00756	0.01497	0.04855
H_2^*	0.00421	0.00493	0.02158	0.00751	0.01455	0.05074
area:	4342	4529	4634	4680	4915	5066
$\kappa_{\min}$:	-0.396	-1.336	-10.219	-1.082	-3.564	-33.958
$\kappa_{\max}$:	0.115	0.448	3.478	0.244	1.128	8.877
$\nabla^2\kappa_{\min}$:	-0.235	-5.378	-141.152	-1.452	-35.139	-910.978
$\nabla^2\kappa_{\max}$:	0.617	10.593	383.713	4.118	89.318	3216.437

obtaining the first few asymptotic coefficients makes the derivation of the higher-order terms practical only through the application of mechanized algebraic methods, at which point, the adoption of special frame orientations confers little practical advantage.

In 3D, intrinsic curvature is entirely expressible in terms of a symmetric second-rank 'Ricci tensor'. This and other other important tensors, and the related ideas of differential geometry, are explained in several standard texts, such as Lovelock and Rund (1989), and are reviewed in the companion, Part II, of this note. The 3D Ricci tensor possesses six components that are independent at a given point, but its covariant derivatives conform to three additional constraints, the 'differential Bianchi identities'.

In future it may be required to develop the computational machinery to produce inhomogeneous and anisotropic quasi-Gaussian covariances of model error in both time and space, although there is no urgent requirement to do this at present. Nevertheless, the same approach can be applied. In 4D, however, it is no longer possible to condense all the curvature information into second-rank Ricci tensor, since the full curvature at a given point possesses more (20) degrees of freedom than can be accommodated in a symmetric rank-two tensor (10). In 4D, or in higher dimensions, it is the rank-four Riemann curvature tensor that completely expresses the necessary information.

In two dimensions, we used specialized geometrical arguments to derive a second order asymptotic expansion, expressed only in terms of intrinsic curvature parameters, which is complete up to this order. We have used analytical methods for the special geometries of constant curvature to establish that the expansions are necessarily asymptotic rather than convergent and that, when the expansions do converge, they do not necessarily converge to the true ampli-

tude quotient. The constant curvature examples of section 3 nicely illustrate these points. For uniform negative curvature, the isotropic geometry in N dimensions is the unique 'hyperbolic space' (sometimes referred to as the 'pseudo-sphere'). These families of constant-curvature isotropic geometries lend themselves to analytic descriptions of the true amplitude quotients and their asymptotic approximations. These therefore provide concrete examples in which the potential of the asymptotic methods can be appraised at different degrees of implied curvature, thereby helping to establish realistic expectations about what the asymptotic method is capable of estimating, and how large the corresponding errors are likely to be.

A tool used in the construction of asymptotic coefficients from the point of view provided by the positive-curvature geometries is the one-sided form of the classical Euler-Maclaurin summation formula (Abramowitz and Stegun 1970), a formula usually associated with problems of quantifying truncation errors in numerical analysis (e.g., Johnson and Riess 1982) but which has other widespread application in both pure and applied mathematics. The summation in this case leads to a convergent result, the true amplitude quotient, and it is through the Euler-Maclaurin method that we identify this solution with an approximating integral, plus an asymptotic series of correction terms. (This reverses the usual roles of summation and integration in numerical analysis applications, where it is normally the integral that is assumed 'true' and the summation assumed to be 'approximate'.) The summation in our context comes from the summation of the discrete eigenmodes (at the 'pole') of the Laplacian operator– an application of the theorem of Mercer (1909)[‡] . In the hyperbolic geometries, where the eigenmodes form a continuum, the analogous construction of the true solution is in the form of an integral representation.

In experiments to implement these methods in a practical way, we have found techniques that help us to keep the residual errors acceptably small. The three practical methods 'E_1^*', 'E_2^*' and 'H_2^*', described in the previous section, which prevent the asymptotic correction terms from becoming wildly erroneous remain perfectly consistent with the first- and second-order asymptotic expansions. By evaluating all their parameters purely locally, they remain true to the spirit of the parametrix expansion method, which also consider only local diagnostic of curvature and its higher derivatives. The defects of the naive methods, which we have been able to remedy through the employment of 'saturation functions' introduced in the last section, can be traced to changes in the aspect tensor (which we now interpret as the metric tensor) occurring over a spatial distance smaller than the aspect tensor itself (that is, smaller than a non-dimensional unit distance, in the Riemannian metric interpretation of the aspect tensor). The more general form of the parametrix expansion method has been studied by mathematicians and statisticians for reasons unrelated to our own practical applications (for example, see Gilkey 1984, Lafferty and Lebanon, 2005, Rosenberg 1997). These technical refinements are beyond the scope of this Part I, since they require a facility with the tensorial tools of general Riemannian geometry. The necessary geometry is reviewed, and the chosen notational conventions (which tend to vary with authors' preferences) established in the sequel, Purser (2008; 'Part II'), to which we also defer discussion of the construction of the first few terms of the full asymptotic expansion. Thus, it might be imagined that an alternative way to make the asymptotic expansion robust for such challenging implied geometries would be to

---

[‡] I am grateful to Dr. John Norbury, Oxford University, for this reference.

condition the asymptotic formulas used on diagnostics of curvature that have themselves been filtered and smoothed by the exact same covariance filters that we are trying to normalize (a 'Cressman' style of empirical smoothing needs no exotic normalization, so we are not entering a vicious circle by attempting this). This approach might well be worth further investigation, but it is beyond the scope set for the present study.

The essential structure of the parametrix method needs no fundamental change when we relax the restriction of radial symmetry in two dimensions; the single radial powers index, p, of section 4 is merely replaced by the pair of spatial indices, p, q, for the powers of x and y respectively, where these coordinates are the Riemann normal coordinates we define in Part II. In 3D, the third exponent, r, would be for the third coordinate z. The slab index then becomes $h = 2s + p + q$ (in 2D), or $2s + p + q + r$ (in 3D), but the decomposition of terms into: those linear in T, grouped as '$\mathcal{T}$'; the purely geometrical fixed forcings, $\mathcal{G}$; and the bilinear combination, $\mathcal{Q}$, of geometric terms with modulating function terms; persists with only the relatively superficial changes to their composition. Obviously, the geometrical terms in the more general case are more complicated. But the principle underlying the method of solution for the successive approximations, $A^{(h)}$ in the general case, in any number of spatial dimensions, remains the same. Thus, to develop the general form of the parametrix expansion method unrestricted by special symmetries of the kind we have dealt with in this Part I, it is actually more convenient to deal with the generic case of non-specific dimensionality and to embrace the full generality of Riemannian geometry. This includes more general representations of the concept of 'curvature', which involve 'shape tensors' of even higher rank than the well-known rank-four Riemann-Christoffel curvature. Part II sets out the relevant tensorial and geometrical ideas and notations, and takes the necessary steps in the construction of the generic second-order parametrix expansion. The results of the experiments of section 5 of this Part I suggest that, in practice, there would be very little (if any) advantage in carrying out the parametrix expansion beyond the second order. This would only be justified were it necessary to obtain amplitude estimates: (i) of very high numerical precision; (ii) in circumstances where the given variations of aspect tensor are extremely gradual and smooth enough to enable the higher order terms to be beneficial. Nevertheless, we hope that the tensorial development of the theory we set out in Part II will make the necessary steps involved in higher order extensions of the parametrix expansion a little easier to formulate and excecute, should this extension ever be needed.

Factorial and Gamma function identities

The factorial function on integers is defined recursively,

$$
0! \;=\; 1, \tag{A.1a}
$$

$$
N! \;=\; N(N-1)! \tag{A.1b}
$$

and, through its identification with Euler's Gamma function,

$$
N! \equiv \Gamma(N+1), \tag{A.2}
$$

defined by the integral,

$$
\Gamma(a) = \int_0^\infty z^{a-1} e^{-z} dz, \tag{A.3}
$$

it is consistently generalized to other real or complex arguments (except where the a in (A.3) is a non-positive integer). The non-integer extension is nicely exemplified by the evaluation of $\Gamma(\frac{1}{2}) = (-\frac{1}{2})!$:

$$
\begin{aligned}
\left(-\frac{1}{2}\right)! &= \int_0^\infty z^{-\frac{1}{2}} e^{-z} dz \\
&= 2 \int_0^\infty e^{-x^2} dx \\
&= \left[\int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(x_1^2 + x_2^2)} dx_1\, dx_2 \right]^{\frac{1}{2}} \\
&= \left[\pi \int_0^\infty 2r e^{-r^2} dr \right]^{\frac{1}{2}} \\
&= \pi^{\frac{1}{2}}.
\end{aligned} \tag{A.4}
$$

Extending the same style of construction gives us the measure, S_N, of the unit N-sphere:

$$
\begin{aligned}
\left[\left(-\frac{1}{2}\right)! \right]^{(N+1)} = \pi^{(N+1)/2} &= \int_{-\infty}^\infty \left(e^{-x_1^2} \right) \dots \left(e^{-x_{N+1}^2} \right) dx_1 \dots dx_{(N+1)} \\
&= S_N \int_0^\infty r^N e^{-r^2} dr \\
&= S_N \int_0^\infty \frac{1}{2} z^{(N-1)/2} e^z dz \\
&= S_n \frac{\left(\frac{N-1}{2} \right)!}{2},
\end{aligned}
$$

or,

$$
S_N \equiv \frac{2\pi^{(N+1)/2}}{\left(\frac{N-1}{2} \right)!}. \tag{A.5}
$$

An important identity, easily verified for integer/half-integer arguments, is:

$$\frac{(2N)!}{N!(N-\frac{1}{2})!} = \frac{2^{2N}}{\pi^{\frac{1}{2}}}. \tag{A.6}$$

For special values of the complex argument (where traditionally the 'Gamma' version of the notation is employed) with integer or half-integer real parts, we can always use iterations of:

$$|\Gamma(m+1+ik)|^2 = (m^2+k^2)\,|\Gamma(m+ik)|^2\,, \tag{A.7}$$

to obtain a formula for $|\Gamma(m/2+ik)|^2$, for any integer m, using one of the standard forms:

$$|\Gamma(ik)|^2 = \frac{\pi}{k\sinh(\pi k)}, \tag{A.8a}$$

$$\left|\Gamma\left(\frac{1}{2}+ik\right)\right|^2 = \frac{\pi}{\cosh(\pi k)}. \tag{A.8b}$$

References

Abramowitz, M., and I. A. Stegun   1970   *Handbook of Mathematical Functions*, Dover, New York. 1046 pp.

Derber, J. C., and A. Rosati   1989   A global ocean data assimilation system. *J. Phys. Ocean.,* **19,** 1333–1347.

Erdelyi, A.   1955   *Asymptotic Expansions* Dover, New York.

Gilkey, P. B.   1984   *Invariance Theory, the Heat Equation, and the Atiyah-Singer Index Theorem.* Publish or Perish, Wilmington, DE, 345 pp.

Hilbert, D.   1912   *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Teubner, Leipzig and Berlin, 282 pp.

Hogg, R. V., and A. T. Craig   1978   *Introduction to Mathematical Statistics, Fourth Ed.* Macmillan, New York, 438 pp.

Johnson, L. W., and R. D. Riess   1982   *Numerical Analysis, Second Ed.* Addison-Wesley, Reading, MA. 563 pp.

Kreyszig, E.   1991   *Differential Geometry* Dover, New York. 352 pp.

Lafferty, J., and G. Lebanon   2005   Diffusion kernels on statistical manifolds. *J. of Machine Learning Research,* **6,** 129–163.

Levi, E. E.   1907   Sulle equazioni lineari totalmente ellittiche alle derivate parziale. *Rend. Circ. Mat. Palermo,* **24,** 275–317.

Lovelock, D., and H. Rund   1989   *Tensors, Differential Forms, and Variational Principles*, Dover, New York. 366 pp.

Mercer, J.,   1909   Functions of positive and negative type and their connection with the theory of integral equations. *Phil. Trans. Roy. Soc. London,* **A 209,** 415.

Purser,R. J.   2008   Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry; Part II: Riemannian geometry and the generic parametrix expansion method. NOAA/NCEP Office Note 457.

Purser, R. J., and L. M. Leslie   1988   A semi-implicit, semi-Lagrangian finite difference scheme using high-order spatial differencing on a non-staggered grid. *Mon. Wea. Rev.,* **116,** 2069–2080.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts   2003   Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.,* **131,** 1536–1548.

Rosenberg, S.   1997   *The Laplacian on a Riemannian Manifold; London Mathematical Society Student Texts 31.* Cambridge, 174 pp.

Weaver, A., and P. Courtier   2001   Correlation modelling on the sphere using a generalized diffusion equation. *Quart. J. Roy. Meteor. Soc.,* **127,** 1815–1846.