Covariance modelling in a grid-point analysis

R. James Purser,

SAIC at NOAA/NCEP/EMC

Manuel de Pondeca (SAIC/EMC),

David F. Parrish (EMC),

Dezso Devenyi (CIRES/NOAA/ESRL).

Topics we discuss in this talk

- The recursive filter technique for simulating a quasi-diffusive process
- The “TRIAD” and “HEXAD” algorithms for anisotropic filtering
- “Blending” refinements to improve continuity and smoothness
- The problem of filter normalization in inhomogeneous cases
- Applications of differential geometry and the “Parametrix” method to improving the normalization
- Covariance synthesis by multi-filter superposition in a multigrid framework
- Analysis error estimation and characterization, and preconditioning
- Analyzing variables with abrupt gradients
- Conclusions

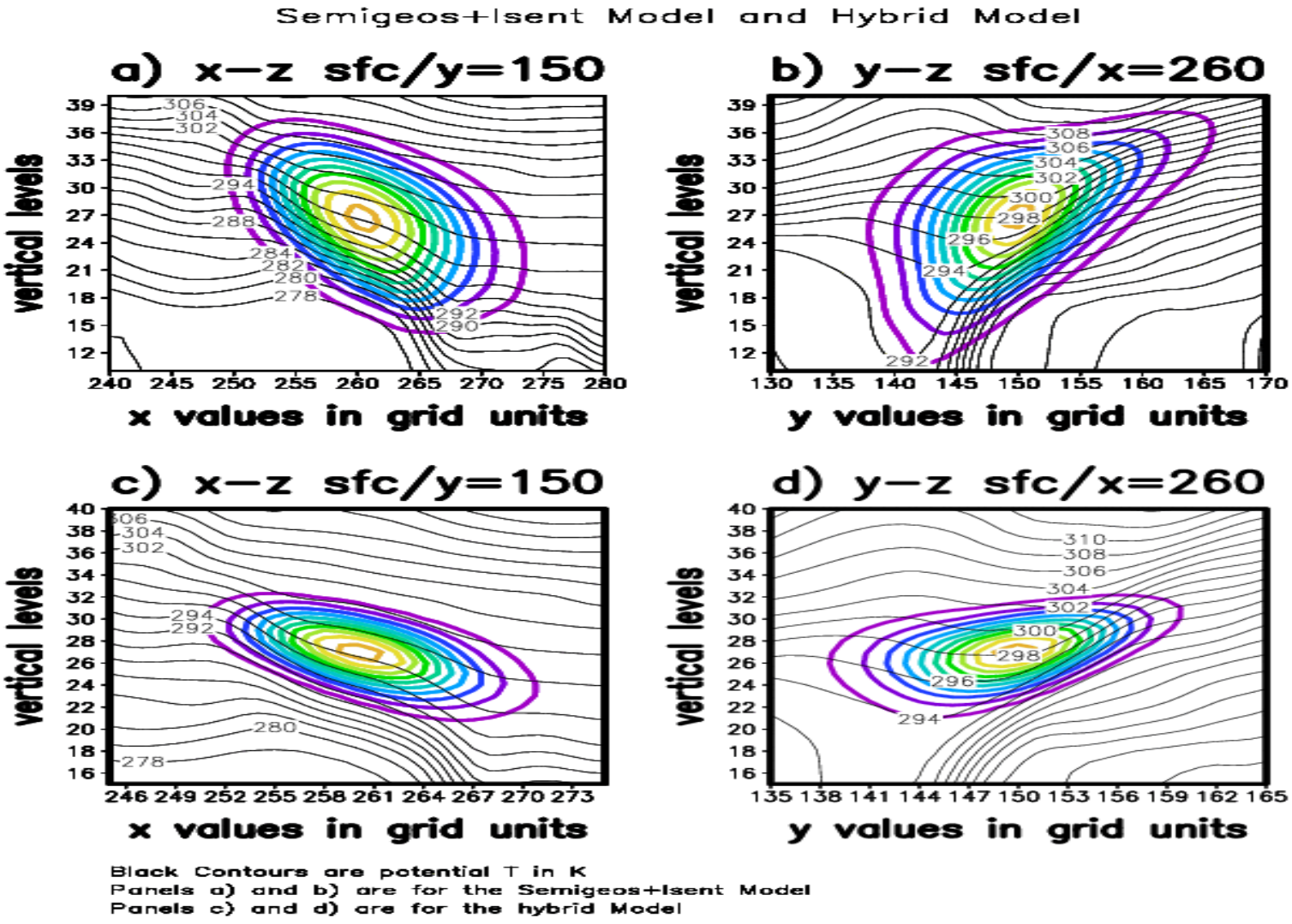
The Recursive Filter

The recursive filter is an efficient numerical technique for simulating diffusion along a grid line. Covariance operators can be constructed with it, as is done in NCEP's Gridpoint Statistical Interpolation (GSI) and other assimilation schemes.

The idea is to take a finite-degree Taylor series approximation to the Exponential function. Then apply this function to an effective Diffusivity*Laplacian*Duration. In finite difference terms, this is a band-matrix operator. Invert the operator, and you have an approximation to the result of diffusion applied for the given Duration. The centered and normalized second moment (like squared "radius of gyration") gives the spread or "Aspect Tensor" which, if the Duration is chosen to be a half, and the diffusivity is uniform, is simply the Diffusivity itself.

The idea of using a diffusive process to generate a covariance contribution is made explicit in the assimilation techniques of Derber and Rosati (1989), extended to anisotropic diffusion on the sphere by Weaver and Courtier (2001).

Examples of vertical sections through 3D covariances generated through adaptations of the Desroziers and Riishojgaard methods.



The general diffusion operation in N dimensions can be formally broken down into N line operations applied sequentially (as in a splitting method). Some of the lines may be oblique in the anisotropic case.

Then the individual line operations which, we recall, take the form of the inversions of band-matrix systems, can be solved recursively in two sweeps, since the symmetric band matrix admits an “L-U” Cholesky decomposition. (There are minor technicalities involving boundary conditions, but these are most conveniently swept away by assuming mirror-reflection conditions at boundaries, possibly slightly outside the physical domain of interest.)

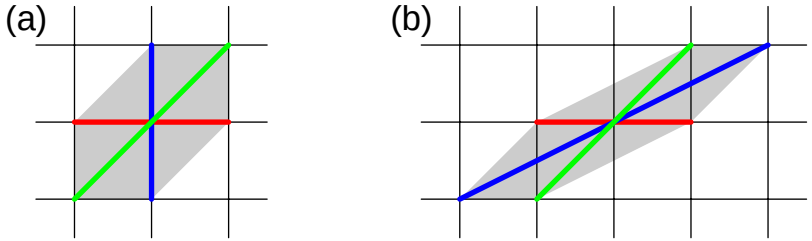
Difficulties can arise if the degree of the Exponential-approximating polynomial is too high and the aspect/diffusivity tensor too large compared to the scale of the grid; the system inverted then becomes ill-conditioned. But such a polynomial can always be factored into real-coefficients polynomials of degree not exceeding two, and this factorization solves the conditioning problem.

Triad and Hexad Methods

Although a general N-dimensional anisotropic diffusion operator may be broken into N sequential oblique operators, the directions cannot all be aligned with the grid in general. An attractive alternative is to employ a larger set of directions, 3 in two dimensions, 6 in three dimensions, such that even the oblique lines do at least become generalized grid lines.

This is the essential idea behind the Triad and Hexad methods (and their generalizations in higher dimensions).

The best choice for triads involves three alignments whose smallest steps, or “generators” form the smallest possible grid triangle. Equivalently, triad directions lie parallel to the diameters of a smallest possible grid-hexagon.

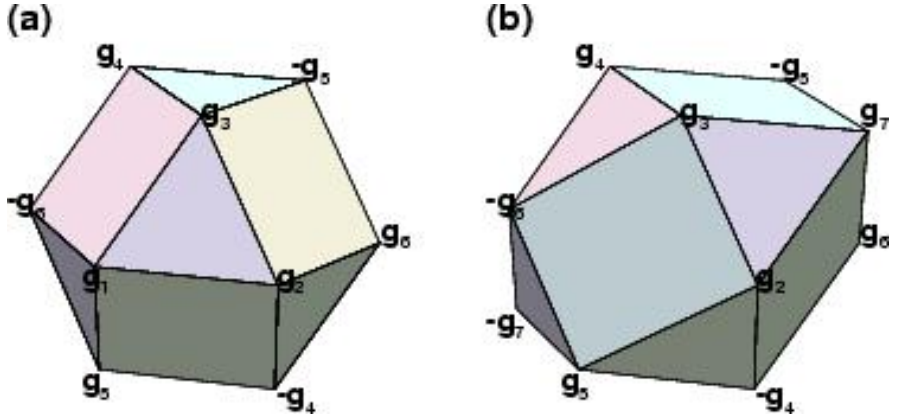


It is easy to see that, for such a grid hexagon, keeping two vertex-diameters fixed leaves two alternatives for the third (always!). The “projection” of the given aspect tensor onto the “spreads” of the line operators is essentially a linear one, since normalized centered second moments add under pure convolution (sequential filters) and approximately so even when the inhomogeneities render the convolutions “impure”. But, in general, the three spreads resulting from the projection of the desired aspect tensor into any given triad are likely to leave one of these spreads negative, which cannot correspond to any sensible smoothing or diffusing operator.

The Triad Algorithm responds to the occurrence of negative spread in a trial triad by replacing the offending spread’s direction by that given by the alternative diameter that will allow a minimal grid-hexagon to appear. The new spread (from a new projection) is then always non-negative, although one of the other two spreads will typically have become negative through this transition; in that case, the process is iterated until a valid solution (three spreads non-negative) is found. Such a solution always exists and is unique at each point.

The Hexad algorithm abides by the same general principle, but now six directions are needed to provide the necessary degrees of freedom in the symmetric aspect tensor in three dimensional space. Geometry dictates that these directions correspond to those of the diameters of the smallest skewed grid “cuboctahedron” (a figure with 12 vertices, 6 quadrilateral faces and 8 triangular faces.

Again, lattice geometry forces unique rules for the transition between one hexad and its alternative, once it is decided which diameter is free to change while the remaining five others are kept fixed.

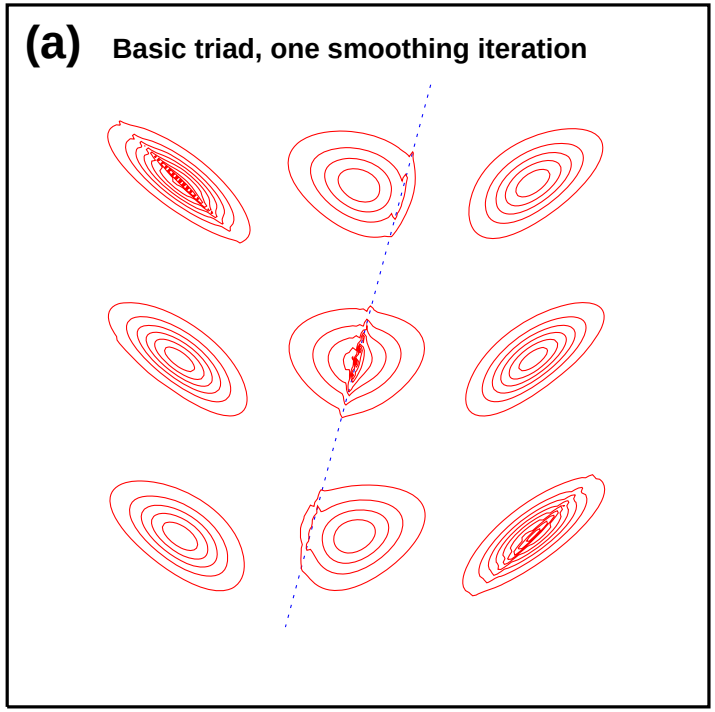


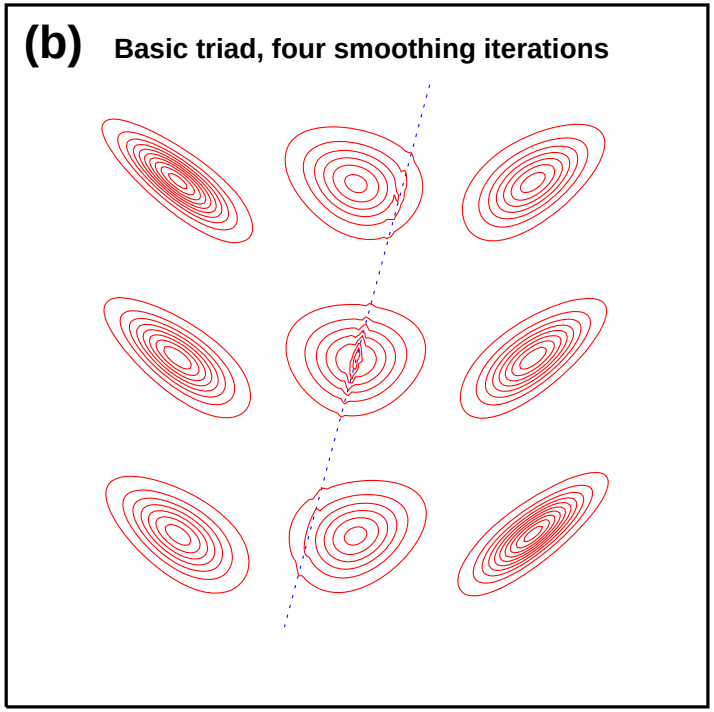
The triads or hexads and their accompanying spreads will clearly change with geographical location in the usual case of an inhomogeneous covariance. Sequencing the line operations involved among the various triads or hexads so that there is no mutual interference of the filters whose lines of operation intersect is therefore a nontrivial problem when these operations are multi-processed.

The solution is provided by an idea from abstract algebra. The generators of a lattice can be “colored” using the non-vanishing elements of a suitable “Galois Field”. The colors form a periodic pattern (the period being some prime number) over the generator lattice. For example, every generator of a basic triad is a different non-null element of GF(4). In three dimensions, the simplest relevant Galois field is GF(8), which provides seven “colors”, of which, the generators of any hexad will have six.

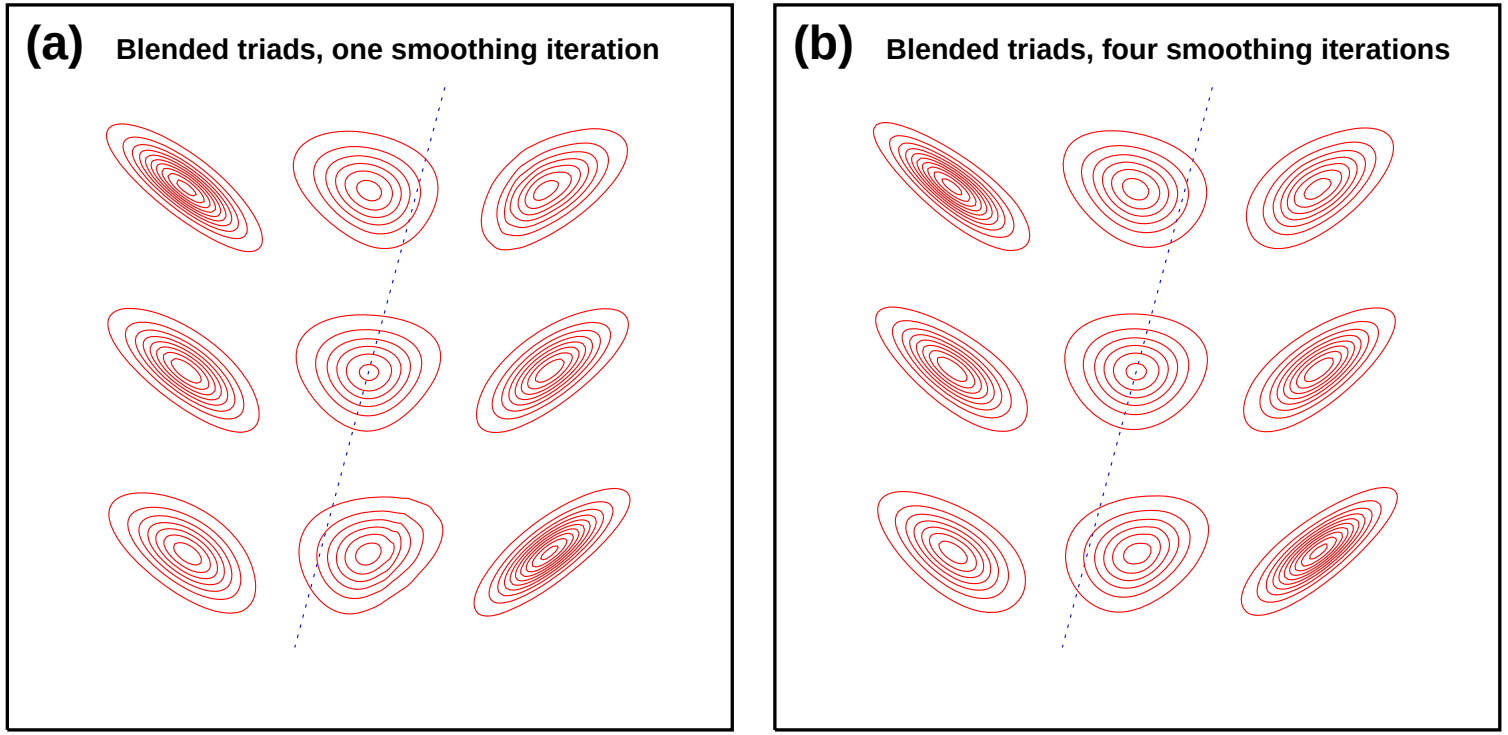
Thus, it is enough to perform all operations of each color concurrently, and to run through the colors in sequence, to avoid a numerical clash of filters.

Early experiments with the Triad and Hexad methods revealed ugly numerical artifacts in the resulting covariances at places where triad or hexad changed. The cause was an excessive rate of change in the line-spreads on approaching the transition.





The cure involves enlarging the allowed set of smoothing directions at each point so that the single triad or hexad can be replaced by an equivalent symmetric superposition of triads or hexads that form a “ball” in aspect tensor space centered on the original aspect tensor “point”. Again, we invoke the (approximate) linearity of spreads to justify this remedial measure.



The result is the “Blended Triad” and “Blended Hexad” methods. To accommodate the larger sets of directions needed for conflict-free parallel algorithms, we now find that the eight non-null elements of GF(9) can be paired up to provide four “colors” for the blended triad, while the 26 non-null elements of GF(27) pair up to provide the 13 “colors” for a blended hexad.

Details of all these methods are available in Wu et el. (2002, MWR), Purser et al. (2003a, b, MWR) and in Purser (2005, NOAA/NCEP Office Note 447).

Normalization

A homogeneous diffusive process in Euclidean geometry and without boundaries is easy to normalize whether or not it is isotropic – the final function has a standard Gaussian form and the amplitude, for aspect tensor A, will be |2*pi*A|^{-N/2} in N dimensions. Weaver and Courtier address the problem in the spherical case. The difficulty occurs when the diffusivity varies or the domain is significantly non-Euclidean. At present we are using a randomized trace Monte Carlo method, but the cost is substantial if we want to adapt to a different covariance from one analysis time to the next.

In fact, the case of varying diffusivity can be mapped into an equivalent problem of **uniform** and **isotropic** diffusivity in a non-Euclidean (Riemannian) space, albeit with a spatial “capacitance” appearing in the transformed diffusion problem. (Capacitance may be thought of as a varying gauge of metal forming a thin smoothly curved shell in the 2D heat-flow analogy.) But what is gained by such an algebraic manoeuvre?

First, we note that, since we are not simulating “real” diffusion but merely taking advantage of a convenient numerical way to generate self-adjoint quasi-Gaussian covariance contributions, we are at liberty to choose the capacitance in the **untransformed** problem that anticipates a uniformity of this quantity once transformed. The original aspect tensor effectively assumes the role of contravariant metric tensor in the transformed problem in which diffusivity is unity isotropically. What we gain is that the amplitude can now only depend upon the distribution of (intrinsic) **curvature** of the transformed space – a second-derivative of the metric when examined in local (Riemann) “normal coordinates”. In a way, we have simplified the factors upon which the amplitude can depend by identifying the source of the amplitude variation as the curvature of the transformed space.

Note that some choices of aspect tensor distribution, such as those which would appear homogeneous in semi-geostrophic space, as suggested by Desroziers (1997), will actually solve the normalization by implying a transformed space that is still almost flat!

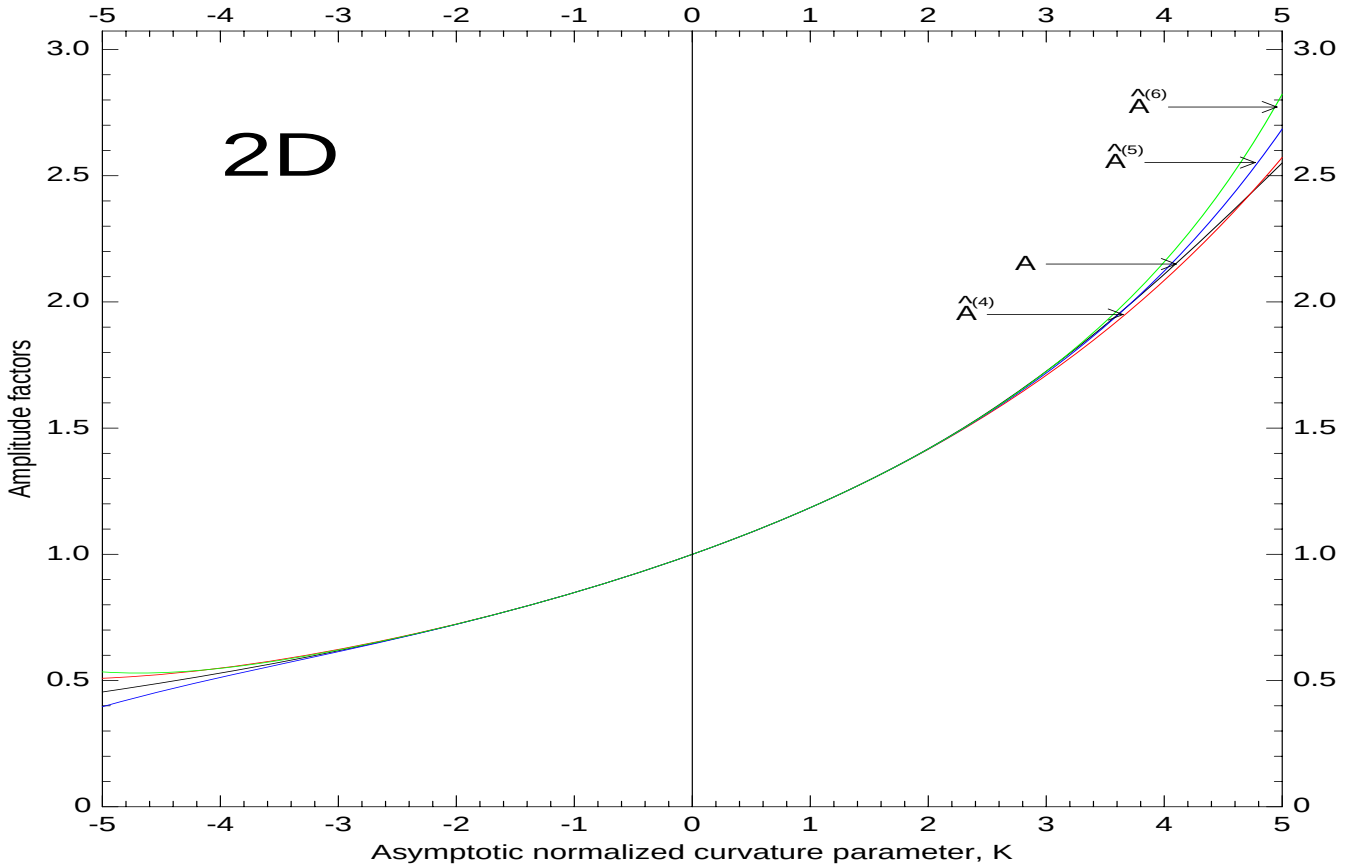
Differential Geometry and the "Parametrix" Method

In the general case, though, the transformation to a Riemannian geometry has not made the amplitude estimation problem an easy one. However, there does exist a literature on the estimation of solutions to the "Heat Kernel" in curved spaces (for example, Rosenberg 1997: The Laplacian on a Riemannian manifold, C.U.P.) and one approach, although it can only lead to an asymptotic approximation at each point, looks promising.

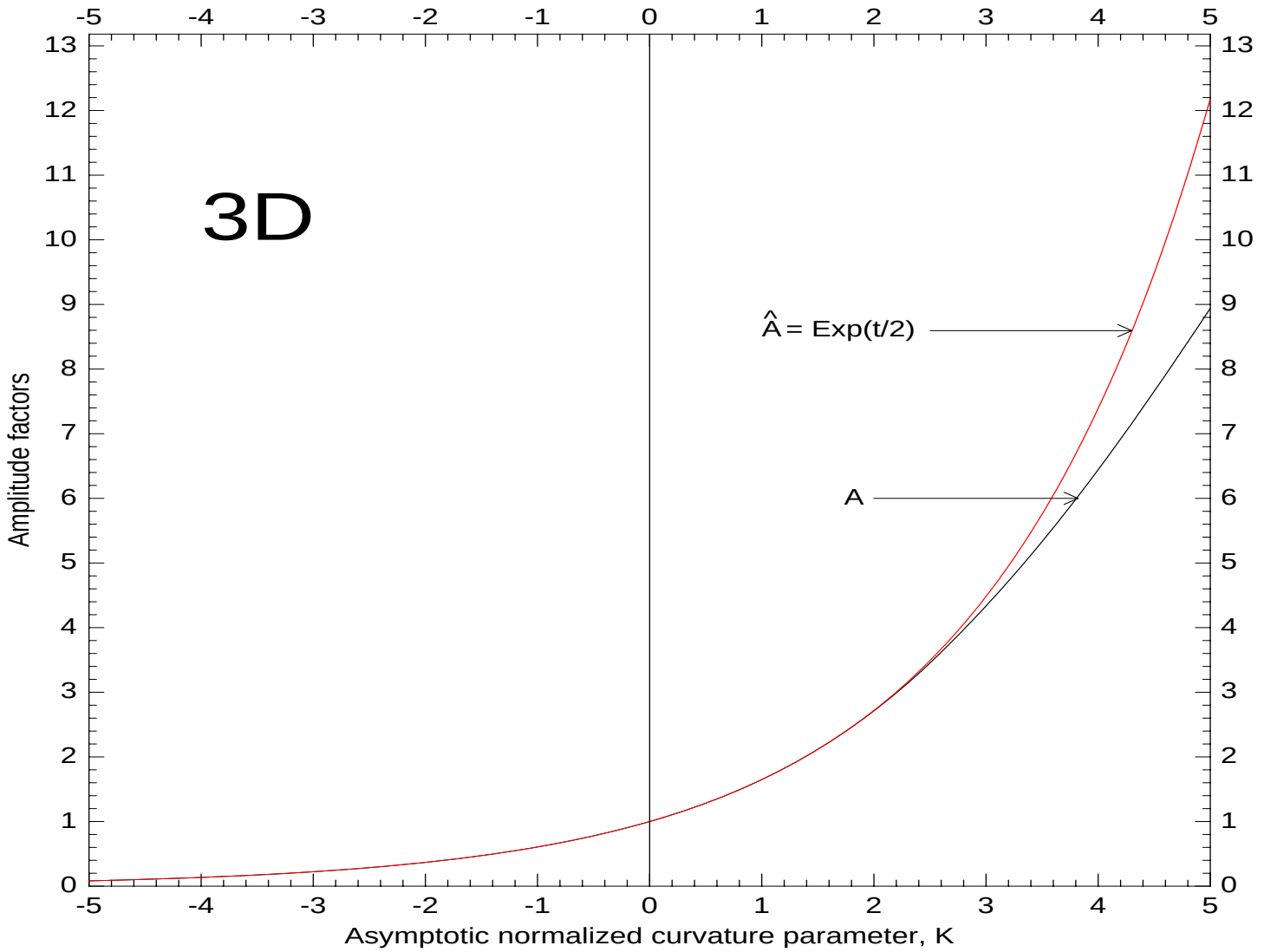
This approach is the "Parametrix method". Normal coordinates are constructed to be "as Cartesian as possible locally"; axes are orthogonal and the radials through the local origin are geodesics on which distances measure true. The idea behind the parametrix method is to represent the evolving solution of the diffusion problem as the corresponding Euclidean solution in the normal coordinates, multiplied by a modulating function that is smooth in "time" and space.

Then, for sufficiently small “times” we should be able to approximate the modulating function in a series of powers of the normal coordinates and “time”. The only part of the solution we shall end up being interested in is the solution at the origin and at “time” = ½ (when the aspect tensor corresponds with the diffusivity). There is a systematic recursive algorithm for generating successive approximations in finite powers where one order of solution, fed back into original diffusion equation, provides the next order of correction.

While, in principle, this process can be continued without limit, in practice it rapidly leads to extremely complicated algebra and the successive approximations, typical of many asymptotic series, do not converge to the true solution for any finite time. Nevertheless, provided the original aspect tensor varies sufficiently smoothly and gradually, the approximate solutions obtained by this approach should be adequate. Most of the algebraic operations involved can actually be reduced to forms that lend themselves to mechanization. We give two idealized examples of the results obtained from the asymptotic method but only give an outline here of some of the steps required for the general case.



These graphs show the 2D results comparing the asymptotic expansion for the amplitude quotient with the true solution in the special case where the Gaussian curvature K is uniform. Even out to a curvature of +/- 5 non-dimensional units, the asymptotic method with a few terms should give a very good approximation, as shown. However, the expansion is formally divergent. The true amplitude quotient is denoted “A”; other graphs show asymptotic expansions truncated to the degrees indicated by the superscript.



Corresponding results in 3D, with uniform sectional curvature, K. But now, the asymptotic expansion converges to exp(K/2). For negative curvature (hyperbolic geometry) this is the exact solution; for positive curvature, there is an error that grows with K.

An outline of the treatment in the case of general Riemannian geometry follows.

First, we observe that, in normal coordinates we can express the covariant metric tensor as a Taylor series (starting with the identity, and with vanishing first-order term). The construction of the normal coordinate system, in which radials are geodesics measuring true at all distances out, implies that the radial vector is an eigenvector, with eigenvalue=one, of both metric tensors at every point. This imposes important constraints on all the quadratic, and higher, Taylor series coefficients of the metric; these become the celebrated “Bianchi identities” (algebraic and differential) when translated into the implications for the Riemann curvature tensor. (In fact, it is possible to express the array of quadratic Taylor series coefficients for the metric directly as simple linear multiples of the Riemann tensor.)

The Bianchi, and other symmetries, restrict the actual number of degrees of freedom at each degree of the Taylor series. For example, in 3D, we might expect the Taylor series for a symmetric metric tensor (6 components) to require 6*6=36 independent coefficients at second degree, but only 6 are actually needed. At 3rd degree, the Bianchi differential identities come into play to keep the independent coefficients at only 15.

The second derivatives of the covariant metric (in normal coordinate) provide the Riemann tensor. Successive covariant differentiations of the Riemann tensor produce new tensors which can be equated to tensorial expressions that involve only the Taylor coefficient arrays of the metric tensor up to a finite degree. However, it is the origin-evaluated derivatives of the metric tensor – i.e., these same Taylor coefficients, that force the successive terms in the parametrix method for approximating the modulating function relating the non-Euclidean solution of the diffusion problem to its standard Euclidean counterpart.

It therefore becomes possible to express the successive terms in the series expansion of the amplitude adjustment quotient directly in terms of the Riemann tensor and its first few covariant derivatives. These latter quantities are straightforward to evaluate on the original grid and, being tensorial, are therefore relatively easy to convert to normal coordinate representations if needed. However, contracted versions of the curvature: the "**Gaussian curvature**" in 2D; the "**Ricci curvature**" in 3D; should make it more convenient to express the approximations to the amplitude quotients directly in terms of these simpler quantities.

Covariance synthesis by multi-filter superposition in a Multigrid framework

Although discussion up to this point has focused on the technicalities involved in the construction of Gaussian contributions to the covariance, we should not ignore the fact that the simple quasi-Gaussian that results from a **single** application of a diffusion-simulating filter is probably **not** a very realistic model of background uncertainty by itself.

If we can afford to construct a superposition of such quasi-Gaussians, each at a different scale and relative amplitude, then we can sculpt a more promising and realistic covariance shape. (Actually, it is better to use this approach to construct the asymmetrical **square-root** of the covariance, as this way, it becomes much easier to ensure the final result is positive definite.)

The computational expense of applying the recursive filters repeatedly, but at different scales, could be greatly mitigated by adopting a multigrid strategy. Then, filtering at the coarse scales is done (cheaply) on only a correspondingly coarse grid. The contribution is interpolated to the next finer grid and added to a quasi-Gaussian contribution appropriate to this scale, and so on. Of course, the quantity being smoothed must also be carried consistently **up** this hierarchy in the first place, using the adjoint of the grid-refinement interpolators.

This is another area of covariance filter development in which we are beginning to make some headway, with preliminary tests in 2D.

This synthesis is closely related to discrete inverse Laplace transformation, as can be verified immediately in the case where the contributions are true Gaussians.

The method therefore enables a vast range of covariance shapes to be synthesized, and suggests that we might even approximate the analysis error covariance this way.

Analysis error estimation and characterization. Preconditioning

These topics are somewhat speculative at present and it may eventually turn out that the kinds of filters we have discussed are not adequate to characterize the analysis error structure. The problem is that, while it seems reasonable to represent a background error covariance in a way that implicitly assumes that, locally, it could be characterized simply by its power spectrum, the degree of spatial inhomogeneity in a typical analysis error covariance, so strongly dominated by equally inhomogeneous observations, make it seem unreasonable that the same kind of characterization of the analysis error covariance would work.

However, even with its many imperfections, it would be at least informative to find out how well a filtering approach to representing analysis covariances might be made to work, once the necessary drastic simplifications in the measurement precision operators are made.

In classical data assimilation theory, we understand the analysis precision to be simply the sum of the background precision and the measurement precision. If observations were spatially homogeneous, we could ascribe to them a local error power spectrum and obtain the corresponding analysis power spectrum fairly directly.

We ask: with realistic observation distributions, can we usefully impose this assumption anyway, and obtain anything of value from the resulting analysis covariance?

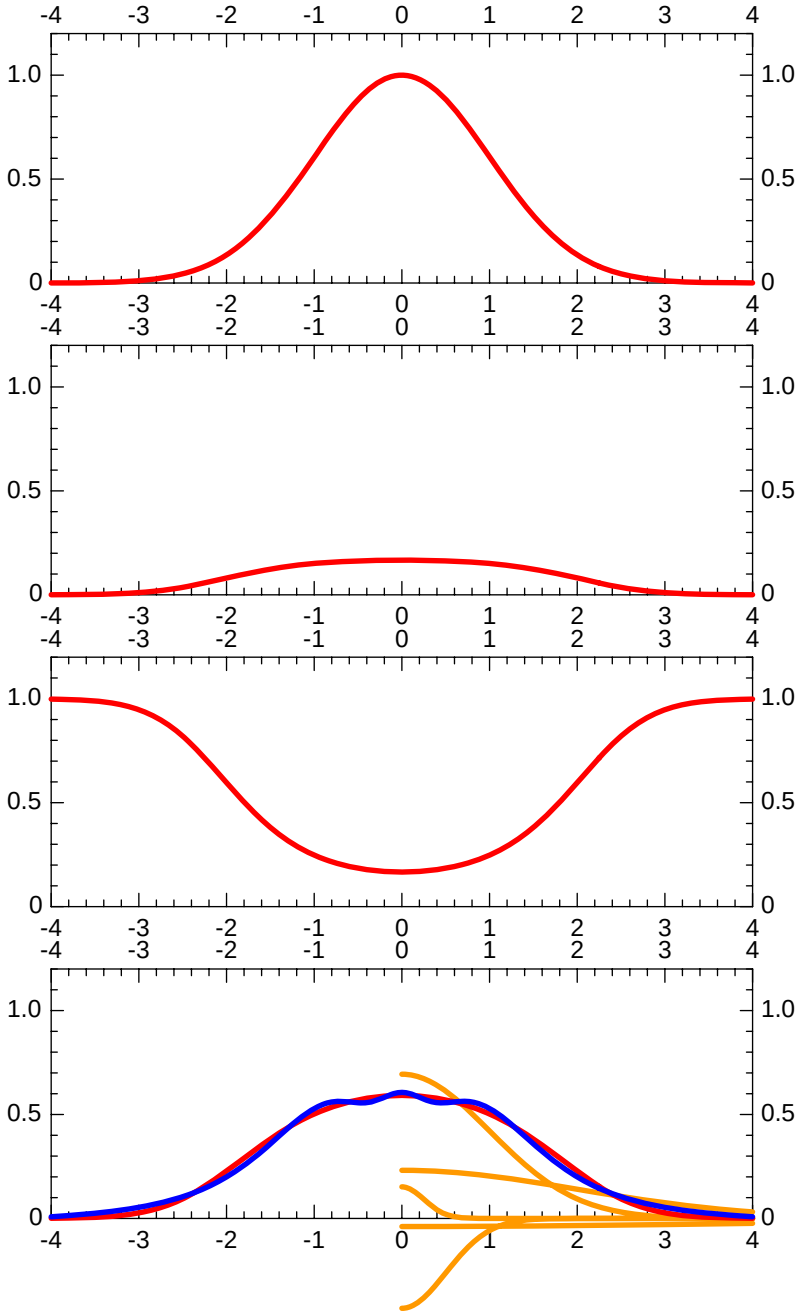
If so, filters might be useful for geographically smoothing a locally valid representation of each observation type's H^T*R^{-1}H operator to cause enough overlap among discretely distributed measurements of the same type to allow their measurement operators to **look** homogeneous locally and to make possible a local assessment of the density of such data.

For example, a scatter of point-observations of the analysis variable would then give (locally) a white-noise distribution for the measurement precision. A horizontal scatter of satellite radiance measurements would, however, result in a very different appearance for the vertical part of the local precision spectrum, with a bell-shaped peak concentrating the precision at only the small vertical wavenumbers (the squared-absolute magnitude of the Fourier transform of the “typical” transmittance function). Each of these two sets of data would have, locally, a different spectral impact on the analysis that could be estimated, and hence represented by a combination of filters.

If it is possible to obtain a reasonable model of the analysis error covariance in terms of a filter, then that same filter might be a useful preconditioner for the analysis itself.

Also, it would provide a valuable tool for normalizing the spread of ensemble members in a way that reflects the influence of new data, but sidesteps the expense of adding an assimilation to each ensemble component.

IN SITU MEASUREMENTS OF THE CONTROL VARIABLE



<- Spectral background covariance, B

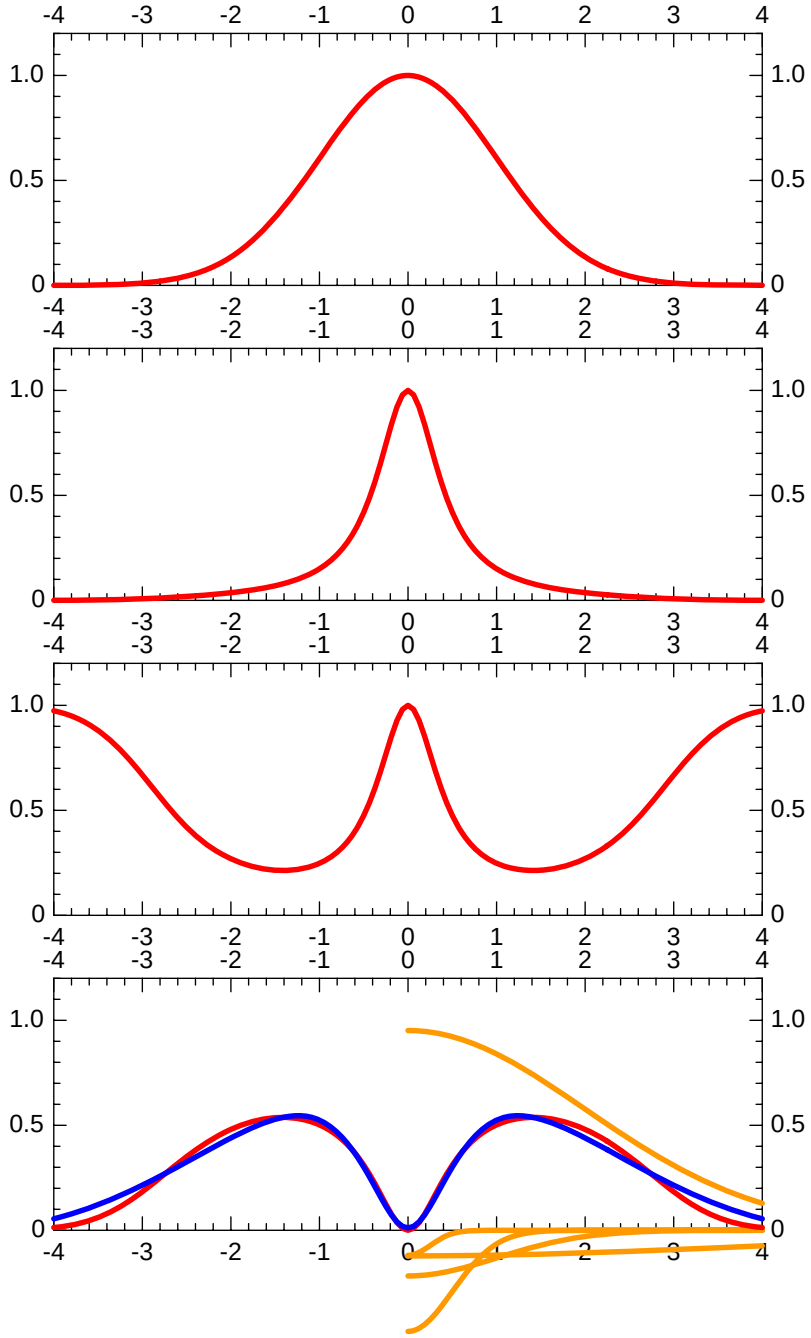
<- Spectral analysis covariance, A

<- Preconditioner, A/B

<- G = 1-sqrt(A/B) and
$\hat{G}$ = sum { Gaussians } (blue)

<- Gaussians (half-curves)

IN SITU MEASUREMENTS OF THE DERIVATIVE



<- Spectral background covariance, B

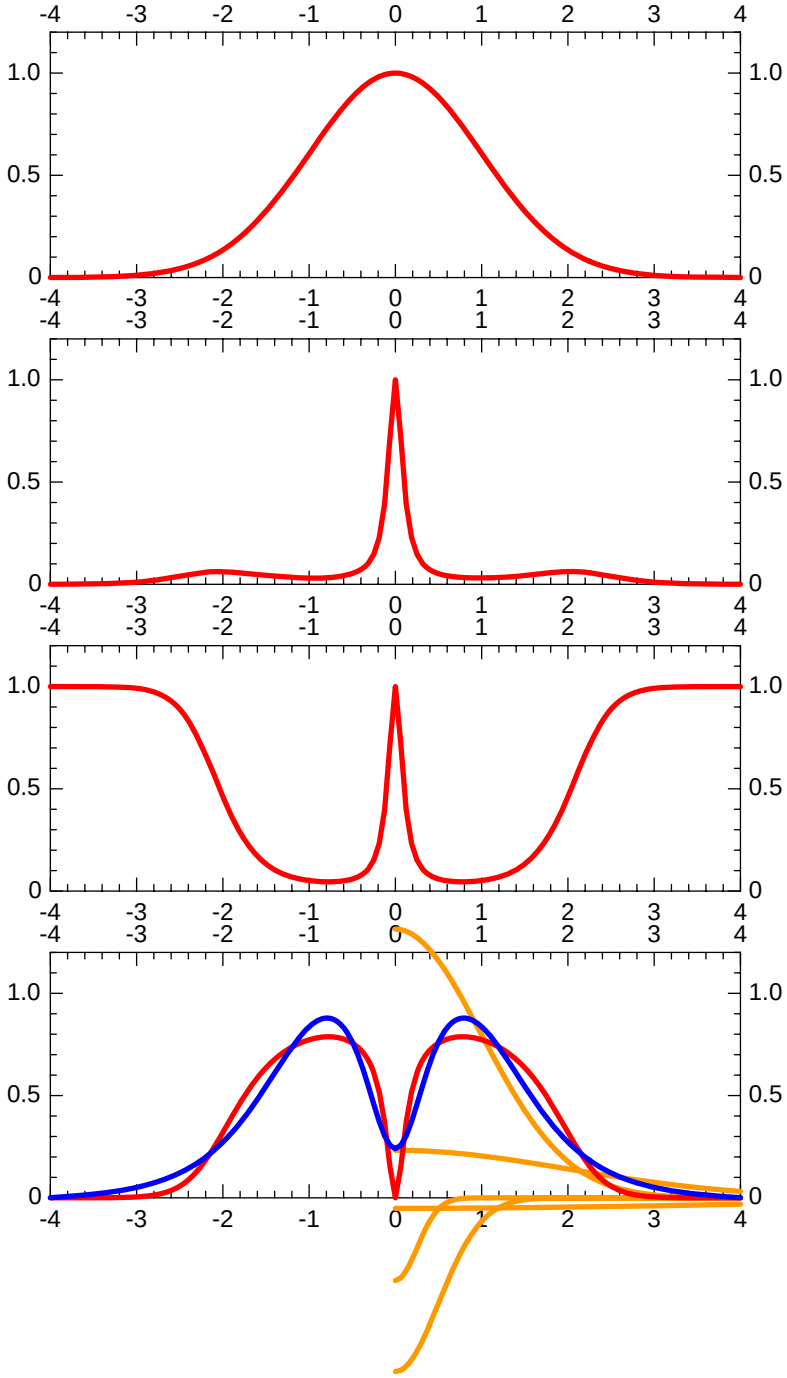
<- Spectral analysis covariance, A

<- Preconditioner, A/B

<- G = 1-sqrt(A/B) and
$\hat{G}$ = sum { Gaussians } (blue)

<- Gaussians (half-curves)

IDEALIZED RADIANCE MEASUREMENTS OF THE DERIVATIVE



<- Spectral background covariance, B

<- Spectral analysis covariance, A

<- Preconditioner, A/B

<- G = 1-sqrt(A/B) and
$\hat{G}$ = sum { Gaussians } (blue)

<- Gaussians (half-curves)

Analyzing variables with abrupt gradients

Earlier, we discussed the problem of adequately normalizing the amplitudes of filters when the aspect tensors vary gradually and smoothly with location, and mentioned that asymptotic methods show promise in performing this task efficiently.

In cases where the covariances change abruptly, the parametrix method will be of no use. However, some abrupt changes are tied to the changes of topographic parameters, such as elevation, or the land/water distinction.

NCEP now carries out an operational 2D “Real-Time Mesoscale Analysis” (RTMA) of surface analyzed parameters over the US where such topographically-tied covariance changes do occur (De Pondeca and co-authors 2007, 22nd Conf. WAF/18th Conf. on NWP, Park City, UT, AMS). One approach we are giving some attention is based on a variant of Riishojgaard’s (1998, Tellus.) method.

The analysis could be carried out once for an elevation (say) that is somewhat higher than the local average obtained by heavily smoothing the true elevation, so that the covariances on this hypothetical surface can themselves now be taken to be smooth. Another, completely independent analysis is performed at a parallel smoothed elevation that is somewhat lower than the actual local average. The extra vertical interpolations are computed in accordance with the prescribed part of the covariance which, as in the Riishojgaard method, depends upon the extra variable (in this case, elevation). Thus the observations are only permitted to influence the respective analyses in proportion to how close their actual elevations are relative to the nominal elevations of the two hypothetical surfaces.

The final analysis is similarly the interpolation between the two surfaces' analyses to a target at the true elevation at each geographical position. The result is expected to give realistically rapid changes in surface analyzed variables at places where the topography changes rapidly, while the underlying covariances seen by the generating filters remain perfectly smooth.

It might be possible to extend these ideas to 3D if the duplicated analysis domains are confined to only the lowest levels.

Conclusions

The numerical grid-smoothing technique of “recursive filters” has shown itself to be a versatile tool for the efficient construction of approximate models of background error covariances. In combination with the “Triad” and “Hexad” methods, we can now generate smooth quasi-Gaussian covariance contributions with arbitrary degrees of anisotropy. Asymptotic methods based on differential geometric ideas are being developed and tested in the GSI at NCEP to make the normalizaton of these filters in inhomogeneous cases more accurate and efficient, while multigrid methods are being developed to facilitate the efficient construction of covariances having more general profile shapes than can be obtained with only a few Gaussians.

Finally, we have discussed some more speculative ideas concerning how future developments of these techniques might address the problems of analysis error characterization, preconditioning and covariance modelling where abrupt changes occur.