Solving the Eigenvalue Problem for a Covariance Kernel with Variable Correlation Length

Wynand Verwoerd and Don Kulasiri
Applied Computing, Mathematics and Statistics Group, Lincoln University, Canterbury, New Zealand.

Abstract
In stochastic modelling of flow in porous media, the medium properties that produce random trajectories of fluid elements are modelled by the assumed correlation kernel. For numerical simulation of the flow, the stochastic differential equation (SDE) is expanded using a Karhunen-Loeve expansion in terms of eigenvalues and eigenfunctions of the correlation function. In real world problems such as contaminant transport in aquifers, the medium properties are themselves variable necessitating the solution of the eigenvalue integral equation with a variable correlation length. We investigate this problem by comparing several approximate approaches for an assumed 1-dimensional exponential kernel with variable correlation length b. In matrix methods, the known solutions of the fixed b equation for a representative value b^0, are used to expand the variable b solutions and hence convert the integral equation to a matrix eigenvalue equation. It is shown that calculation of the matrix is the computational bottleneck and two approximations are introduced that speeds this up by two orders of magnitude. One of these is to use a piecewise constant kernel; this leads also to a non-matrix method where the eigenfunction itself is approximated as a piecewise function. The performance of these approximations is investigated in detail by applying them to a model problem and it is found that both the best speed and accuracy is achieved by a method that uses the fact that the correlation function is strongly localised. It is shown that this approximation, called the diagonal correlation length matrix method, gives virtually identical eigenvalues and eigenfunctions to an exact calculation of matrix elements. Its use to give a basis for expanding stochastic quantities is illustrated by showing an expansion of the covariance function in terms of the variable b eigenfunctions.

1. INTRODUCTION

In a previous article by Kulasiri and Verwoerd [1999], it was shown that a stochastic model of flow in a porous medium (e.g. in an aquifer) can be formulated in terms of a flow equation that is perturbed by a spatially distributed Wiener process. The tortuous trajectory of a fluid element is modelled as a response to random fluctuations in medium properties such as porosity and hydraulic conductivity that it encounters as it moves through the medium. Thus the medium properties can be considered to vary smoothly on a macroscopic scale, but to have random fluctuations over a microscopic scale superimposed on it, reflecting the granularity of the medium. The scale of this granularity appears in the mathematical description as a spatial correlation between the stochastic perturbations at nearby points; i.e. a correlation length b appears in the description as an important parameter describing medium properties.

The way in which the correlation length enters the stochastic differential equation (SDE), is contained in a Karhunen-Loeve expansion of the Wiener term [Ghanem, 1991]. In the Karhunen-Loeve expansion, independent Wiener processes at different points are replaced by a single process modulated by a function of position. It turns out that if this modulation is constructed from the eigenfunctions of some assumed covariance function q(x,y), the stochastic variation of the Wiener processes are indeed correlated as specified by q within a range determined by the correlation length that appears as a parameter in q.

By an eigenfunction of q, we mean a function $f_n(x)$ that satisfies the following integral equation:

$$\int_p^q q(x_1, x_2) f_n(x_2) dx_2 = \lambda_n f_n(x_1) \qquad (1)$$

Here, p and q are constant limits describing an interval of interest, e.g. the boundaries of the flow region. For each value of index n, a discrete function is to be found satisfying this equation for a corresponding real number λ_n, the eigenvalue.

It remains to choose a covariance function. In this article, we only consider a function of the form:

$$q(x_1, x_2) = e^{\frac{|x_1 - x_2|}{b(x_1, x_2)}} \qquad (2)$$

In terms of integral equation terminology, the covariance function plays the role of a kernel in equation (1). The solution of this equation for an exponential kernel of the form of equation (2) but with a constant value of b, is well known [Ghanem, 1991]. Reducing the integral equation to a differential equation yields a function of the form:

$$f_n(x) = A_n \cos(\omega_n \tfrac{x}{t}) + B_n \sin(\omega_n \tfrac{x}{t}) \qquad (3)$$

As equation (3) indicates, it is convenient to rescale all spatial variables, including b, to dimensionless variables by dividing by the interval length t = q-p.

The coefficients A_n and B_n are determined by boundary conditions at x = p and x = q that are dictated by the integral equation. This, in turn, only has non-trivial solutions for the discrete values ω_n that satisfy the equation:

$$\tan \omega = \frac{2b\omega}{b^2\omega^2 - 1} \qquad (4)$$

The eigenvalues of the integral equation are related to the ω_n by:

$$\lambda_n = \frac{2b}{b^2\omega_n^{\,2} + 1} \qquad (5)$$

In the present context, the outlined solution would describe a homogeneous medium. However, the purpose of this article is to extend that to the case that b is also a function of position, in order to apply the stochastic model to non-homogeneous media.

The λ_n and f_n are the quantities needed for expanding spatially varying Wiener process amplitudes, and as shown elsewhere [Ghanem, 1991] the covariance function itself can also be expanded in terms of them:

$$q(x_1, x_2) = \sum_n \lambda_n f_n(x_1) f_n(x_2) \qquad (6)$$

To set the scene for the various approximations to be discussed, it is useful to establish some plausible properties of the covariance function. Firstly, the Wiener process is perfectly self-correlated, i.e. q(x,y) = 1 for x = y (a condition relaxed for a general Wiener process), and the correlation must decrease as |x-y| increases. This means that q(x,y) is a peaked function along the line x = y in the X-Y plane. Equation (2) clearly exhibits this behaviour. Secondly, it may be assumed that the peak width is small compared to the dimensions of the interval of interest, i.e. the (scaled) b(x,y) << 1. This is because the correlation length is much smaller than the macroscopic dimensions of the aquifer. Finally, the fact that changes in b are to be used to model changes in the medium properties from one part of the aquifer to another, is plausibly represented by the assumption b(x,y)=b((x+y)/2), i.e since only nearby points are significantly correlated, the correlation length may be evaluated at the midpoint between them. These characterisations are summarised by the schematic plot of a plausible covariance function in Figure 1.

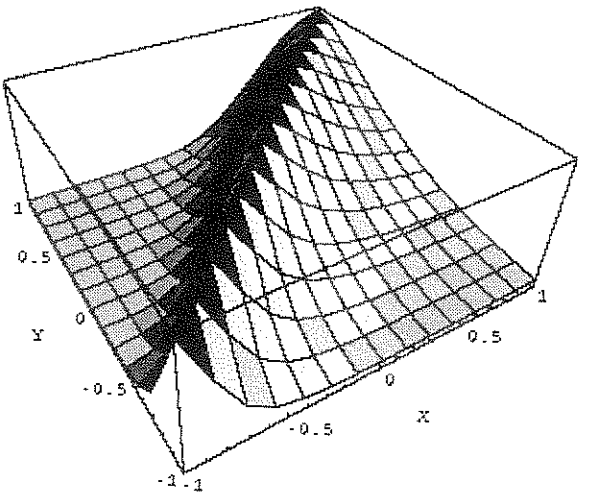

Figure 1. Schematic view of a variable correlation length covariance function; b increases in the direction away from the viewer.

Section 2 will be devoted to an outline of various proposed methods to solve the variable b eigenvalue problem. Section 3 compares the accuracy and computational effort when these are applied to a specific example of an assumed position dependence of b. Section 4 discusses the conclusions that can be drawn from the comparison.

2. APPROXIMATE SOLUTIONS

Equation (1) is classified as a Fredholm integral equation of the second kind [Morse and Feshbach, 1953]. The exact solution for constant b discussed above was obtained by applying the standard technique to reduce an equation of this kind to a differential equation. However, when b is variable, this does not deliver a differential equation that is easily solved, and moreover in

the applications envisaged b may only be known as a table of numerical values derived from measured media properties. Hence analytical methods are ruled out, and we resort to numerical solutions.

There are also well documented standard techniques for numerical solution of Fredholm equations of the second kind [Press et al, 1992]. Prominent among these is the Nystrom method, which uses Gauss-Legendre integration on the kernel integral to reduce the integral equation to a matrix eigenvalue problem of dimension equal to the number of integration points. To evaluate the method, it was applied to equation (1) for a fixed value b=0.2 for which the analytical solution is known. Comparing the eigenvalues found with the exact ones, improvements were found up to about 40 integration points, after which numerical inaccuracies set in. Eigenvalues could be obtained to within 10%, but the eigenfunctions are highly irregular and do not resemble the smooth exact functions given by equation (3). The reason for this failure is that the simple Nystrom method only works well for a smooth kernel. The exponential kernel however, is nearly singular - while it does remain finite, its derivative across the diagonal line x = y is discontinuous and it is highly localised around this line.

For the treatment of a kernel with a diagonal singularity, the Nystrom method is often extended by making use of the smoothness of the solution to subtract out the singularity [Press et al, 1992]. When applied to the present case, this is found to give some improvement for a low number of integration points but it is actually worse for more than about 12 points. The best accuracy obtained is no better than for the simple Nystrom method. More elaborate methods to deal with diagonal singularities have been used; for example, methods that construct purpose made integration grids to take the singular behaviour into account [Press et al, 1992]. However, we aim to construct a method which does not require a detailed prior knowledge of the kernel, and so these methods do not appear promising.

Moreover, if a specialised method is anyway required, a more direct approach is to make use of the known analytical solution for the fixed b case. This is supported by noting that the solutions in equations (3)-(5) do not, in fact, depend strongly on the value of b. That is illustrated by Figure 2, which shows the behaviour of the n=4 eigenfunction for 0.001<=b<=0.5, a variation over more than 2 orders of magnitude.

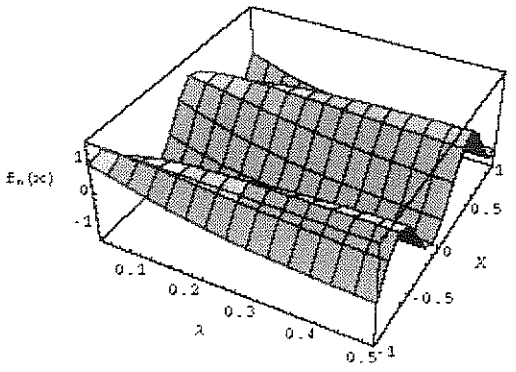

Figure 2. The n=4 eigenfunction of a fixed correlation length kernel, as the constant value b=λ, ranges from λ = 0.001 to λ = 0.5

An obvious way to exploit this observation, is to expand the eigenfunctions for variable b in terms of those calculated for some fixed typical correlation length b^0, e.g. the average value of b(x,y) over the integration interval:

$$f_n(x) = \sum_m F_{nm} f_m{}^0(x) \qquad (7)$$

When this is substituted into equation (1), the integral eigenvalue equation for the function q(x,y) is transformed to a matrix eigenvalue equation for the matrix Q defined by:

$$Q_{nm} = \int_p^q dx_1 \int_p^q dx_2 q(x_1, x_2) f^0{}_n(x_1) f^0{}_m(x_2) \qquad (8)$$

The dimension of the matrix is equal to the cutoff value M that has to be introduced as upper limit of the expansion over m in equation (7). Once the matrix has been diagonalised, the elements F_{nm} of its eigenvector matrix can be substituted back into equation (7) to get the first M of the desired eigenfunctions and its eigenvalues are identical to the first M eigenvalues of the integral equation.

The eigenfunctions of the kernel with a fixed correlation length b^0 can be shown to form a complete orthogonal basis. Therefore this method to solve the variable b case is exact up to the introduction of the finite cutoff M. Because the eigenfunctions are relatively insensitive to the value of b it is reasonable to expect a fast convergence of the expansion, so for practical purposes it should be possible to keep M fairly small. Nevertheless this solution is computationally intensive, not only because each of the M^2 elements of Q requires a multiple integral, but because the near singularity in q requires a large number of integration points for accurate numerical integration. The approximate

methods described below are intended to overcome this problem.

A key observation in this regard is that the double integration in equation (8) can be reduced to a single integral if b is a constant. By splitting the inner integral into two subranges the absolute value in the exponent in q can be eliminated, and in each subrange a factor $\exp(\pm x_1/b)$ can be factored out of the integral provided that b does not depend on x_2. The remaining integrand can be analytically integrated because of the simple form of the $f^{(0)}_n$ as specified by equation (3), leaving only the outer integral to be done numerically.

A direct way to take advantage of this idea is to approximate $b(x_1,x_2)$ as piecewise constant. The behaviour of $q(x_1,x_2)$ limits significant contributions to the integral to the vicinity of the diagonal line $x_1 = x_2$. Thus in a subdivision of the region of integration into a grid of square blocks, the dominating contribution will come from those blocks strung along the diagonal. In each of these q is approximated by using a fixed value of b, e.g its value in the centre of the block. The matrix element integral is reduced to a sum of integrals over the diagonal blocks, in each of which a different constant value of b is used to reduce it to a one-dimensional integral. We refer to this as the piecewise kernel matrix (PKM) method.

Having decided to use a piecewise kernel, one can go a step further by also constructing piecewise eigenfunctions. In fact, in this framework it is plausible to do away with the matrix problem altogether. Since a formula for the eigenfunction corresponding to any one of the piecewise constant values of b is known, this solution may be used within the subinterval, and the complete eigenfunction constructed by linking up all the solutions across the subinterval boundaries. The comparison between this approach and the matrix approach is somewhat like that between a spline function interpolation and a Fourier expansion of a function. However, in the present context the eigenfunctions to be linked up are already largely determined and there are not enough free parameters available to ensure that the function and its derivative are continuous across the subinterval boundary (as is done by spline functions). In fact, a problem in applying piecewise eigenfunctions is to determine the relative amplitudes of the functions used in neighbouring subintervals. As the eigenvalue equation is independent of amplitude, the only guideline is the overall normalisation over the entire interval. In practice, the insensitivity of the eigenfunctions to b ensures that discontinuities remain insignificant if subintervals are chosen to allow only moderate change of b from one subinterval to the next.

The elimination of the need to calculate and diagonalise a matrix in the piecewise eigenfunction (PE) method, is a major conceptual simplification. However, in computational terms it is not so much simpler. If there are M subintervals, for each eigenfunction M sets of coefficients in each subinterval need to be kept, and that is similar to keeping coefficients for an expansion over M basis functions in a matrix method. Also, all subsequent manipulations with piecewise eigenfunctions require the complexity of breaking up operations into subintervals, while in the matrix method a single function valid over the whole interval is obtained even when it was calculated from a piecewise kernel.

Returning to the matrix methods, there is another way to obtain the benefits of a constant b in calculating the matrix element integral. We write:

$$b(x_1,x_2) \approx b(\tfrac{x_1+x_2}{2}) \approx b(x_1) \qquad (9)$$

i.e. the correlation length b is kept variable, but only its value on the diagonal is used, because the behaviour of q limits the effective region of integration to $x_1 \approx x_2$. Equation (9) is enough to allow the factorisation of the kernel that leads to one-dimensional matrix element integrals. This is described as the diagonal correlation length matrix (DCLM) method.

3. RESULTS

To evaluate the relative performance of the various methods, we take p = -1, q = 1 and:

$$b(x) = 0.1 + 0.05x \qquad (10)$$

Thus b varies by a factor of three over the interval, but is always small compared to the total interval size t=2. None of the methods make any use of the assumed functional dependence; only numerical values are used, either pointwise or per subinterval. Therefore the simple linear variation of b should be representative of any other smooth variation within a similar range of values.

As a first measure of performance, it is noted that equation (1) can be interpreted as an integral operator that when operating on an eigenfunction, gives back the same function apart from an amplitude factor which is the corresponding eigenvalue. Therefore a candidate eigenfunction can be judged by how closely it resembles the resultant function obtained from it by the action of the integral operator. Figure 3

shows the n=4 eigenfunction and its (rescaled) resultant, for the 3 approximations:

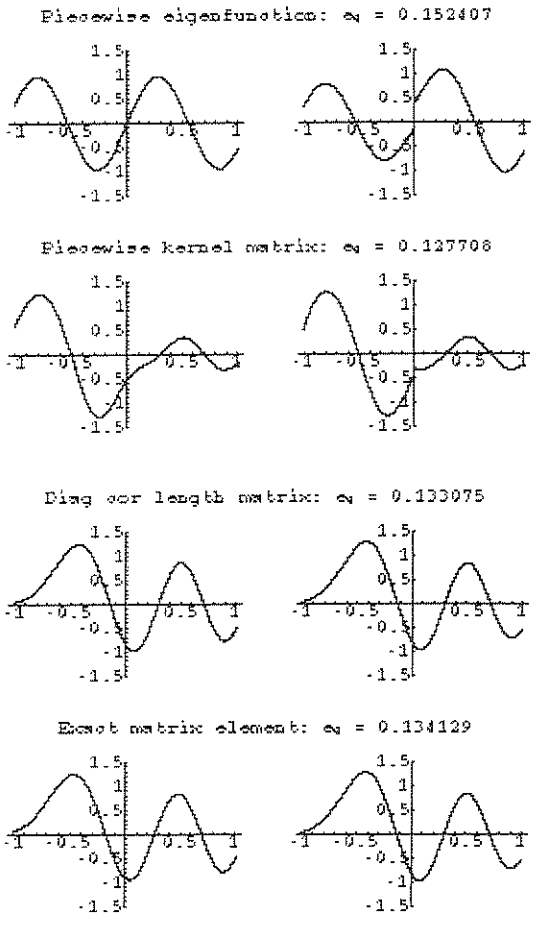

Figure 3. Eigenfunctions (left) and resultant (right) from action of kernel operator.

A very coarse piecewise grid consisting of only two subintervals (-1,0) and (0,1) was chosen for the piecewise approximations shown in the first two sets of graphs in figure 3, while a dimension of M=9 eigenfunctions was used for all the matrix methods. The calculations have also been repeated for a piecewise grid of 9 subintervals (which makes the complexity comparable to that of the 9x9 matrix methods) but despite the smaller step in kernel values from one subinterval to the next, the accuracy is worse because for the particular choice of parameters in equation (10) the assumption b<< subinterval is not satisfied for 9 subintervals.

The calculated eigenfunctions in Figure 3 all seem to satisfy the requirement that they should remain unchanged by the action of the kernel integral, reasonably well. Nevertheless the functions look quite different, apart from a qualitative agreement in the number of peaks, and also the eigenvalues differ substantially.

An obvious defect of the piecewise eigenfunction is that its amplitude in the two subintervals is the same whereas all the others show a plausible response to the different value of the kernel in the subintervals. The PE method might be improved if a rule can be found to choose relative amplitudes in the subintervals, but that would clearly introduce a discontinuity into the function.

The PKM method on the other hand appears to exaggerate the amplitude difference. By contrast, the DCLM approximation gives an eigenfunction that is virtually indistinguishable from the exact one in the last row of Figure 3, and also shows an excellent agreement of the eigenvalue.

A more quantitative measure of the performance of the eigenfunctions can be constructed by making use of the fact that the matrix elements of the kernel operator, defined as in equation (8) but using the calculated variable b eigenfunctions, should be the diagonal eigenvalue matrix. For each of the approximate methods, this 9x9 matrix was calculated and each element Q_{nm} scaled by dividing it by the harmonic mean of the diagonal elements at positions n and m. This produces a matrix with unit elements on the diagonal and off-diagonal elements should be 0 on a scale of 0 to 1. The largest scaled off-diagonal element is used as a measure of the overall eigenvector accuracy. Similarly, the eigenvalue accuracy is measured by the maximum difference from the exact value scaled to a fraction of the exact value. Unlike the visual comparison presented above, these measures cover the complete set of eigenfunctions and eigenvalues that were calculated. The result of the comparison is shown in Table 1, along with a comparison of the computing times for the methods.

Table 1. CPU time in seconds, and accuracy comparison, for various approximation methods.

Approx	CPU time	Eigenval dev, %	Eigenfun dev, %
Exact	4400	-	-
PE, 2 sub	27	21	89
PE, 9 sub	116	47	77
PKM, 2 sub	9	8	14
PKM, 9 sub	40	36	13
DCLM	20	3	3

It is seen that all the approximate methods are faster than an exact calculation by about two orders of magnitude, but only the DCLM method achieves this without a substantial loss of accuracy, particularly regarding the eigenfunctions. The "exact" calculation used as reference, uses matrix elements that are

calculated without approximation to within the numerical accuracy of the underlying quadrature routines, but is itself approximate to the extent that the matrix is truncated to 9x9. However, to confirm that this is not in practice a significant limitation, the DCLM calculation was repeated with the cutoff increased to a dimension of M=20, and agreement was found within the convergence limits of the quadrature routines for the first 9 eigenvalues and -functions.

The purpose for calculating the eigenfunctions in the context of this article is to be used to expand stochastic amplitudes. Therefore it is also of interest to see how they perform as basis functions for expanding a known function; the obvious candidate is the covariance function itself, for which the expansion is known to be given by equation (6). Figure 4 shows the result of that comparison.

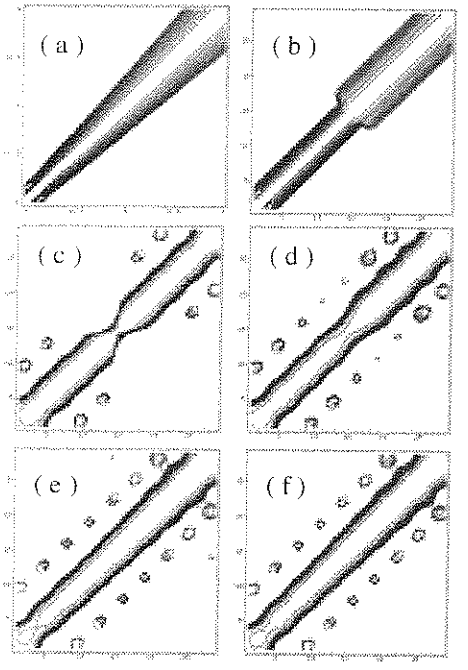

Figure 4: Contour plots of various representations of the exponential covariance function.

a) Exact kernel	b) stepwise kernel
c) 9-term PE expansion	d) 9-term PKM exp.
e) 9-term DCLM exp.	f) 9-term exact exp.

Again it is seen that the DCLM expansion gives the best representation of the covariance function, hardly distinguishable from that given by the exact eigenfunctions. Even the latter has some difficulty in representing the exponential rise in value near the diagonal because of the truncation of the expansion. However, it should be noted that for the sake of clarity, the contour spacing in the diagrams is logarithmic rather than equally spaced, so that the outlying contours showing the most deviation represent very small function values.

4. CONCLUSION

Of the various approximations investigated, the DCLM method gives the most accurate eigenfunctions and eigenvalues, virtually identical with the ones calculated from exact matrix elements but requiring less computing time by two orders of magnitude.

The actual calculations reported here, were done using the Mathematica® program [5] on a 150 MHz Pentium computer. The absolute timing values can no doubt be improved on by faster machines and using compiled code, but the relative values should remain similar.

It is concluded that the DCLM is a practical method for solving the eigenvalue problem for an exponential covariance kernel with a correlation length that is position dependent, and the application of this method to stochastic modelling of flow problems in porous media will be further investigated.

5. REFERENCES

Ghanem, R.G. and P.D. Spanos, Stochastic Finite Elements: a spectral approach, Springer Verlag, 1991.

Kulasiri, D and W.S. Verwoerd, A stochastic model for solute transport in porous media: mathematical basis and computational experiments, paper presented at MODSIM 99, Hamilton, New Zealand, 1999.

MATHEMATICA Version 3.0, Wolfram Research, Inc, 100 Trade Center Drive, Champaign, IL 61820, USA.

Morse, P.M and H. Feshbach, Methods of Theoretical Physics, Part I, McGraw-Hill, 1953.

Press, W.H., S.A. Teukolsky, W.T.Vetterling and B.F. Flannery, Numerical Recipes in C, Cambridge University Press, 1992.