Comparison between Karhunen–Loeve and wavelet expansions for simulation of Gaussian processes

K.K. Phoon *, H.W. Huang, S.T. Quek

Department of Civil Engineering, National University of Singapore, Singapore 117576, Singapore

Accepted 5 March 2004
Available online 9 April 2004

Abstract

The series representation consisting of eigenfunctions as the orthogonal basis is called the Karhunen–Loeve expansion. This paper demonstrates that the determination of eigensolutions using a wavelet-Galerkin scheme for Karhunen–Loeve expansion is computationally equivalent to using wavelet directly for stochastic expansion and simulating the correlated random coefficients using eigen decomposition. An alternate but longer wavelet expansion using Cholesky decomposition is shown to be of comparable accuracy. When simulation time dominates over initial overhead incurred by eigen or Cholesky decomposition, it is potentially more efficient to use a shorter truncated K–L expansion that only retains the most significant eigenmodes.

Keywords: Wavelets; Karhunen–Loeve; Eigen decomposition; Cholesky factorisation; Gaussian process

1. Introduction

For time dependent reliability assessment using simulation, one of the principal challenges is to develop efficient computer algorithms that can generate realistic sample functions satisfying incomplete probability information in a theoretically consistent way. Phoon et al. [1] emphasized that inexpensive implementation on a modest computing platform is a practical necessity because practitioners are unlikely to have access to computational resources beyond a desktop PC in routine design.

There are several ways to describe a stochastic process but a parametric form involving at-most countable random variables is probably most convenient for computer simulation and should be sufficiently general

for practical applications. The most widely known example in structural reliability literature relating to simulation is the spectral expansion [2,3]. The shape of the sample functions are specified by trigonometric functions. Other shape functions such as eigenfunctions [4], Legendre polynomials [5], and wavelets [6] have been used. While the shape functions are orthonormal, the random coefficients (variables) are not orthogonal in \( L^2 \) (i.e., they are correlated) for finite length processes unless eigenfunctions are used.

The series representation consisting of eigenfunctions as the orthogonal basis is called the Karhunen–Loeve (K–L) expansion. This expansion is known to produce the most economical representation among all orthogonal bases for the Gaussian case [6,7]. It has also been widely exploited by Schüeller et al. in the evaluation of stochastic responses of large finite element models [8]. However, it is not necessary that the most economical representation will be the most inexpensive, in computational terms. In other words, it may be more costly to obtain the coefficients of a short expansion than a long expansion. For simulation, repeated use of a short expansion may or may not offset this initial overhead.

*Corresponding author. Tel.: +65-68746783; fax: +65-67791635.
E-mail address: cvepkk@nus.edu.sg (K.K. Phoon).
Indeed, the K–L expansion is rarely used for simulation because the determination of the eigenfunctions requires the costly solution of a Fredholm integral equation, and this is frequently assumed to be the governing factor.

Recently, Phoon et al. [9] attempted to redress this disadvantage in K–L expansion by applying an efficient wavelet-Galerkin scheme for numerical solution of the Fredholm integral equation. It is natural to compare a wavelet-Galerkin K–L expansion with one based on direct application of wavelets as orthogonal bases. A theoretical comparative study is presented in this paper. The wavelet expansion is gaining interest in recent years [10]. Even for the Gaussian case, it is not obvious if a shorter K–L expansion with uncorrelated coefficients will necessarily out-perform the longer wavelet expansion with correlated coefficients. This practical computational aspect is studied in this paper using a stationary Gaussian process with an exponential covariance function as an illustrative example.

2. Theory

2.1. Karhunen–Loeve expansion

A random process \( \sigma(x, \theta) \) indexed on a bounded domain \( D \), having zero mean and finite variance, can be approximated using a finite Karhunen–Loeve series

\[
\sigma(x, \theta) = \sum_{k=1}^{M} \sqrt{\lambda_k} \xi_k(\theta)f_k(x)
\]

where \( \lambda_k \) and \( f_k(x) \) are the eigenvalues and eigenfunctions of the covariance function \( C(x_1, x_2) \). \( \xi_k(\theta) \) is a set of uncorrelated random variables, and \( M \) is the number of K–L terms. The deterministic eigenfunctions \( f_k(x) \) are obtained from the spectral decomposition of the covariance function via solution of the homogeneous Fredholm integral equation of the second kind

\[
\int_{D} C(x_1, x_2) f_k(x_2) \, dx_2 = \lambda_k f_k(x_2)
\]

Eq. (2) can be solved efficiently using a wavelet-Galerkin scheme by first approximating the eigenfunction \( f_k(x) \) as a truncated series of Haar wavelets [9]

\[
f_k(x) = \sum_{i=0}^{N-1} d^{(k)}_i \psi_i(x) = \Psi^T(x)D^{(k)}
\]

where \( N = 2^m \), \( m \) is the wavelet level, \( d^{(k)}_i \) are the wavelet coefficients, and

\[
\psi_0(x) = 1 \\
\psi_i(x) = \psi(2^i x - k) \\
i = 2^i + k; \quad k = 0, 1 \ldots, 2^i - 1; \quad j = 0, 1 \ldots, m - 1
\]

The extended set \( \Psi(x) \) forms a complete set of orthogonal functions over the domain \([0,1]\) and the orthogonality condition can be written as

\[
\int_0^1 \Psi \Psi^T \, dx = H
\]

where \( H \) is a diagonal matrix

\[
H = \begin{bmatrix} h_0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & h_{N-1} \end{bmatrix}
\]

and \( h_0 = 1, \ldots, h_i = 2^{i'}, \quad i = 2^i + k; \quad k = 0, 1 \ldots, 2^i - 1; \quad j = 0, 1 \ldots, m - 1 \). For convenience, we defined a normalized basis vector as

\[
\hat{\Psi} = H^{-\frac{1}{2}}\Psi
\]

It is easy to show that

\[
\int_0^1 \hat{\Psi} \hat{\Psi}^T \, dx = I
\]

Eq. (3) can be re-written as

\[
f_k(x) = \Psi^T(x)H^{-\frac{1}{2}}H^{\frac{1}{2}}D^{(k)} = \Psi^T(x)D^{(k)}
\]

Using the Galerkin method, the solution of Eq. (2) reduces to the determination of unknown vector \( D^{(k)} \) from the following finite-dimensional eigenvalue problem

\[
\lambda^{(k)}_k D^{(k)} = AD^{(k)}
\]

where \( A = H^2 \hat{A} H^2 \) and \( \hat{A} \) is the 2-D discrete wavelet transform (DWT) of \( C(x_1, x_2) \) sampled over a \( N \times N \) grid. Substituting Eq. (9) into (1), we have

\[
\sigma(x, \theta) = \sum_{k=1}^{M} \sqrt{\lambda_k} \xi_k(\theta)\Psi^T(x)D^{(k)}
\]

The same procedure applies for other Daubechies’ wavelets.

2.2. Wavelet expansion

If the same DWT is applied on an ensemble of the random process \( \sigma(x, \theta) \), each sampled at \( N \) points, we can write

\[
\sigma(x, \theta) = \Psi^T(x)G(\theta)
\]

where \( G(\theta) \) is a \( N \times 1 \) vector of random wavelet coefficients. It is easy to demonstrate that \( G(\theta) \) is a correlated random vector by observing that

\[
E[\sigma(x_1, \theta)\sigma(x_2, \theta)] = C(x_1, x_2)
\]

\[
= \Psi^T(x_1)E[G(\theta)G(\theta)]\Psi(x_2)
\]
and the covariance function is related to its 2-D wavelet transform $\tilde{A}$ by [11]

$$C(x_1, x_2) = \Psi^T(x_1) \tilde{A} \Psi(x_2) = \Psi^T(x_1) A \Psi(x_2)$$

(14)

The vector $G(\theta)$ is correlated because $A$ is usually not diagonal.

It is well known that correlated Gaussian vectors can be generated from eigen or Cholesky decomposition of the covariance matrix [12]. The former is more efficient for highly correlated vectors and is simply given as

$$m(x, \theta) = \Psi^T(x) \sum_{k=1}^{N} \sqrt{\lambda_k} \xi_k(\theta) D^{(k)}$$

(15)

Eq. (15) can be truncated to the first $M$ terms by ignoring higher order eigenmodes. In addition, it is more efficient to apply the inverse discrete wavelet transform on $D$ within the summation, because it involves only $N$ transforms. In contrast, the number of inverse transforms is equal to the number of realisations in Eq. (15). If the above two aspects are considered, Eqs. (11) and (15) are identical. Hence, the determination of eigen-solutions using a wavelet-Galerkin scheme is equivalent to using wavelet directly for stochastic expansion and simulating the correlated random coefficients using eigen decomposition. This argument applies for other orthogonal bases (Legendre polynomials, trigonometric functions, etc.,).

It is also possible to simulate a correlated random vector as follows:

$$G(\theta) = Q \begin{bmatrix} \xi_1(\theta) \\ \xi_2(\theta) \\ \vdots \\ \xi_N(\theta) \end{bmatrix}^T$$

(16)

where $Q$ is a lower triangular matrix obtained from Cholesky decomposition of $A$, i.e., $A = QQ^T$. Substituting Eq. (16) into (12), we have

$$m(x, \theta) = \Psi^T(x) \sum_{k=1}^{N} \xi_k(\theta) Q^{(k)}$$

(17)

where $Q^{(k)}$ are the column vectors in $Q$. Unlike eigen decomposition, it is not possible to obtain a more parsimonious expansion in a straightforward way by simply keeping the first $M$ terms in the general case. Although wavelet transforms are known to produce sparse representations, the significant non-zero terms are generally scattered across all levels, rather than conveniently consolidated within a group of contiguous levels [13]. Hence, additional housekeeping expenses are required to keep track of both magnitudes and positions of non-zero coefficients to achieve truncation. On the other hand, Cholesky factorization only requires $N^3/3$ operations while the Householder method (an intermediate tri-diagonalisation step in eigensolution) already requires $2N^3/3$ operations [13]. For efficiency, the inverse discrete wavelet transform is applied on $Q^{(k)}$ within the summation as discussed above.

3. Numerical results

This section presents the accuracy and runtime results for the simulation procedures described above. A zero-mean unit-variance stationary Gaussian process defined over the interval $[-a, a]$ with the following exponential covariance function is used for illustration

$$C(x_i, x_j) = e^{-j|x_i-x_j|/b}$$

(18)

where $b$ is the correlation parameter. Runtime required to compute realisations of this process is recorded using a standard Intel Pentium III, 450 MHz machine. Differences between the target and simulated covariance function are measured using the eigenvalues, $\lambda_k$. The relative error for each $\lambda_k$ is defined as $|\lambda_{\text{target}} - \lambda_{\text{simulated}}|/|\lambda_{\text{target}}|$, where $\lambda_{\text{target}}$ are eigenvalues determined numerically by sampling the target covariance function at a dense grid of $N = 1024$ points.

3.1. Accuracy

Fig. 1 shows the relative eigenvalue errors versus number of expansion terms ($M$) for wavelet-based K–L expansion (Eq. (15)) with $M = N$, where $N$ is the number of wavelet basis functions. Note that $a = 5$ and $b = 1$ in this example. The relative error behaves erratically after $M = 2^5$ terms for 10,000 realisations and more importantly, degrades significantly from the accuracy theoretically achievable as $M$ increases (Fig. 1a). The “theoretical” values are determined from Eq. (10) without involving simulation. It is possible to approach the minimum theoretical error limit by increasing the sample size as shown in Fig. 1b, thus identifying the source as statistical in nature.

For the large sample sizes shown in Fig. 1, the most likely reason for the above degradation in eigenvalue accuracy is the presence of minor but non-zero statistical correlations between the simulated standard Gaussian variates $\xi_k(\theta)$ appearing in the K–L expansion. In this study, standard uniform realisations are produced by the pseudo-random number generator ran1 from numerical recipes [13]. The routine ran1 is based on a simple multiplicative congruential algorithm with shuffling to remove low-order serial correlations. Realisations of the standard Gaussian variates, $\xi_k(\theta)$, are computed from ran1 using the standard Box-Muller transform and stored in a $n \times M$ matrix $X$ where the $k$th column contains $n$ realisations of $\xi_k(\theta)$. Statistical covariances between $\xi_k(\theta)$ can be computed as

$$C = \frac{X^T X}{n} - \frac{X^T U U^T X}{n^2}$$

(19)
where $C$ is the \( M \times M \) covariance matrix and $U$ is an \( n \times 1 \) vector containing ones. The correlation matrix $R$ is obtained by normalizing each entry in $C$ as follows:

$$r_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}} \quad (20)$$

It is possible to verify that the magnitudes of these correlations are consistent with the null hypothesis of zero correlation between $\zeta_k(\theta)$ by using the well-known $t$ statistics [13]

$$t = r \sqrt{\frac{n-2}{1-r^2}} \quad (21)$$

where $r$ is any off-diagonal entry in $R$ and $t$ follows the student’s $t$-distribution with $n-2$ degrees of freedom. Table 1 shows that the number of non-zero off-diagonal entries in $R$ exceeding the critical $r$ value from a two-sided test with 5% level of significance converges to 5% as $M$ increases. For large $n$, the critical $r$ value for this test can be computed from Eq. (21) as

$$r_{\text{crit}} = \sqrt{\frac{1.96^2}{n-2 + 1.96^2}} \quad (22)$$

Hence, the small correlations produced by ran1 and Box–Muller transform are not anomalous from the statistical sense and one must conclude that eigenvalues are very sensitive to this statistical departure from the ideal correlation matrix for uncorrelated random variables (identity matrix).

The practical significance of the above discussion is that it is necessary to increase both sample size ($n$) and number of expansion terms ($M$) to achieve improvement in the accuracy of eigenvalues. However, sample size is typically governed by the event probability in the problem of interest. It would be uneconomical to select a larger sample size to suit a longer expansion. From this viewpoint, it may be worthwhile to remove statistical correlations by performing the following pre-processing steps on the realisations for $\zeta_k(\theta)$:

1. Compute the covariance of $X$ using Eq. (19).
2. Perform Cholesky decomposition on $C$ to obtain the Cholesky lower triangular factor $Q$ using \texttt{cholsl} from numerical recipes [13].
3. Obtain a new $n \times M$ matrix of realisations $X' = X(Q^T)^{-1}$ by orthogonalization.

It can be proven that the correlation matrix for $X'$ would be exactly identity. The effect of orthogonalization is very significant as shown in Fig. 2. The relative eigenvalue errors are almost identical to the minimum theoretical limit, reaffirming the sensitivity of eigenvalues to minor statistical correlations arising from finite sample size. Similar eigenvalue accuracy also can be achieved using only 1000 realisations if the orthogonalization pre-processing step is used. This procedure for removing undesirable coupling between sample size and number of expansion terms can be applied to the alternate wavelet-Cholesky simulation approach (Eq. (17)) as well.

### 3.2. Runtime

There are two basic components in the simulation runtime. One is a fixed overhead consisting of eigensolution or Cholesky decomposition. It has been noted in Section 2 that the latter operation is more time efficient. Table 2 shows that this overhead difference can be considerable when $M$ is large. The orthogonalization overhead has been ignored because it is applicable to both expansions. The second cost component is variable and depends on the number of realisations ($n$) required for the problem of interest. In civil engineering prob-
problems, it is not uncommon to perform simulation studies using 10,000 realisations. Table 2 shows that the difference in total runtime is modest because most of the time is expended in the repetitive evaluation of the expansion, rather than in the one-off computation of coefficients using eigen or Cholesky decomposition.

The practical significance of this simple comparison is that more attention should be focused on the variable cost component, rather than on the relative efficiency of evaluating the coefficients in different expansions, when the number of realisations required is sufficiently large. Relatively minor computational details can exert a disproportionate influence on the total runtime. An example discussed in Section 2 is the application of the inverse discrete wavelet transform within (Eq. (11)) or without (Eq. (15)) the summation. Both are theoretically equivalent, but the former formulation requires only $N$ transforms ($N = 512$ in this example), while the latter formulation requires $n$ transforms ($n = 10,000$ in this example). Table 2 highlights a time savings factor of more than 3 when this computational detail is properly implemented.

Another observation is that the ordinal nature of eigenvalues permits the K–L expansion to be easily truncated in a systematic way. Fig. 3 shows that the number of significant eigenmodes is relatively small for various normalized length of the process. As to be expected, the rate of decay is more rapid for a ''short'' or strongly correlated process ($a = 5$, $b = 1$). It is possible to take advantage of this to reduce the variable cost component by calculating the coefficients accurately using a dense $N/C^2$ wavelet grid ($N = 512$ in this example) for eigensolution and retaining only the significant modes in the K–L expansion. Note that the sample paths simulated using a truncated K–L expansion are still densely defined at $N$ points because the eigenvectors $D_{(k)}$ in Eq. (11) are $N/C^2$ vectors. If the orthogonalization step is performed on the simulated $n_{(h)}$ prior to computation of the truncated K–L expansion, the relative eigenvalue errors associated with the simulated covariance function for the first $M$ modes would be essentially identical to the corresponding errors achievable using a K–L expansion or Cholesky expansion with $M = N = 512$.

Table 3 shows a runtime comparison between the truncated K–L expansion with $M = N$ and K–L expansion.
with $M = N$ and Cholesky expansion with $M = N$. For the truncated expansion, $M$ is selected to satisfy $\sum_{i=1}^{M} \lambda_i / \sum_{i=1}^{N} \lambda_i > 0.95$. The significant time savings achievable by simply computing fewer expansion terms are perhaps not well emphasized in the literature. In this numerical example, the Cholesky expansion is only competitive in terms of total runtime for “long” or weakly correlated processes (say $a/b = 100$) when the truncated K–L expansion is only marginally shorter than its full length. Needless to say, the inexpensive overhead incurred by Cholesky decomposition can result in lower total runtime, regardless of truncation, when the number of realisations required is small. In the example shown in Table 3, this happens for $n = 1000$ realisations, which is below the usual range of practical interest for $n$.

It may be possible to use less than $N$ terms for Cholesky expansion if the sparsity structure produced by wavelet transform is exploited. It has been mentioned previously that the significant non-zero terms are gener-

Table 2
Comparison of runtime (s) required for simulating 10,000 realisations using K–L expansion (Eq. (15)) and Cholesky expansion (Eq. (17)) with $M = N$

<table>
<thead>
<tr>
<th>$M = N$</th>
<th>K–L</th>
<th>Cholesky</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overhead</td>
<td>Total</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>128</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>256</td>
<td>4</td>
<td>44</td>
</tr>
<tr>
<td>512</td>
<td>39</td>
<td>252</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$M = N$</th>
<th>K–L</th>
<th>Cholesky</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overhead</td>
<td>Total</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>128</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>256</td>
<td>4</td>
<td>44</td>
</tr>
<tr>
<td>512</td>
<td>39</td>
<td>252</td>
</tr>
</tbody>
</table>

Table 3
Comparison of runtime (s) for generating different normalized length of processes using the truncated K–L expansion with $M \ll N$, K–L expansion with $M = N$ and Cholesky expansion with $M = N$ ($N = 512$ for all cases)

<table>
<thead>
<tr>
<th>$a/b$</th>
<th>Truncated K–L $M \ll N$</th>
<th>K–L $M = N$</th>
<th>Cholesky $M = N$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M^a$</td>
<td>$M/N$</td>
<td>Overhead</td>
</tr>
<tr>
<td>$n = 10,000$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>41</td>
<td>0.08</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>0.16</td>
<td>39</td>
</tr>
<tr>
<td>20</td>
<td>151</td>
<td>0.29</td>
<td>39</td>
</tr>
<tr>
<td>50</td>
<td>291</td>
<td>0.57</td>
<td>39</td>
</tr>
<tr>
<td>100</td>
<td>386</td>
<td>0.75</td>
<td>39</td>
</tr>
</tbody>
</table>

| $n = 1000$ |
| 5     | 41 | 0.08 | 39 | 42 | 39 | 62 | 3 | 24 |
| 10    | 80 | 0.16 | 39 | 44 | 39 | 62 | 3 | 24 |
| 20    | 151 | 0.29 | 39 | 47 | 39 | 62 | 3 | 24 |
| 50    | 291 | 0.57 | 39 | 56 | 39 | 62 | 3 | 24 |
| 100   | 386 | 0.75 | 39 | 62 | 39 | 62 | 3 | 24 |

$^aM$ is truncated to satisfy $\sum_{i=1}^{M} \lambda_i / \sum_{i=1}^{N} \lambda_i > 0.95$. 

Fig. 3. Eigenvalues of exponential covariance function evaluated using $N = 512$ wavelet basis functions for various normalized length of the process ($a/b$).
ally scattered across all wavelet levels and it is necessary to keep track of both magnitudes and positions of these non-zero coefficients, potentially leading to onerous housekeeping expenses. Note that the sparsity structure (which affects housekeeping expenses involved in truncation) is related to the type of wavelet basis function used (which requires different computational efforts for DWT) [9]. There is an interesting optimization possibility here that is currently under study. The practical computational issue posed in this paper clearly encompasses broader aspects beyond the relative efficiency of evaluating the coefficients in different expansions, which is not always the governing factor as demonstrated in this paper.

4. Conclusions

This paper demonstrates that the determination of eigensolutions using a wavelet-Galerkin scheme for Karhunen–Loeve expansion is computationally equivalent to using wavelet directly for stochastic expansion and simulating the correlated random coefficients using eigen decomposition. An alternate but longer wavelet expansion using Cholesky decomposition is shown to be of comparable accuracy. When simulation time dominates over initial overhead incurred by eigen or Cholesky decomposition, it is potentially more efficient to use a shorter truncated K–L expansion that only retains the most significant eigenmodes.

References