The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0264-4401.htm

A novel joint diagonalization approach for linear stochastic systems and reliability analysis

Feng Wang, Chenfeng Li and Jianwen Feng
College of Engineering, Swansea University, Swansea, UK
Song Cen
Department of Engineering Mechanics, Tsinghua University, Beijing, China and Key laboratory of Applied Mechanics, School of Aerospace, Tsinghua University, Beijing, China, and
D.R.J. Owen
College of Engineering, Swansea University, Swansea, UK

Received 12 May 2011
Revised 30 May 2011
Accepted 9 June 2011

Abstract

Purpose – The purpose of this paper is to present a novel gradient-based iterative algorithm for the joint diagonalization of a set of real symmetric matrices. The approximate joint diagonalization of a set of matrices is an important tool for solving stochastic linear equations. As an application, reliability analysis of structures by using the stochastic finite element analysis based on the joint diagonalization approach is also introduced in this paper, and it provides useful references to practical engineers.

Design/methodology/approach – By starting with a least squares (LS) criterion, the authors obtain a classical nonlinear cost-function and transfer the joint diagonalization problem into a least squares like minimization problem. A gradient method for minimizing such a cost function is derived and tested against other techniques in engineering applications.

Findings – A novel approach is presented for joint diagonalization for a set of real symmetric matrices. The new algorithm works on the numerical gradient base, and solves the problem with iterations. Demonstrated by examples, the new algorithm shows the merits of simplicity, effectiveness, and computational efficiency.

Originality/value – A novel algorithm for joint diagonalization of real symmetric matrices is presented in this paper. The new algorithm is based on the least squares criterion, and it iteratively searches for the optimal transformation matrix based on the gradient of the cost function, which can be computed in a closed form. Numerical examples show that the new algorithm is efficient and robust. The new algorithm is applied in conjunction with stochastic finite element methods, and very promising results are observed which match very well with the Monte Carlo method, but with higher computational efficiency. The new method is also tested in the context of structural reliability analysis. The reliability index obtained with the joint diagonalization approach is compared with the conventional Hasofer Lind algorithm, and again good agreement is achieved.

Keywords Stochastic processes, Programming and algorithm theory, Mathematics, Least squares, Real symmetric matrices, Steepest descent method, Conjugate gradient method, Reliability analysis

Paper type Research paper

The authors would like to thank the Royal Academy of Engineering for support from the Research Exchanges with China and Indian Award; the National Natural Science Foundation of China (Grant nos 10872108 and 10876100); the Program for New Century Excellent Talents in University (Grant no. NCET-07-0477) and the National Basic Research Program of China (Grant no. 2010CB832701).

Engineering Computations: International Journal for Computer-Aided Engineering and Software
Vol. 29 No. 2, 2012
pp. 221-244

0264-4401
DOI 10.1108/02644401211206061

1. Introduction

In stochastic finite element analysis, without loss of generality, the global stochastic linear equation of the system can be expressed as follows:

$$(\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_m A_m)x = b \tag{1}$$

in which $\alpha_i (i = 1, \ldots, m)$ are random variables, $A_i (i = 1, \ldots, m)$ are real symmetry deterministic matrices (e.g. stiffness matrix), b a deterministic/random vector denoting the load, and x is the unknown response vector. The solution of the above stochastic linear system has been extensively studied in the past two decades and many methods have been proposed, which includes the Monte Carlo method, the perturbation method, the Neumann expansion method, the polynomial chaos/generalized polynomial chaos method, and very recently the joint diagonalization method. In this paper, a novel gradient based joint diagonalization method is presented and a number of examples are set up to examine the performance of this new algorithm.

The results of stochastic finite element analysis are often presented in the form of the first- and second-order moments of the response (i.e. mean and standard derivation) (Sudret and Der Kiureghian, 2000, 2002; Sudret *et al.*, 2003; Xu and Rahman, 2005; Evans, 1992; Thoft-Christensen and Baker, 1982; Huh *et al.*, 2003; Neville and Kennedy, 1964; Melchers, 1999; Nakagiri, 1989; Bayer *et al.*, 2007). However, as demonstrated in Sudret and Der Kiureghian (2000, 2002) and Sudret *et al.* (2003) an important application of the stochastic finite element method is structural reliability analysis, where the satisfaction of the response depends on the design requirements. These include safety tests of the structure against collapse, damage, deflection, etc. and each such requirement may be termed as a limit state. In this paper, reliability analysis for structures will also be investigated.

2. Literature review

The Monte Carlo method

Monte Carlo methods are often used in simulating physical and mathematical systems (Fishman, 1995). Because of their reliance on repeated computation of random or pseudo-random numbers, these methods are most suited to calculation by a computer and tend to be used when it is unfeasible or impossible to compute an exact result with a deterministic algorithm. Monte Carlo simulation techniques involve "sampling" at "random" to simulate artificially a large number of random variables and to observe the result. In the case of analysis for stochastic finite element linear equations, this means, in the simplest approach, sampling each random variable $a_i (i = 1, \ldots, m)$ randomly to give N sets of sample values, for each set, we have:

$$K_i X_i = b \quad (i = 1, \ldots, N) \tag{2}$$

where K_i is a deterministic matrix according to the ith set of samples. Finally, the N sets of $X_i (i = 1, \ldots, N)$ can be employed for calculating the statistical properties of responses, e.g. the expectation and covariance of responses, and even the joint distribution quantities. Finally, practical rules can be established for giving the necessary number of trials. The more trials, the more accurate the results. Due to the reliability of Monte Carlo methods, they are always referred to as a reference when comparing with other methods.

The perturbation method
Perturbation methods comprise mathematical methods that can be used to find an approximate solution to a problem which cannot be solved exactly, by starting from the exact solution of a related problem (Cropper, 2004). It is applicable if the problem at hand can be formulated by adding a "small" term to the mathematical description of the exactly solvable problem.

The Neumann expansion method
To date, research results show that the Monte Carlo is not suitable for large complex structural analysis. Later in the 1980s, Shinozuka and Yamazaki coupled the Neumann expansion with the Monte Carlo simulation method to produce an efficient algorithm (Ghanem and Spanos, 2003). When the variability of random variables is greater than 0.2, the first-order perturbation methods cannot satisfy the precision requirements, but the Neumann method can also provide good results under this condition.

The polynomial chaos expansion method
The Neumann expansion methods in the previous section may be considered as a useful modification of existing techniques (Sachdeva *et al.*, 2006). However, given the range of applications encountered in practice it is desirable to devise a formulation which is more versatile. The polynomial chaos expansion is a technique that uses a polynomial-based stochastic space to represent and propagate uncertainty in space. This concept was first introduced by Wiener (1938) as "Homogeneous chaos". The theory evolved into the Wiener-Askey polynomial chaos with the extension of the theory to the entire Askey scheme of orthogonal polynomials. It has been applied to numerous different fields of study including fluid dynamics and circuit simulation (Xiu, 2010). The stochastic space described by polynomial chaos expansion method must be limited to a more reasonable size to allow for the evaluation of the equations required. The truncation of the polynomial chaos expansion is influenced by two factors, i.e. the number of random variables present in the system and the order of the polynomial basis.

The generalized polynomial chaos method
Xiu and Karniadakis (Xiu, 2010) have in fact shown that for a large number of common probability laws, the corresponding families of polynomials are determined using the Askey scheme (Askey and Wilson, 1985).

The joint diagonalisation method
Recently, a novel method named "the joint diagonalisation" that was first proposed by Li *et al.* (2006a, 2009) has been attracting more and more attention. In the next section, derivations about a new joint diagonalization strategy will be presented.

3. Proposed algorithm
Consider the following least-squares problem:

$$C(P, \Lambda_k) = \arg\min \sum_{k}^{N} \|A_k - P\Lambda_k P^T\|_F^2 \tag{3}$$

in which, Λ_k are real and diagonal. This is the cost function considered for joint diagonalization. Before proceeding, it is important to note some ambiguities existing

in the joint diagonalization. Without additional a priori information, matrix P can be at best identified up to permutation and scaling of its columns. Let Π, Δ and ε denote, respectively, a permutation matrix, a diagonal matrix whose entries are positive and a sign matrix (a diagonal matrix whose entries equal ± 1), the relation:

$$P\Lambda_{kl}P^T = (P\Delta\Pi\varepsilon)(\varepsilon\Pi^T\Delta^{-1}\Lambda_j\Delta^{-1}\Pi\varepsilon)(\varepsilon\Pi^T\Delta P^T) \tag{4}$$

shows clearly that P and Λ_k are not uniquely defined. Without loss of generality, in order to eliminate the ambiguity associated with the permutation factor Π, the columns of P are arranged in decreasing order according to their first elements; for the scaling factor Δ, we can put a norm constraint on the columns of P, say $\|p_i\| = 1$ to resolve the scaling problem; and for the sign factor ε, we can simply fix it as positive. Then, the cost function can be rewritten as follows:

$$C(P, \Lambda_k) = \sum_{i=1}^{N} \|A_k - P\Lambda_k P^T\|_F^2 \tag{5}$$

For each $P\Lambda_k P^T$, $k = 1, \ldots, N$, we have:

$$\begin{aligned} & P\Lambda_k P \\ &= p_1 p_1^T \lambda_1^k + p_2 p_2^T \lambda_2^k + \cdots + p_d p_d^T \lambda_d^k \\ &= \sum_{i=1}^{d} \left(p_i p_i^T\right)\lambda_i^k \end{aligned}$$

where λ_i^k is the ith element of the kth diagonal matrix Λ_k, and p_i the ith column of matrix P. Let $vec(\bullet)$ denote the column vector form of a matrix, the cost function can be rewritten as:

$$\begin{aligned} & \sum_{k=1}^{N} \left\| A_k - \sum_{i=1}^{d} \left(p_i p_i^T\right)\lambda_i^k \right\|_F^2 \\ &= \sum_{k=1}^{N} \left\| vec(A_k) - \sum_{i=1}^{d} vec\left(p_i p_i^T\right)\lambda_i^k \right\|_F^2 \\ &= \|\underline{A} - Q\underline{\Lambda}\|_F^2 \end{aligned} \tag{6}$$

in which:

$$\begin{aligned} \underline{A} &= [vec(A_1), vec(A_2), \ldots, vec(A_N)], \\ Q &= \left[vec\left(p_1 p_1^T\right), vec\left(p_2 p_2^T\right), \ldots, vec\left(p_d p_d^T\right)\right], \\ \underline{\Lambda} &= [diag(\Lambda_1), diag(\Lambda_2), \ldots, diag(\Lambda_N)]. \end{aligned}$$

This is recognized as a subspace fitting problem where the N columns of $\underline{A}$ are considered to span a subspace. It is clear that given the matrix Q, the cost function (6) is

minimized by $\underline{\Lambda} = Q^{\dagger}\underline{A}$ where $Q^{\dagger}$ is the pseudo-inverse (Moore-Penrose inverse) of Q. Thus, the minimization problem can be further simplified as:

$$\begin{aligned} C(\underline{A}) &= \|\underline{A} - QQ^{\dagger}\underline{A}\|_F^2 = \|(I - QQ^{\dagger})\underline{A}\|_F^2 \\ &= \|H\underline{A}\|_F^2 = f^T f \end{aligned} \tag{7}$$

where $H = I - QQ^{\dagger}$ and $f = vec(H\underline{A})$.

Define the Jacobian matrix as:

$$F = \left[\frac{df}{dP_{11}}\frac{df}{dP_{12}}\cdots\frac{df}{dP_{ij}}\right] \tag{8}$$

in which P_{ij} are elements of the transformation matrix P.

The derivative of f with respect to P_{ij} is Golub and Pereyra (1973):

$$\begin{aligned} \frac{df}{dP_{ij}} &= vec\left(\frac{dH}{dP_{ij}}\underline{A}\right) \\ &= -vec\left(\left(H\frac{dQ}{dP_{ij}}Q^{\dagger} + \left(H\frac{dQ}{dP_{ij}}Q^{\dagger}\right)^T\right)\underline{A}\right). \end{aligned} \tag{9}$$

We propose to ignore the second term in equation (9), since the $H\underline{A}$ occurring in that term corresponds to the residual and is typical very small in the neighborhood of the optimum. Thus, the approximate expression of df/dP_{ij} is:

$$\frac{df}{dP_{ij}} = -vec\left(H\frac{dQ}{dP_{ij}}Q^{\dagger}\underline{A}\right) \tag{10}$$

Noting that $Q = [vec(p_1p_1^T), vec(p_2p_2^T), \ldots, vec(p_dp_d^T)]$, the Jacobian matrix F can be calculated straightforwardly. Specifically, the derivative matrix dQ/dP_{ij} has non-zero values only in the jth column $vec(p_jp_j^T)$; and within the jth column of Q, only those entries that are related to the ith row and ith column of the matrix $p_jp_j^T$ have non-zero values.

In most circumstances the properties of equation (7) make it worthwhile to use methods designed specifically for the least-squares problem. In particular, the gradient:

$$g = F^T f \tag{11}$$

and the Hessian matrix:

$$G = F^T F + R \tag{12}$$

both have special structures. Least-squares methods are typically based on the premise that eventually the first-order term F^TF of equation (12) will dominate the second-order term R (Gill *et al.*, 1981; Fletcher, 1987). This assumption is not justified when the residuals at the solution are very large, i.e. roughly speaking, when the residual $\|f^*\|$ is comparable to the largest Eigen-value of F^TF. For many problems, however, the residual at the solution is small enough to ignore the term R.

4. Numerical methods

Based on the gradient g and the Hessian matrix G, there are many numerical methods for the solution of the nonlinear least-squares problem (6), such as the steepest descent method, the Gauss-Newton method, and the conjugate gradient method, etc. In this section, the detailed formulation and the convergence properties of these algorithms will be discussed in the context of the joint diagonalization problem.

4.1 The steepest descent method

The gradient descent method can work in spaces of any number of dimensions, even in infinite-dimensional cases. To find a local minimum of a function, one takes steps proportional to the negative of the gradient of the function at the current point. In the proposed algorithm, the iterative formulation has the form as follows:

$$P_{(k+1)} = P_{(k)} + \eta_k r_k \quad r_k = -g_k \tag{13}$$

where η_k is the kth step size.

It is well known that the gradient descent method has problems with pathological functions such as the Rosenbrock (1960) function. The Rosenbrock function has a narrow curved valley which contains the minimum. The bottom of the valley is very flat. Because of the curved flat valley the optimization is "Zig-Zagging" slowly with small step size towards the minimum. For poorly conditioned convex problems, gradient descent increasingly "Zigzags" as the gradients point nearly orthogonally to the shortest direction to a minimum point (Shewchuk, 1994). However, the gradient decent method remains a popular choice for many mathematical applications due to its simplicity. A main advantage of this method is that it is guaranteed to find the minimum through iterations as long as it exists (Wang, 2008). A number of studies have focused on the convergence performance of the steepest descent method (Oldenburg *et al.*, 1993; Battiti, 1992).

The new algorithm with steepest descent directions is denoted as "DPSDM", and within each iteration, the operations are summarized below.

Summary of DPSDM method

Do

$$\begin{aligned} P &= P_k \\ g_k &= F^T f \\ P_{(k+1)} &= P_{(k)} + \eta_k r_k \end{aligned}$$

Until convergence

4.2 Gauss-Newton algorithm

The Gauss-Newton algorithm is a method usually used for solving nonlinear least-squares problems. It can be seen as a modification of Newton's method for finding a minimum of a function. Unlike Newton's method, the Gauss-Newton algorithm can only be used to minimize a sum of squared function values, but it has the advantage that second derivatives, which can be challenging to compute, are not required (Bjorck, 1996; Fletcher, 1987; Nocedal and Wright, 1999).

The iterative process with the new algorithm is:

$$P_{(k+1)} = P_{(k)} - \eta_k G_k^{-1} g_k. \tag{14}$$

The convergence rate of the Gauss-Newton algorithm can approach quadratic. The algorithm may converge slowly or not at all if the initial guess is far from the minimum or the matrix G_k is ill-conditioned.

When dealing with large-scale problems, two related difficulties may occur: the computational time required may become too long to justify solving the problem; and, more critically, there may not be enough computer memory, either in core or on auxiliary storage, to contain the matrix needed to compute the direction of search. The new algorithm with Gauss-Newton search directions is denoted as "DPGNA", and within each iteration, the operations are summarized as below.

Summary of DPGNA method

Do

$P = P_k$

$g_k = F^T f$

$G_k = F^T F$

$P_{(k+1)} = P_{(k)} - \eta_k G_k^{-1} g_k$

Until convergence

4.3 Conjugate gradient method

As discussed in Sections 4.1 and 4.2, the steepest descent method often finds itself taking steps in the same direction during early searching steps and the Gauss-Newton method generates directions of search with a huge demand of memory. As an alternative approach, conjugate gradient type methods compute search directions without storing a matrix and along each search direction, the method takes exactly one step. In mathematics, the conjugate gradient method is an algorithm for the numerical solution of particular systems of linear equations, such as quadratic functions. The conjugate gradient method can be used not only to find the minimum point of a quadratic form, but it also minimizes any continuous function f for which the gradient f' can be computed. Applications include a variety of optimization problems, such as engineering design, neural net training, and nonlinear regression. In nonlinear cases, the only significant modification is that the step size η_k must be computed by an iterative process rather than in a closed form.

The conjugate gradient method is formulated as:

$$P_{(k+1)} = P_{(k)} + \eta_k r_k \tag{15}$$

where the search direction for the next iteration is defined as:

$$r_{k+1} = -g_{k+1} + \beta_{k+1} r_k \tag{16}$$

There are two options for β in the nonlinear conjugate gradient method, i.e. the Fletcher-Reeves (Al-Baali, 1985; Al-Baali and Fletcher, 1996; Gilbert and Nocedal, 1992; Dai and Yuan, 1996, 1998; Fletcher and Powell, 1963; Fletcher and Reeves, 1964) formula and the Polak-Ribiere formula (Dai and Yuan, 1995; Grippo and Lucidi, 1997; Khoda *et al.*, 1992). The Fletcher-Reeves method converges if the starting point is

sufficiently close to the desired minimum, whereas the Polak-Ribiere method can, in rare cases, cycle infinitely without converging. However, Polak-Ribiere often converges much more quickly (Shewchuk, 1994). The new algorithm with the Polak-Ribiere conjugate gradient directions is denoted as "DPCGM".

In general, the nonlinear conjugate gradient method comes with fewer guarantees for convergence compared with the linear case. The less similar f is to a quadratic function, the more quickly the search directions lose conjugacy. Fortunately, preconditioning and restart schemes can be employed to help relieve difficulties in convergence (Shewchuk, 1994). The iterative process for DPCGM is summarized as follows.

Summary of the DPCGM method

$$r_0 = -g_0$$

Find η_k that minimizes the cost function along r_0:

$$P_{(k+1)} = P_{(k)} + \eta_k r_k$$

$$r_{k+1} = -g_{k+1} + \beta_{k+1} r_k$$

in which:

$$\beta^{FR}_{(k+1)} = \frac{g^T_{(k+1)} g_{(k+1)}}{g^T_{(k)} g_{(k)}}$$

or:

$$\beta^{PR}_{(k+1)} = \frac{g^T_{(k+1)} (g_{(k+1)} - g_{(k)})}{g^T_{(k)} g_{(k)}}$$

4.7 Step size control

The line search strategy first finds a descent direction along which the value of the objective function will be reduced and then computes a step size that decides how far the objective function should move along that direction. The step size can be determined either exactly or approximately. The exact line search (e.g. the explicit step size in Newton's method) is often valid only for quadratic models. Typically, an effective line search only looks towards $\eta_k > 0$ since a reasonable method should guarantee that the search direction is a descent direction. However, approximate line search schemes are very popular for most practical applications. In the unconstrained minimization problem, the Wolfe condition (Nocedal and Wright, 1999; Wolfe, 1969) defines a set of inequalities for performing approximate line search. The idea is to find:

$$\min f \tag{17}$$

for some smooth $f: R^n \rightarrow R$. Each step often involves approximately solving the sub problem:

$$\min_{\eta} f(P_k + \eta_k r_k) \tag{18}$$

in which P_k is the current best guess, $r_k \in R^n$ is a search direction, and $\eta_k \in R$ is the step length. Approximate line searches provide an efficient way of computing an

acceptable step length η that reduces the objective function "sufficiently", rather than minimizing the objective function over $\eta \in R^+$ exactly. A step length η_k is said to satisfy the Wolfe conditions if the following two inequalities hold:

(1) $f(P_k + \eta_k r_k) \leq f(P_k) + c_1 \eta_k r_k^T \nabla f(P_k)$

(2) $r_k^T \nabla f(P_k + \eta_k r_k) \geq c_2 r_k^T \nabla f(P_k)$

with $0 < c_1 < c_2 < 1$. The constant c_1 is usually chosen to be quite small while c_2 is much larger. Nocedal (1991, 1995, 1996) gives example values of $c_1 = 10^{-4}$ and $c_2 = 0.9$ for Newton or quasi-Newton methods and $c_2 = 0.1$ for the nonlinear conjugate gradient method. The inequality:

(1) is known as the Armijo rule, that ensures that the step length η_k decreases f; and

(2) as the curvature condition, that ensures that the slope has been reduced sufficiently.

5. Numerical simulation

The experiments in this section are intended to compare the DPSDM, DPGNA and DPCGM methods for simultaneous diagonalization. We present the results of three progressively more complex experiments. First, a "Sanity check" experiment is performed on a relatively easy set of perfectly diagonalizable matrices. This experiment is intended to demonstrate the convergence performance of these methods for small scale diagonalizable matrices. The second experiment is designed to investigate the impact of non-diagonalizability of the set of matrices on the performance of these three methods. In the third experiment, a stochastic numerical modeling of a composite plate will be presented. All the computations are performed on a PC platform with an Intel (R) Xeon (TM) CPU 2.4 GHZ and 3.5 GB of RAM.

5.1 "Sanity check" experiment

The test data in this experiment are generated as follows. We use $K = 15$ diagonal matrices Λ_i of size 5×5 where the elements on the diagonal are drawn from a uniform distribution in the range $[-1,\ldots,1]$. These matrices are "mixed" by an orthogonal matrix P according to $P\Lambda_i P^T$ to generate the set of target matrices $\{A_i\}$ to be diagonalized. The mathematical form is:

$$A_i = P\Lambda_i P^T \quad i = 1, \ldots, 15 \tag{19}$$

The DPSDM, DPGNA and DPCGM methods are initialized with the identity matrix $P_{(0)} = I$, and the diagonalization error is measured by the function $\sum_{i=1}^{15} off(A_i)$, in which the $off(A_i)$ is defined as:

$$off(A_i) = \sum_{i=1}^{5} \sum_{\substack{j=1 \\ j \neq i}}^{5} (A_i)_{ij}^2 \quad (i = 1, \ldots, 15)$$

The step size is controlled by the Wolfe conditions for all these three algorithms. The convergence of these three methods means that the cost function must be reduced to zero (Figure 1).

It is seen from the graph that both DPSDM and DPGNA can converge towards the required accuracy. But DPCGM failed to converge, and stopped at the third iteration due to the cost function not being a quadratic type. As discussed in Section 4.3, the less similar the cost function is to a quadratic function, the more quickly the search directions lose conjugacy. Much research work has been undertaken to exploit the convergence properties of the conjugate gradient method for non-quadratic problems, but until now there has not been a clear answer (Al-Baali and Fletcher, 1996; Al-Baali, 1985; Dai and Yuan, 1996, 1998; Gilbert and Nocedal, 1992). For the performance of DPSDM and DPGNA, the DPGNA converges to solution with only seven iterations, while DPSDM requires 16 iterations. Clearly, the DPGNA was faster than the DPSDM due to its Gauss-Newton search direction. The convergence rate of the Gauss-Newton algorithm can approach quadratic. However, it is noted that the DPGNA method may converge slowly or not at all if the initial guess is far from the minimum or the Hessian matrix is ill-conditioned. As stated in Section 4.1, due to the inefficient search direction, the DPSDM always converges to solutions slower than the other methods. The steepest descent direction often finds itself taking steps in the same direction during early search steps, because the so-called "steepest descent direction" represents only a local property of the cost function. Although the steepest descent direction is rarely used in practical applications, it can be used as the initial steps of other algorithms, since it always finds a descent direction at a fixed point.

5.2 Non-diagonalizable matrices

We now investigate the impact of non-diagonalizability of the set of matrices on the performance of DPSDM and DPGNA. The DPCGM will not be discussed here due to its limitations for non-quadratic type problems. Non-diagonalizability is modeled by adding a random non-diagonal symmetrical "noise" matrix to each of the input matrices:

$$A_i = P\Lambda_i P^T + \sigma^2 (R_i)(R_i)^T \quad i = 1, \ldots, m \tag{20}$$

in which the elements of R_i are drawn from a standard normal distribution. The parameter σ allows control of the impact of the non-diagonalizable component.

Figures 2 and 3, respectively, show the convergence plots of DPSDM and DPGNA for two values of σ. The experimental set up is the same as in Section 5.1, apart from the additive noise. The impact of the latter can be quantified by computing the $off(\cdot)$ function on the noise terms only. In the case of $\sigma = 0.02$, both DPSDM and DPGNA converged to a point that is very close to zero, while in the case of $\sigma = 0.1$ the solutions only converge to a level that is somewhat further from zero.

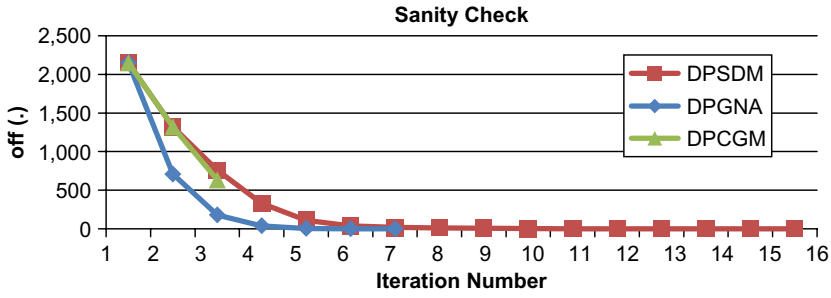

Figure 1.
Sanity check of DPSDM, DPGNA, DPCGM

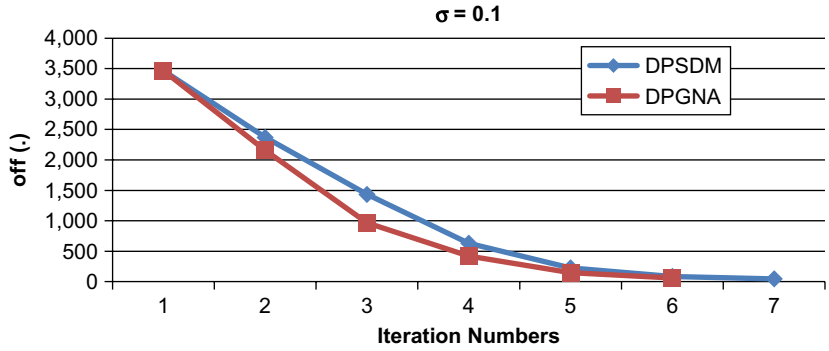

Figure 2.
Convergence behavior in the case of $\sigma = 0.1$

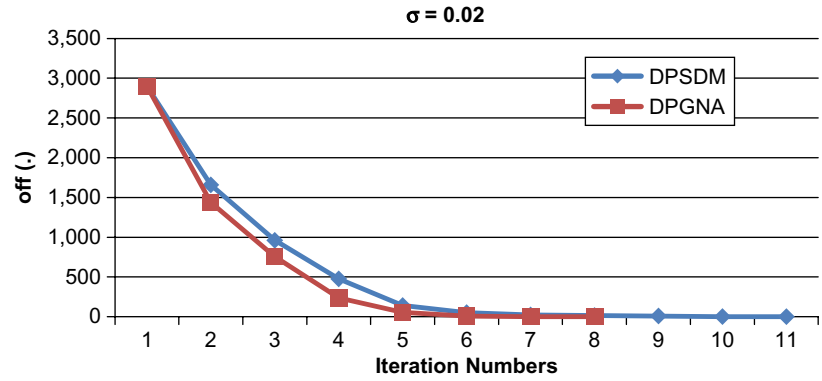

Figure 3.
Convergence behavior in the case of $\sigma = 0.02$

5.3 Composite material (plate) modeling (random Young's modulus)

In this section, a numerical modeling for elastostatic problems of composite material (plate) is presented to verify the joint diagonalization strategy, and the results will be compared with Monte Carlo solutions and Ansys modeling (for constant material properties).

Composite materials are engineered or naturally occurring materials made from two or more constituent materials with significant different physical or chemical properties which remain distinctly separated at the macroscopic or microscopic scale within the finished structure. This example considers an elastic composite plate under plane stress conditions with isotropic random Young's modulus. Figure 4 shows the configuration of this modeling.

A static loading of 10^8 N/m (tension) is applied on the top edge of the plate, and all the degrees of freedom (Ux, Uy) are fixed for the bottom nodes. The composite material is assumed to be isotropic and material properties including Poisson's ratio $v = 0.2$, and density $\rho = 1,500\ \text{kg/m}^3$ are assumed to be constant values. The Young's modulus $E = E(x, \theta)$ is assumed to be random and approximately modeled as independent stationary Gaussian stochastic fields (Note: the joint diagonalization strategy does not have any limitation on the type of probability distribution of random variables, which means that any probability distribution can be used). The mean value of Young's modulus and the corresponding covariance function are defined as follows:

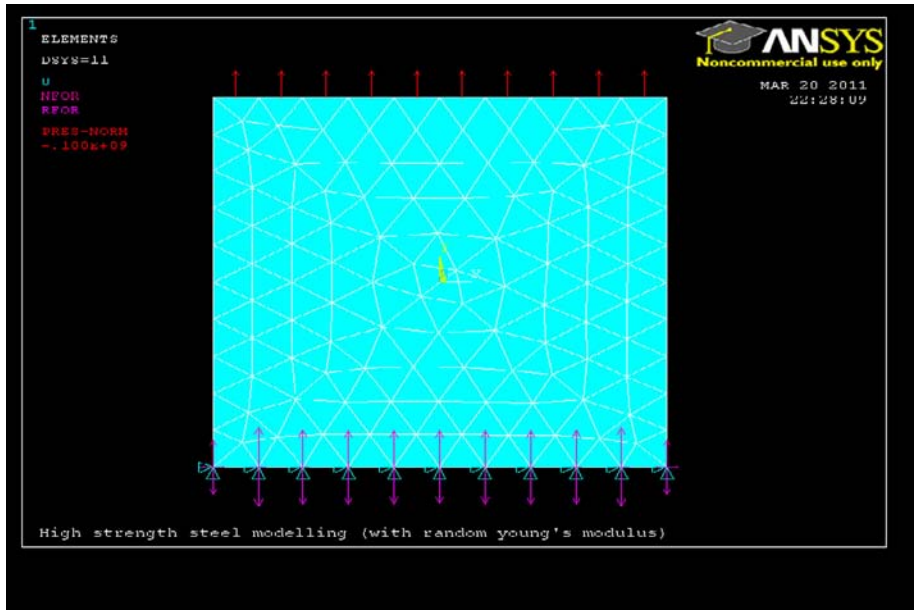

Figure 4.
The configurations of the modeling

$$E = E(x, \theta) = 205.9\,\text{Gpa} \quad \forall x \in D \tag{21}$$

$$\begin{aligned} &Cov(E(x_1, \omega), E(x_2, \omega)) \\ &= 1695.7924 e^{-(((x_2-x_1)^2+(y_2-y_1)^2)/0.6325^2)}\text{Gpa}^2 \end{aligned} \tag{22}$$

Following the F-K-L representation scheme and SFEM formulation (Li *et al.*, 2006b), a stochastic linear equation is obtained, which consists of $m = 9$ real symmetric matrices (264×264) with a set of mutually independent standard Gaussian random variables $a_i (i = 1, \ldots, 9)$, in which, the first variable a_1 always equals 1 because the first matrix is deterministic. Owing to the large demand of memory for calculating the Hessian matrix G in the DPGNA method, the joint diagonalization method (DPSDM) is selected to compute a total number of 2,000 sample solutions of the resulting stochastic linear system. In order to verify the accuracy of the joint diagonalization strategy, the samples above are compared with Monte Carlo solutions. As shown in Figure 5, the total value of the off-diagonal entries, i.e. $\sum_{i=1}^{m} off(K_i)$, is significantly reduced by the DPSDM method. From the resulting average transformation structure, the ratio of the off-diagonal entries to the Frobenius norm is $2.9E-2$, i.e.:

$$\sum_{i=1}^{m} off\left(K_i^*\right) : \sum_{i=1}^{m} \left\|K_i^*\right\|_F^2 = 2.9 \times 10^{-2} : 1 \tag{23}$$

At the same time, the Jacobi-like joint diagonalization strategy (Li *et al.*, 2006a) can only reduce this ratio to 9.68×10^{-2}, which induces large errors when compared with Monte Carlo solutions.

The random distribution of the two principal stresses (σ_1, σ_2) solved by the DPSDM method are shown in Figures 6 and 7, while the Monte Carlo solutions are shown in Figures 8 and 9 as reference solutions.

As shown in Figures 6-9, results calculated by the DPSDM method match very well with Monte Carlo solutions. The 1st principal stress σ_1 distribution as shown in Figure 6 varies over the plate from a minimum value of 0.095 Gpa to a maximum value of 0.126 Gpa, while the 1st principal stress σ_1 varies from a minimum of 0.096 Gpa to a maximum value of 0.127 Gpa for the Monte Carlo solution. In addition to values of principal stress σ_1, positions that the maximum and minimum values occur are almost the same in Monte Carlo solutions and solutions calculated from DPSDM. However, there are some minor differences in the distribution of σ_1 at the macro scale due to errors produced by the joint diagonalization strategy (Non-diagonalized part of matrices set K_i,

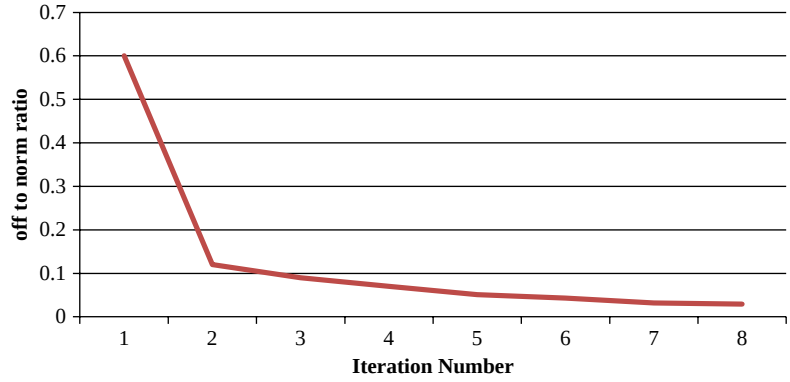

Figure 5.
Convergence history of the DPSDM method

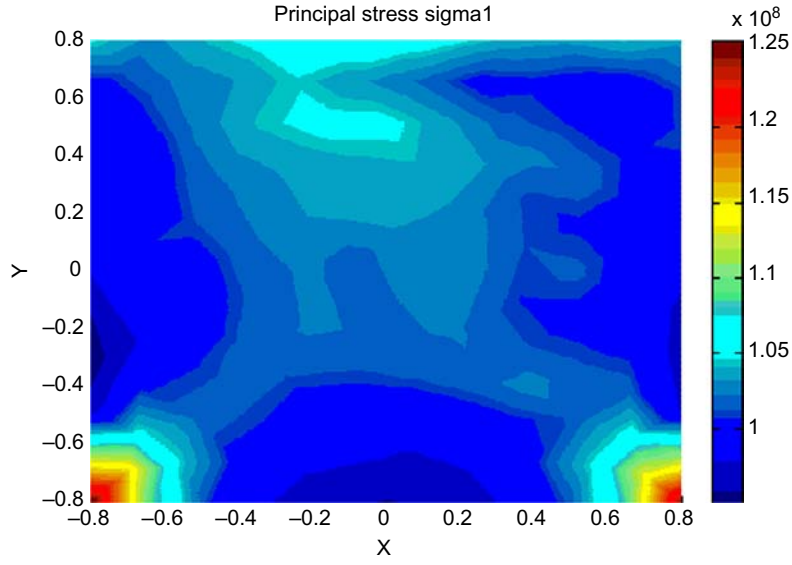

Figure 6.
Distribution of principal stress σ_1 produced by DPSDM

$i = 1, \ldots, m$). From Figures 7-9, the distribution of the 2nd principal stress σ_2 fluctuates from a minimum value of 0.2×10^7 pa at the top edge of the plate to a maximum value of 2.59×10^7 pa at the bottom of plate for both the DPSDM and Monte Carlo method.

Besides the comparison of results, the corresponding time costs of 2,000 solutions are recorded in Table I. It can be seen that the joint diagonalization solution strategy exhibits better performance than the Monte Carlo method in terms of efficiency.

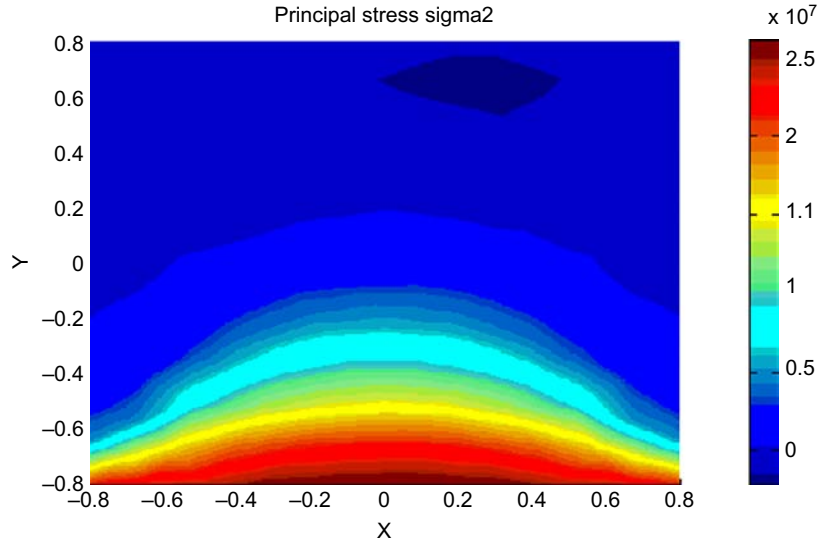

Figure 7.
Distribution of principal stress σ_2 produced by DPSDM

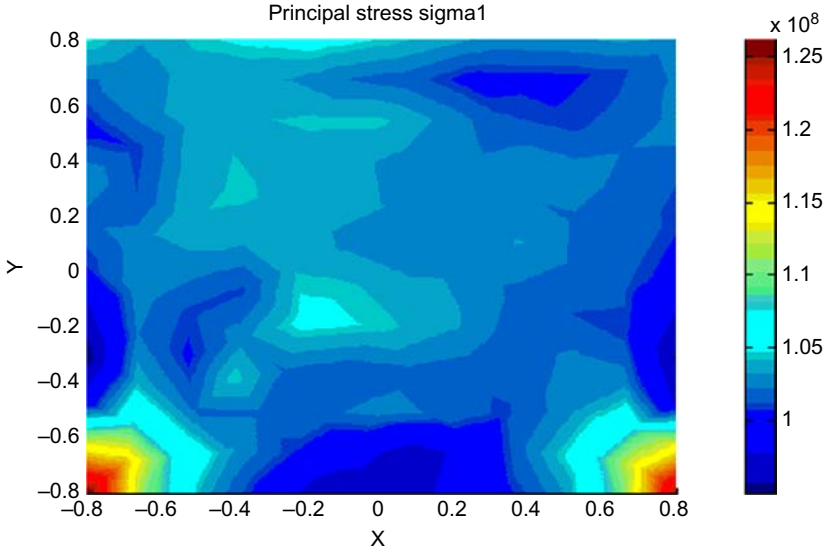

Figure 8.
Distribution of principal stress σ_1 produced by Monte Carlo method

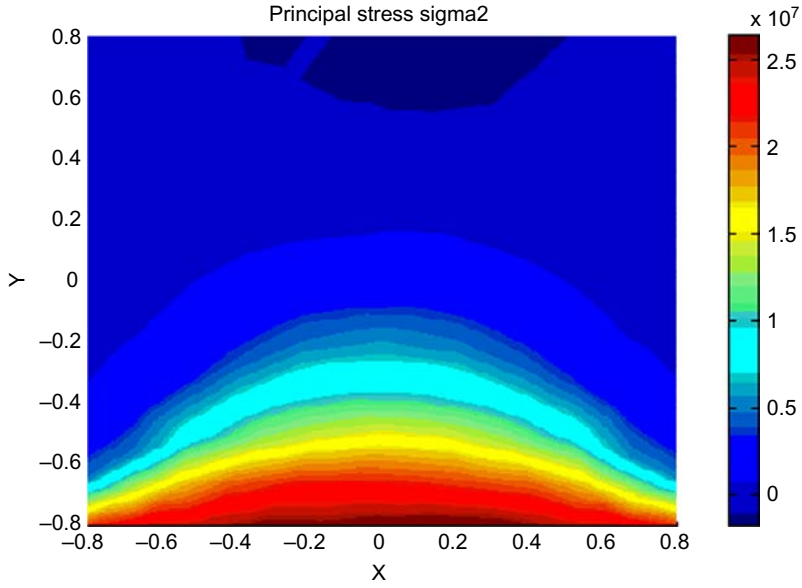

Figure 9.
Distribution of principal stress σ_2 produced by Monte Carlo method

	Monte Carlo solutions	DPSDM
Time (s)	570	460

Table I.
Time costs of 2,000 solutions

In addition to modeling with random material properties, a deterministic analysis with constant material properties is performed with Ansys (version 12.1). The model shares the same mesh structures with the modeling of the stochastic one. The distribution of the two principal stresses (σ_1, σ_2) are shown in Figures 10 and 11.

Figure 10 shows that the distribution of the 1st principal stress σ_1 varied from a minimum value of 0.951×10^8 pa at the bottom and two sides of plate to a maximum value of 1.21×10^8 at the bottom corners. As we can see from the figure, there are some differences in the maximum and minimum values between the deterministic model and the model with random material properties, but the overall distribution appears similar. The same phenomenon occurs when undertaking comparison for the second principal stress σ_2.

6. Application to reliability analysis

The principles of reliability analysis have been applied to a very large class of problems, ranging from the design of control systems for complex nuclear and chemical plants to the design of specific mechanical and structural components, as well as more generally in the field of electronics and aerospace. Most structural design is undertaken in accordance with codes of practice, which in many countries is regulated by law.

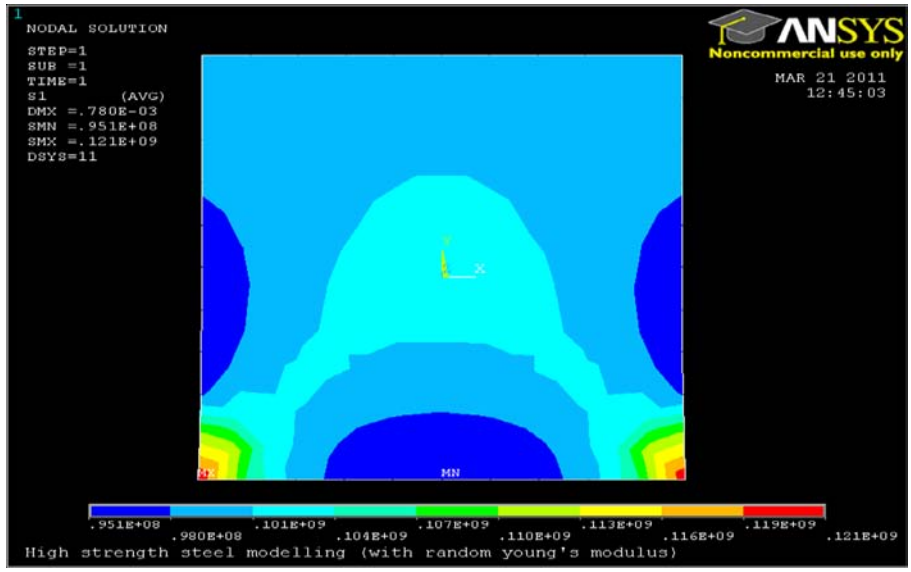

Figure 10.
The first principal stress σ_1 under constant material properties

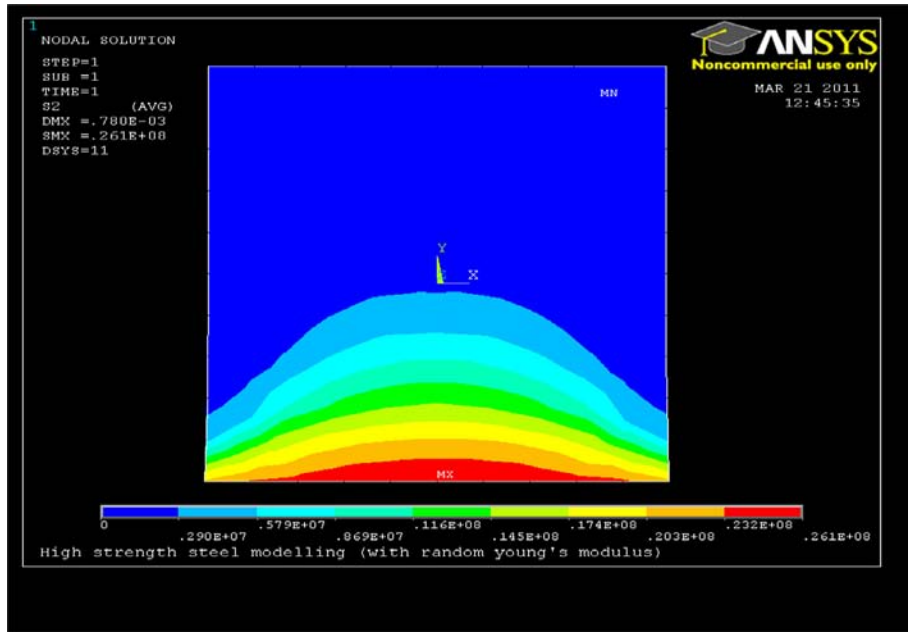

Figure 11.
The second principal stress σ_2 under constant material properties

6.1 Basic reliability theory

6.1.1 Random properties of loading S *and resistance* R. Basic structural reliability analysis considers only one load effect (such as the relevant displacement, principal stresses) S resisted by one resistance R (such as the designed maximum displacement,

the designed maximum principal stresses or strain). The load effect S and the resistance R are described by probability density functions $f_S(x)$ and $f_R(x)$, respectively. In general, S may be obtained from the applied loading Q through structural or mechanical analysis (either deterministic or with random components). It is important that R and S are expressed in the same units.

Without loss of generality, the safety of one structural element is considered here, and the structural element fails if its resistance R is less than the stress (or displacement) resultant S acting on it. The safety margin can be expressed as:

$$M = R - S \tag{24}$$

Let $F_R(x)$ denote the probability distribution function of R and assume that R and S are statistically independent, the failure probability P_f of the structural element can be computed as:

$$P_f = P(R \leq S) = P[M(R,S) \leq 0] = \int_{-\infty}^{\infty} F_R(x) f_S(x) dx \tag{25}$$

The function $M(R, S)$ is termed the "limit state function" and the probability of failure is identical with the probability of limit state violation.

The reliability $\Re$ is the probability that the structure will survive when the load is applied, and is given by:

$$\Re = 1 - P_f = 1 - \int_{-\infty}^{\infty} F_R(x) f_S(x) dx \tag{26}$$

6.1.2 Multiple random variables for a single structural member. For general structures, the load R and the resistance S are both functions of a set of random variables, such as:

R = function (material properties, dimensions).

S = function (applied loads, densities, dimensions).

In this case, the best solution to the problem is to express each limit state equation or failure function in terms of the set of n basic variables $\bar{X}$ which affect the structural performance, such that $M = f(X_1, X_2, \ldots, X_n) \leq 0$, where M is the safety margin sometimes referred to as the failure indicator. It is assumed that a set of basic variables $\bar{X} = (X_1, \ldots, X_n)$ is chosen in such a way that a failure surface can be defined in the n-dimensional basic variable space $\boldsymbol{\omega}$.

The failure surface is the surface dividing the basic variable space into two regions namely a failure region $\boldsymbol{\omega}_f$ and a safe region $\boldsymbol{\omega}_s$ (Thoft-Christensen and Baker, 1982; Neville and Kennedy, 1964; Melchers, 1999). The failure region contains all realizations of $\bar{X}$ that would result in failure, and the safe region contains all realizations of $\bar{X}$ that would not result in failure.

It is convenient to describe the failure surface by an equation of the form as follows:

$$f(\bar{X}) = f(X_1, X_2, \ldots, X_n) = 0 \tag{27}$$

$$f(\bar{X}) = \begin{cases} > 0 & \text{when } \bar{\mathrm{X}} \in \boldsymbol{\omega}_s \\ \leq 0 & \text{when } \bar{\mathrm{X}} \in \boldsymbol{\omega}_{\mathrm{f}} \end{cases} \tag{28}$$

Consequently, the reliability of the structure can be expressed as:

$$\Re = 1 - P_f = 1 - \underbrace{\int\int \ldots \int}_{f(\bar{X}) \leq 0} f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \tag{29}$$

in which $f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)$ is the joint probability density function for the variables $\bar{X}$.

If all the basic variables $\bar{X}$ are statistically independent, equation (29) can be expressed as:

$$\Re = 1 - P_f = 1 - \underbrace{\int\int \ldots \int}_{f(\bar{X}) \leq 0} f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n) dx_1 dx_2 \ldots dx_n \tag{30}$$

In general, there will be many basic random variables $\bar{X} = (X_1, \ldots, X_n)$ describing the structural reliability problem. In the case of complex structures, n could be very large indeed, which can cause a problem for numerical integration. In order to mitigate this "curse of dimensionality", a reliability index is defined as:

$$\begin{gathered} \beta = \min \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \min (x^T \cdot x)^{1/2} \\ \text{Subject to } M(\bar{x}) = 0 \end{gathered} \tag{31}$$

where the x_i represent the realizations of random variables X_i. The reliability index β represents the shortest distance from the origin to the failure surface, and generally speaking, a larger value of β indicates a safer structure.

6.1.3 Reliability index for linear failure functions. For the case of a linear safety margin M and normal basic variables, let the safety margin M be linear with respect to the basic variables $X_1, \ldots, X_n$:

$$M = a_0 + a_1 X_1 + \cdots + a_n X_n \tag{32}$$

Its mean value is given by:

$$\mu_M = a_0 + a_1 \mu_1 + \cdots + a_n \mu_n \tag{33}$$

and variance is:

$$\sigma_M^2 = a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2 + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \rho_{ij} a_i a_j \sigma_i \sigma_j \tag{34}$$

Following equation (31), the reliability index can be expressed as:

$$\beta = \frac{\mu_M}{\sigma_M} \tag{35}$$

This above formulation of the reliability index β was used by Cornell as early as 1969 Cornell (1967, 1969).

6.1.4 Reliability index for nonlinear failure functions.

(1) Hasofer Lind algorithm. If the safety margin M is nonlinear in $\bar{X} = (X_1, \ldots, X_n)$ (Hasofer and Lind, 1974) introduced another way to calculate the reliability index β. The first step in defining Hasofer and Lind's reliability index is to normalize the set of basic variables. The new set $\bar{Z} = (Z_1, \ldots, Z_n)$ is defined as:

$$Z_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}} \tag{36}$$

where μ_{X_i} and σ_{X_i} are the mean and the standard deviation of the random variables X_i. Note that:

$$\mu_{Z_i} = 0 \quad \text{and} \quad \sigma_{Z_i} = 1, \quad i = 1, 2, \ldots, n \tag{37}$$

The linear mapping defined in equation (37) transforms the failure surface in the X-coordinate system into a failure surface in the Z-coordinate system. Hasofer and Lind's reliability index β is defined as the shortest distance from the origin to the failure surface in the normalized Z-coordinate system (Thoft-Christensen and Baker, 1982; Nakagiri, 1989).

In the general case where the failure surface is nonlinear, an iterative method is needed. Provided that the failure functions is differentiable, the distance β can be determined by solving the following $n + 1$ equations iteratively:

$$x_i = \frac{-(\partial f / \partial Z_i)(\beta \bar{x})}{\left[\sum_{k=1}^{n} \left(\partial f / \partial Z_k(\beta \bar{x})\right)^2\right]^{1/2}} \tag{38}$$

$$f(\beta x_1, \beta x_2, \ldots, \beta x_n) = 0 \tag{39}$$

(2) Lagrangian multiplier method. In mathematical optimization, the method of Lagrange multiplier provides a strategy for finding the maxima and minima of a function subject to constraints. Thus, for the optimization problem (31), by introducing a Lagrange multiplier λ, the following Lagrange function can be defined:

$$\Lambda(\bar{x}, \lambda) = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} + \lambda \cdot M(\bar{x}) \tag{40}$$

As constrained optimization problems have been addressed in numerous textbooks (Fletcher, 1987), readers are recommended to refer to this literature.

6.2 Reliability analysis of a simply supported beam

Reliability analysis is based on the definition of a limit state function depending on the output of the analysis. A displacement-based state limit state function is considered for example. That is:

$$M = u - u^i \tag{41}$$

where u^i is a random nodal displacement and u is a prescribed threshold (such as design values according to British Standards).

As shown in Figure 12, a simply supported beam made of concrete is considered here. The concrete beam with a span of 2 m is designed to sustain a static distributed loading on the top edge. The concrete material is assumed to be isotropic, and the density and the Poisson's ratio are assumed to be deterministic constants, i.e. $\rho = 1{,}500\,\text{kg/m}^3$ and $v = 0.2$. The only random material property is the Young's modulus:

$$E = E(x, \theta) = 60\,\text{Gpa} \quad \forall x \in D \tag{42}$$

$$\begin{aligned} &Cov(E(x_1, \omega), E(x_2, \omega)) \\ &= 144 e^{-((x_2 - x_1)^2 + (y_2 - y_1)^2)/0.5^2}\,\text{Gpa}^2 \quad \forall x_1, x_2 \in D \end{aligned} \tag{43}$$

Following the F-K-L representation scheme and SFEM formulation (Li *et al.* 2006b, 2008), a stochastic linear equation is obtained, which consists of $m = 4$ real symmetric matrices with a set of mutually independent standard Gaussian random variables $a_i (i = 1, \ldots, 4)$, in which, the first variable a_1 always equals 1 because the first matrix is deterministic. The proposed joint diagonalization method "DPSDM" is then applied to diagonalize the stochastic matrix set. The total value of the off-diagonal entries after diagonalization, i.e. $\sum_{i=1}^{m} off(K_i)$, is significantly reduced by the DPSDM method. From the resulting average transformation structure, the ratio of the off-diagonal entries to the Frobenius norm is 8.7×10^{-3}, i.e.:

$$\sum_{i=1}^{m} off\left(K_i^*\right) : \sum_{i=1}^{m} \left\| K_i^* \right\|_F^2 = 8.7 \times 10^{-3} : 1 \tag{44}$$

The limit state function is defined in terms of the maximum vertical displacement at the centre of the bottom edge, i.e. node 34, as:

$$M = u - u_{34} \tag{45}$$

where u is an ultimate design displacement which equals span/250 according to Gulvanessian and Holicky (2005) and British Standard Institution (1990a, b). A total number of 1,000 sample solutions of the stochastic linear system are calculated after the DPSDM diagonalization. As discussed in Section 6.1.3, the corresponding calculation of the reliability index can be calculated straightforwardly as:

$$\beta = \frac{\mu_M}{\sigma_M} = \frac{3.1 \times 10^{-3}}{7.21914 \times 10^{-4}} = 4.29414 \tag{46}$$

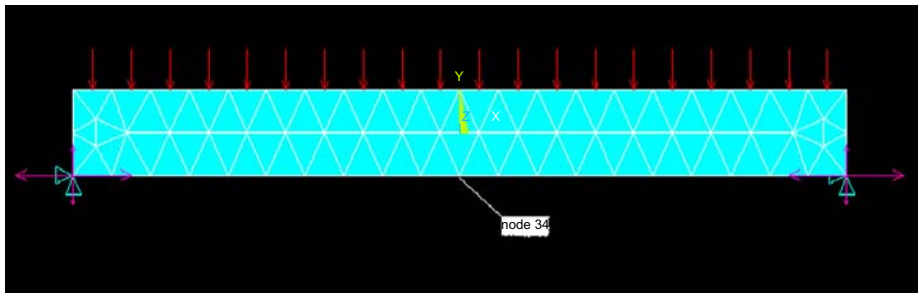

Figure 12.
Configuration of the simply supported beam

In order to verify the accuracy of the joint diagonalization strategy, the averaged sample solutions are compared with the Monte Carlo method. The following solutions calculated by the Monte Carlo method agree very well with solutions calculated by DPSDM. The required reliability index for this concrete beam according to the British Standard Institution (1990a, b) is 3.8, and the result shows that the ultimate design of the concrete beam under static distributed loadings is satisfactory and reliable. Besides the joint diagonalization strategy, as discussed the in literature review, any stochastic responses can be expanded as polynomial chaos, and a generalized polynomial chaos scheme proposed by Xiu (2010) is selected for comparison. The n-variate pth-degree gPC basis functions are the products of the univariate gPC polynomials of total degree less than or equal to p, and its general form is:

$$u(x) = \sum_{j=0}^{p-1} u_j \Psi_j(x) \tag{47}$$

Substituting equation (47) into equation (45), the corresponding limit state function can be rewritten as:

$$\begin{aligned} M &= 0.008 - \sum_{j=0}^{2} u_j \Psi_j(x) \\ &= -1.079 \times 1\hat{0} - 4x_1 + 1.627 \times 1\hat{0} - 4x_2 + 1.401 \times 1\hat{0} - 4x_3 \\ &\quad + 8.29 \times 1\hat{0} - 5x_1^2 - 5.57 \times 1\hat{0} - 5x_2^2 + 4.338 \times 1\hat{0} - 4x_3^2 \\ &\quad + 1.95 \times 1\hat{0} - 4x_1x_2 + 1.761 \times 1\hat{0} - 4x_1x_3 + 9.19 \times 1\hat{0} - 5x_2x_3 + 0.012439 \end{aligned} \tag{48}$$

In the above equation, up to the third order gPC $p = 2$ has been chosen to expand the displacement at node 34, in which each component $u_j (j = 0, \ldots, p-1)$ is calculated by collocation method (Hairer, 1993; Iserles, 1996) and x_1, x_2, x_3 stands for Gaussian random variables. Solving the reliability index β at the design point is a constrained nonlinear optimization problem, and the nonlinear methods discussed in Section 6.1.4 can be employed. A standard Hasofer Lind algorithm is employed here to calculate the reliability index β and the final converged solution of the reliability index β is 4.9. It is greater than the value of 4.2 calculated by DPSDM. The error between the Hasofer Lind method and DPSDM method may be produced by the truncated generalized polynomial chaos expressions of the displacement under consideration.

7. Conclusion

A novel algorithm for joint diagonalization of real symmetric matrices is presented in this paper. The new algorithm is based on the least-squares criterion, and it iteratively searches for the optimal transformation matrix based on the gradient of the cost function, which can be computed in a closed form. Numerical examples show that the new algorithm is efficient and robust. The new algorithm is applied in conjunction with stochastic finite element methods, and very promising results are observed which match very well with the Monte Carlo method, but with higher computational efficiency. The new method is also tested in the context of structural reliability analysis. The reliability index obtained with the joint diagonalization approach is compared with the conventional Hasofer Lind algorithm, and again good agreement is achieved.

References

Al-Baali, M. (1985), "Descent property and global convergence of the Fletcher-Reeves method with inexact line search", *IMA J. Numer. Anal.*, Vol. 5, pp. 121-4.

Al-Baali, M. and Fletcher, R. (1996), "On the order of convergence of preconditioned nonlinear conjugate gradient methods", *SIAM J. Sci. Comput.*, Vol. 17, pp. 658-65.

Askey, R. and Wilson, J.A. (1985), "Some basic hypergeometric polynomials that generalize Jacobi polunomials", *American Mathematical Society*, Vol. 319, pp. 1-55.

Battiti, R. (1992), "1st-order and 2nd-order methods for learning-between steepest descent and Newton method", *Neural Computation*, Vol. 4, pp. 141-66.

Bayer, V., Roos, D. and Adam, U. (2007), "Structural reliability analysis by random field, modeling with robustness methods and adaptive response surfaces", paper presented at 11 th International Conference on Civil, Structural and Enviromental Engineering Computing, St Julians, Malta.

Bjorck, A. (1996), *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA.

British Standard Institution (1990a), *EN 1990: Eurocode: Basis of Structural Design, Annex B*, British Standard Institution, London.

British Standard Institution (1990b), *EN 1990: Eurocode: Basis of Structural Design, Annex C*, British Standard Institution, London.

Cornell, C.A. (1967), "Bounds on the reliability of structural systems", *J. Struct. Div., ASCE*, Vol. 30.

Cornell, C.A. (1969), "A probability based structural code", *J. Amer. Concrete Inst.*, Vol. 66, pp. 974-85.

Cropper, W.H. (2004), *Great Physicists: The Life and Times of Leading Physicists from Galileo to Hawking*, Oxford University Press, Oxford.

Dai, Y.H. and Yuan, Y. (1995), "Further studies on the Polak-Ribiere-Polyak method", Research Report ICM-95-040, Chinese Academy of Science.

Dai, Y.H. and Yuan, Y. (1996), "Convergence of the Fletcher-Reeves method under a generalized Wolfe search", *Journal Computational Mathematics of Chinese Universities*, Vol. 2, p. 7.

Dai, Y.H. and Yuan, Y. (1998), "A class of globally convergent conjugate gradient methods", Research Report ICM-98-030, Institue of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Science.

Evans, D.H. (1992), *Probability and Its Applications for Engineers*, ASQC Quality Press, New York, NY.

Fishman, G.S. (1995), *Monte Carlo: Concepts, Algorithms, and Applications*, Springer, New York, NY.

Fletcher, R. (1987), *Practical Methods of Optimization*, Wiley, New York, NY.

Fletcher, R. and Powell, M.J.D. (1963), "A rapidly convergent descent method for minimization", *The Computational Journal*, Vol. 6, pp. 163-8.

Fletcher, R. and Reeves, C.M. (1964), "Function minimization by conjugate gradients", *The Computational Journal*, Vol. 7, pp. 149-54.

Ghanem, R.G. and Spanos, P.D. (2003), *Stochastic Finite Elemens: A Spectral Approach*, Springer, New York, NY.

Gilbert, J.C. and Nocedal, J. (1992), "Global convergence properties of conjugate gradient methods for optimization", *SIAM Journal of Optimization*, Vol. 2, pp. 21-42.

Gill, P.E., Murray, W. and Wright, M.H. (1981), *Practical Optimization*, Academic Press, New York, NY.

Golub, G.H. and Pereyra, V. (1973), "The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate", *SIAM J. Numer.*, pp. 413-32.

Grippo, L. and Lucidi, S. (1997), "A globally convergent version of the Polak-Ribiere conjugate gradient method", *Mathematics Programming*, Vol. 78, pp. 375-91.

Gulvanessian, H. and Holicky, M. (2005), "Eurocodes: using reliability analysis to combine action effects", *Proceedings of the Institution of Civil Engineers*, pp. 243-52.

Hairer, E. (1993), *Solving Ordinary Differential Equations I: Nonstiff Problems*, Springer, Berlin.

Hasofer, A.M. and Lind, N.C. (1974), "Exact and invariant second-moment code format", *Journal of the Engineering Mechanics Division*, Vol. 100, pp. 111-21.

Huh, J., Lee, S.Y. and Haldar, A. (2003), "Reliability evaluation using finite element method", *Proceedings of the Fourth International Symposium on Uncertainty Modelling and Analysis*.

Iserles, A. (1996), *A First Course in the Numerical Analysis of Differential Equations*, Cambridge University Press, Cambridge.

Khoda, K.M., Lin, Y. and Storey, C. (1992), "Generalized Polak-Ribiere algorithm", *Journal of Optimization Theory and Applications*, Vol. 75, pp. 345-54.

Li, C.F., Feng, Y.T. and Owen, D.R.J. (2006a), "Explicit solution to stochastic system of linear algebraic equations $(a_1A_1 + \cdots a_1A_1)x = \mathrm{b}$", *Computer Methods in Applied Mechanics and Engineering*, Vol. 195 Nos 44-47, pp. 6560-76.

Li, C.F., Adhikari, S., Cen, S., Feng, Y.T. and Owen, D.R.J. (2009), "A joint diagonalisation approach for linear stochastic systems", *Computers & Structures*, Vol. 88, pp. 1137-48.

Li, C.F., Feng, Y.T., Owen, D.R.J., Li, D.F. and Davies, I.M. (2006b), "Fourier representation of random media fields in stochastic finite element modelling", *Engineering Computations*, Vol. 23, pp. 794-817.

Li, C.F., Feng, Y.T., Owen, D.R.J., Li, D.F. and Davies, I.M. (2008), "A Fourier-Karhunen-Loeve discretization scheme for stationary random material properties in SFEM", *International Journal for Numerical Methods in Engineering*, Vol. 73, pp. 1942-65.

Melchers, R.E. (1999), *Structural Reliability Analysis and Prediction*, Wiley, Chichester.

Nakagiri, S. (1989), "Reliability enhancement of portal frame structure by finite element synthesis", *Transactions of the International Conference on Structural Mechanics in Reactor Technology,* Vol. 10, pp. 155-60.

Neville, A.M. and Kennedy, J.B. (1964), *Basic Statistical Methods for Engineers and Scientists*, Harper & Row, London.

Nocedal, J. (1991), "Theory of algorithm for unconstrained optimization", *Acta Numerical*, Vol. 1, pp. 199-242.

Nocedal, J. (1995), "Conjugate gradient methods and nonlinear optimization", *Proceedings of the AMS-IMS-SIAM Summer Research Conference on Linear and Nonlinear Conjugate Gradient-related Methods*.

Nocedal, J. (1996), *Large Scale Unconstrained Optimization*, Northwestern University, Chicago, IL.

Nocedal, J. and Wright, S.J. (1999), *Numerical Optimization*, Springer, New York, NY.

Oldenburg, D.W., McGillivray, P.R. and Ellis, R.G. (1993), "Generalized subspace methods for large-scale inverse problems", *Geophysical Journal International*, Vol. 114, pp. 12-20.

Rosenbrock, H.H. (1960), "An automatic method for finding the greatest or least value of a function", *The Computer Journal*, Vol. 3, pp. 175-84.

Sachdeva, S.K., Nair, P.B. and Keane, A.J. (2006), "Comparative study of projection schemes for stochastic finite element analysis", *Computer Methods in Applied Mechanics and Engineering*, Vol. 195, pp. 2371-92.

Shewchuk, J.R. (1994), "An introduction to the conjugate gradient method without the agonizing pain", Technical Report, August.

Sudret, B. and Der Kiureghian, A. (2000), *Stochastic Finite Element Methods and Reliability*, University of California, Berkeley, CA.

Sudret, B. and Der Kiureghian, A. (2002), "Comparison of finite element reliability methods", *Probabilistic Engineering Mechanics*, Vol. 17, pp. 337-48.

Sudret, B., Berveiller, M. and Lemaire, M. (2003), "Application of a stochastic finite element procedure to reliability analysis", *Proceedings of the 11th IFIP WG7.5 Conference on Reliability and Optimization of Structural Systems, Banff, Canada*.

Thoft-Christensen, P. and Baker, M.J. (1982), *Structural Reliability Theory and Its Applications*, Springer, Berlin.

Wang, X. (2008), "Method of steepest descent and its application", *IEEE Microwave and Wireless Components Letters*, Vol. 12, pp. 24-6.

Wiener, N. (1938), "The homogeneous chaos", *American Journal of Mathematics*, Vol. 60, pp. 897-936.

Wolfe, P. (1969), "Convergence conditions for ascent methods", *SIAM Review*, Vol. 11, pp. 226-35.

Xiu, D. (2010), *Numerical Methods for Stochastic Computations*, University Press, Princeton, NJ.

Xu, H. and Rahman, S. (2005), "Decomposition methods for structural reliability analysis", *Probabilistic Engineering Mechanics*, Vol. 20, pp. 239-50.

Further reading

Cameron, R. and Martin, W. (1947), "The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals", *Ann. Math*, Vol. 48, pp. 385-92.

Corresponding author

Chenfeng Li can be contacted at: c.f.li@swansea.ac.uk