Stochastic Finite Element Approaches for Constraint Optimization

Mihai Anitescu,

LANS Seminar,

November 1, 2006

Estimating the solution of parametric equations

- Consider the parametric nonlinear equation

$$F(x,\omega) = 0, \quad x \in R^n,\ \omega \in \Omega \subset R^m,$$
$$F : R^n \times R^m \to R^n, \text{ a smooth function.}$$

- Question: Can I compute (or at least, efficiently approximate) the solution mapping $\tilde{x}(\omega)$?
- This is the essential step in uncertainty quantification, in cases where the parameters (such as diffusion coefficients, cross-sections, etc) are known only with some error or have a statistical distribution.
- Monte Carlo answer: Yes, if I sample enough. Use parameter PDF and $\{\omega_i\},\ i = 1, 2, \ldots, M \Rightarrow \{\tilde{x}(\omega)_i = F^{-1}(0, \omega_i)\}$

Followed by "histogram" or other post processing.

- At each step I must solve the original problem for one parameter. It is a nightmare if one has many parameters and/or each subproblem solve is expensive

The stochastic finite element (SFEM) setup

The mapping $\tilde{x}(\omega)$ is approximated by a Fourier type expansion with respect to a basis of polynomials, orthogonal with respect to the probability density function (PDF) of ω $P_0(\omega),\ P_1(\omega),\ \ldots P_K(\omega)$ such that

$$E_\omega \left[P_i\left(\omega\right) P_j\left(\omega\right) \right] = \delta_{ij}, \quad 0 \leq i, j \leq K$$

. This stochastic interpretation is the origin of the name *polynomial chaos*

The stochastic finite element (SFEM) answer … that has nothing stochastic in it

Take $x_0, x_1 \ldots x_K \in R^n$. We define the spectral approximation $\tilde{x}(\omega; x_0, x_1, \ldots x_K) = \sum_{i=0}^{K} x_i P_i(\omega)$. SFEM defines the coefficients from Galerkin projection condition:

$$E_\omega \left(F \left(\tilde{x}(\omega; x_0^*, x_1^*, \ldots x_K^*), \omega \right) P_k(\omega) \right) = 0_n, \qquad k = 0, 1, \ldots, K$$

- This results in a system of nonlinear equations, which is K+1 times larger than the original system of equations.
- Once the system is solved, the distribution $\tilde{x}(\omega)$ can be sampled with Monte Carlo without solving any additional systems of nonlinear equations. So sampling is decoupled from the most computationally expensive part, which is deterministic.
- As shown in my previous talk, the computational savings can be tremendous over regular Monte Carlo, if instance problem solves are expensive.

SFEM issues

- Convergence of the method or well posedness of the resulting nonlinear equations is virtually NEVER analyzed in Spanos, Ghanem, and co-authors, unless the SFEM equation itself is linear.
- In the case of optimization problems, it is inconvenient to start with an optimization problem and end with a system of nonlinear equations, since a lot of solver capability is lost (not to mention stability).
- We attempt to resolve these issues in this work.
- In addition, the number of polynomials of a given degree increase exponentially with the dimension of the parameter space, so how to chose them wisely is an important issue not addresed here; nonetheless it is a problem that plagues all SFEM approaches, so we will concentrate on the first two issues.

The scalar product

Define

$$\langle g, h \rangle_W = \int_\Omega W(\omega) g(\omega)^T h(\omega) d\omega.$$

where g, h are continuous functions from $\mathrm{I\ R}^m$ to $\mathrm{I\ R}^p$. Here $\Omega \in \mathrm{I\ R}^m$ is an open set and $w(\omega)$ is a weight function.

We will also use the notation

$$E_\omega[g(\omega)] =< g(\omega) >= \langle g, 1 \rangle$$

though the latter is not a scalar product and may even be multidimenional.

The weight function

1. The weight function $W(\omega) : I\!R^m \rightarrow I\!R$ is nonnegative, that is, $W(\omega) \geq 0$, $\forall \omega \in \Omega$.

2. Any multivariable polynomial function $P(\omega)$ is integrable, that is,

$$\int_{\Omega} W(\omega)|P(\omega)|d\omega < \infty$$

With this definition,, we can define the semi norm

$$||g||_w = \sqrt{\langle g, g \rangle_W}$$

on the space of continuous functions. If, in addition, $||g||_W = 0 \Rightarrow g = 0$, then $||\cdot||_W$ is a norm.

Orthogonal polynomials

We can ortho normalize the set of polynomials in the variable $\omega$and we obtain the orthogonal polynomials $P_i(\omega)$ that satisfy

- $\left\langle P_i, P_j \right\rangle = \delta_{ij}$, $0 \leq i, j$. By convention, we always take P_0 to be the constant polynomial.

- The set $\{P_i\}_{i=0,1,2,\ldots}$ forms the basis of the complete space L^2_W.

- If $k_1 \leq k_2$, then deg $\left(P_{k_1}\right) \leq \left(P_{k_2}\right)$.

Function spaces

We define

$$L^2_{p,W} = \underbrace{L^2_W \otimes L^2_W \otimes \ldots L^2_W}_{p}$$

We can now define the Fourier coefficients as

$$c_k(f) = \langle f, P_k \rangle \in \mathrm{I\,R}^{\,p}, \quad f \in \text{Ł}^2_W,\ k = 0, 1, \ldots$$

The projection operator: $f \in L^2_W$ on to the space of the polynomials of degree at most K

$$\Pi^W_K(f) = \sum_{k,\deg(P_k) \leq K} c_k(f) P_k(\omega)$$

Bounds on the projection

$$f \in H_W^q \Rightarrow \left\| f - \Pi_W^K(f) \right\|_W \leq C \frac{1}{K^{\epsilon q}} \qquad (1)$$

When the function f is analytical q is unbounded and we obtain *exponential convergence*

In addition, a reciprocal also holds, that is, there exists a parameter t such that, that depends only on $W(x)$ such that

$$\sum_{k \in \mathcal{N}} ||c_k(f)|| \operatorname{deg}(P_k)^t < \infty \;\Rightarrow f \in C^1(\Omega) \qquad (2)$$

Example

For the case $m = 1$, $\Omega = [-1, 1]$, and $W(\omega) = \frac{1}{2}$, the ortho normal family are the normalized Legendre polynomial functions

$$P_k = \frac{1}{\sqrt{2}}\sqrt{\frac{2k+1}{2}}\frac{1}{2^k k!}\frac{d^k}{d\omega^k}\left(\omega^2 - 1\right)^k, \quad k = 1, 2, \ldots, K$$

SFEM for unconstrained optimization

$$(UO) \quad \min_{x \in R^n} f(x, \omega)$$

- We have two formulations: One is optimize and project (the nonlinear equation SFEM approach) and the other one is project and optimize (our approach).

$$E_\omega \left(\nabla f \left(\tilde{x}(\omega; x_0^*, x_1^*, \ldots x_K^*), \omega\right) P_k(\alpha)\right) = 0_n, \quad k = 0, 1, \ldots, K$$

$$(SUO(K)) \quad \min_{x_0, x_1, \ldots, x_K} \; \tilde{f}^K(\tilde{x}) = E_\omega \left[f(\tilde{x}, \omega)\right]$$

- Result: If f is smooth, the optimality conditions of (SUO(K)) are the same as the nonlinear equations of the optimize and project approach. THE SAME APPLIES FOR THE CONSTRAINED OPTIMIZATION APPROACH!!

SFEM formulation of eigenvalue problem (for one parameter)

$$\min \quad E_\omega\left[\left(\sum_{i=0}^{K} x_i P_i(\omega)\right)^T (Q+\omega D_Q)\left(\sum_{i=0}^{K} x_i P_i(\omega)\right)\right]$$

$$\text{s.t.} \quad E_\omega\left[\left(\sum_{i=0}^{K} x_i P_i(\omega)\right)^T \left(\sum_{i=0}^{K} x_i P_i(\omega)\right) P_k(\omega)\right] = E_\omega\left[P_k(\omega)\right] \quad k=1,2,\ldots,K$$

- Note that this problem can be *exactly* represented, even when multiple parameters are involved, modulo a series of quadratures

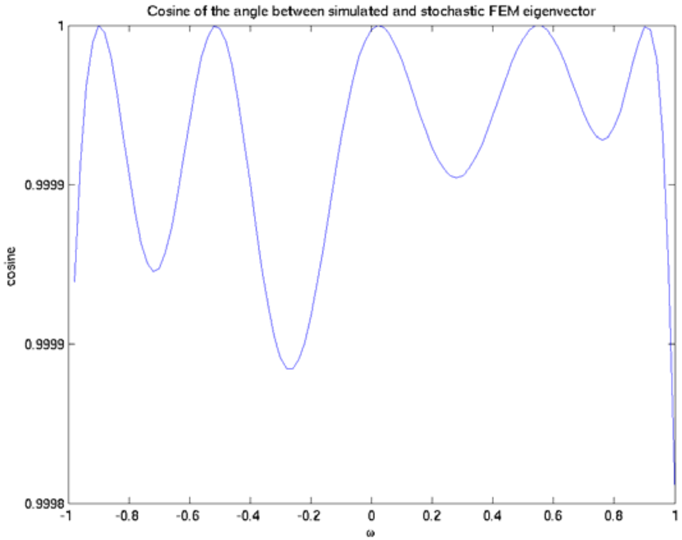
$$\min \quad \sum_{i=0}^{K} x_i^T Q x_i + \sum_{i=0}^{K}\sum_{j=0}^{K} x_i^T D_Q x_i E_\omega\left[P_i(\omega)(\omega)P_j(\omega)\right]$$

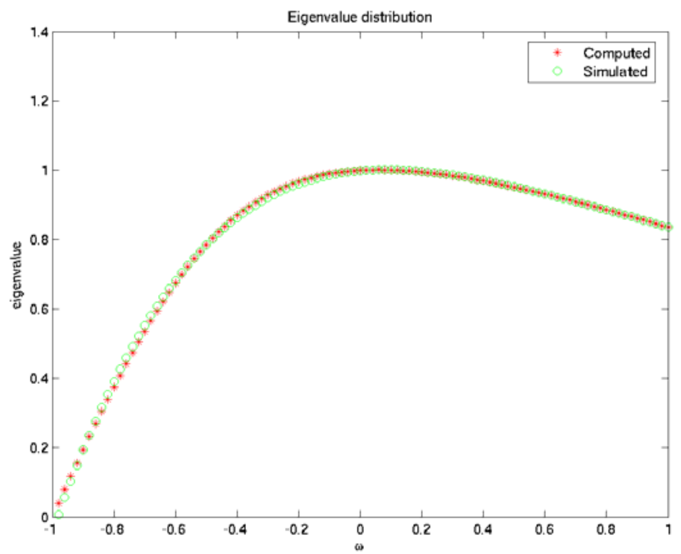
$$\text{s.t.} \quad \sum_{i=0}^{K}\sum_{j=0}^{K} x_i^T x_i E_\omega\left[P_i(\omega)P_k(\omega)P_j(\omega)\right] = E_\omega\left[P_k(\omega)\right] \quad k=0,1,2,\ldots,K$$

Numerical Example for n=2

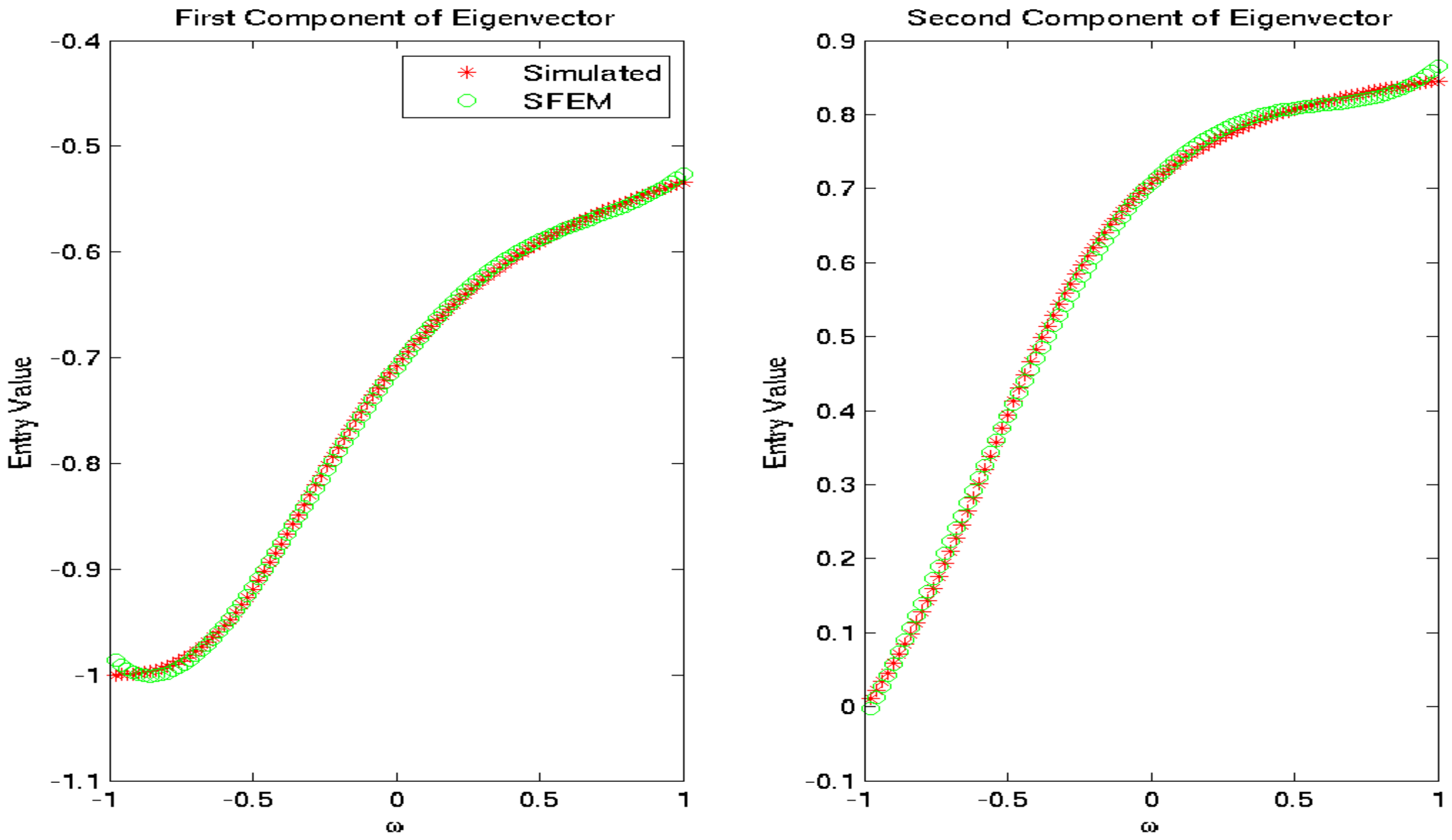
$$Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}; D_Q = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$$

- Use Legendre polynomials with K=4.
- Note that the variations allowed are not small variations ! (though the matrix will stay positive definite).
- The error in the angle of the eigenvector is less than 1%.
- The eigenvalue distribution is excellent fit.



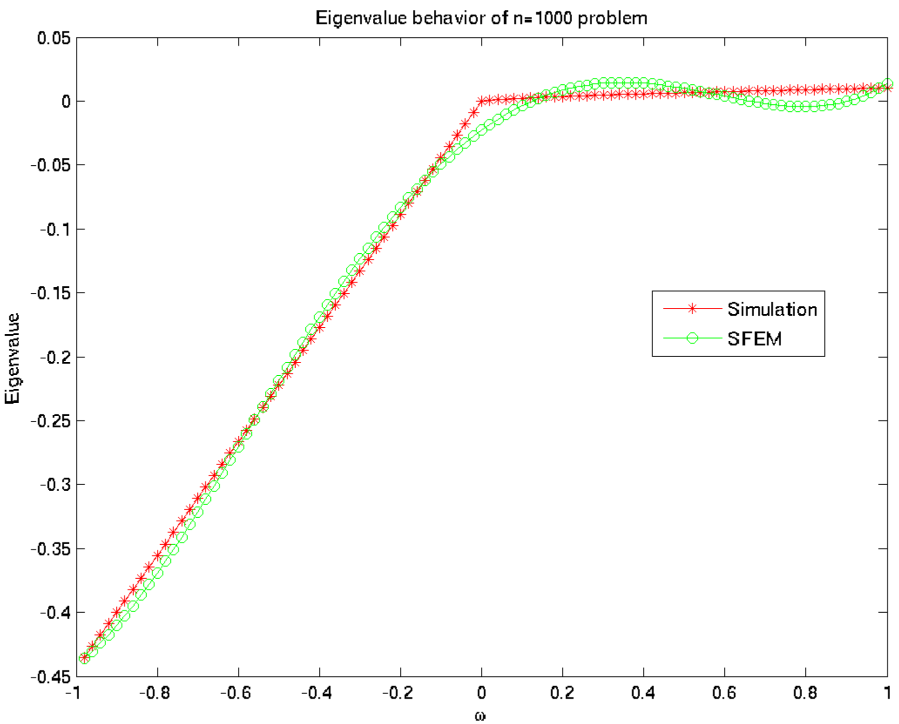


Example for n=2, eigenvector convergence



Numerical Example for n=1000

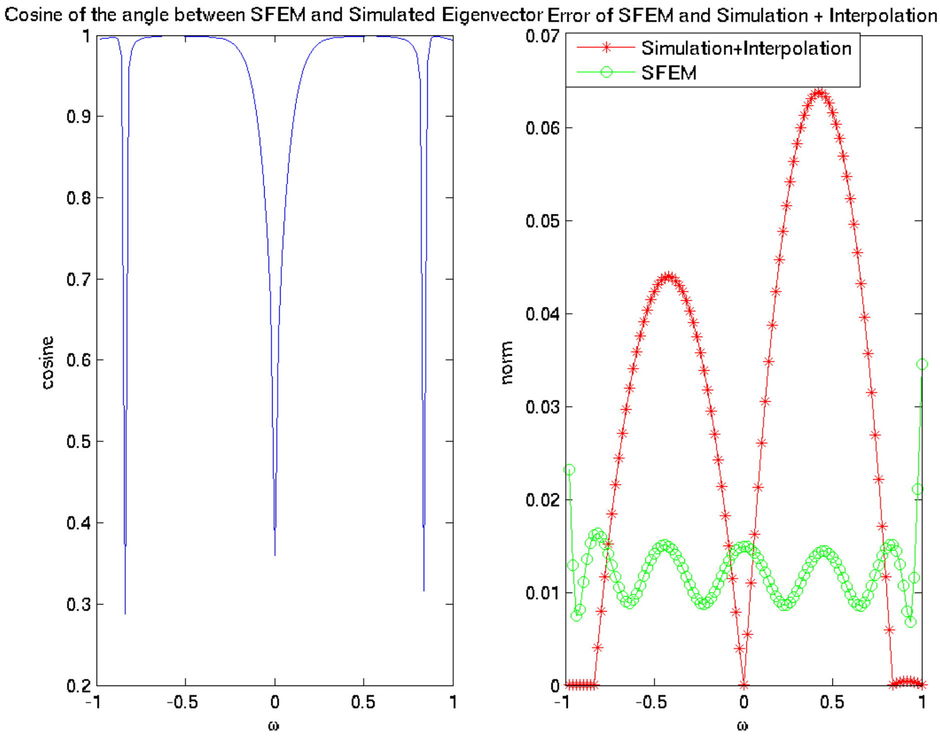
- Mimics the minimum eigenvalue problem for a one-dimensional diffusion equation, with uncertainty in absorption, akin to neutronics problems (where uncertainty in parameters also enters the matrix linearly)
- For K=4 The SCO formulation was solved in 7 seconds by KNITRO (one eig operation on same computer takes 13 seconds!)



$$Q_{i,j} = \begin{cases} 1 & i = j = 1 \\ 1 & i = j = n \\ -1 & |i-j| = 1 \\ 2 & 1 < i = j < n \\ 0 & \text{otherwise} \end{cases}, \quad D_{Q\,i,j} = \begin{cases} 2\dfrac{i}{n}\dfrac{n-i}{n}\cos\left(\dfrac{i}{n}\right) & i = j \\ 0 & \text{otherwise} \end{cases}$$

Numerical Example, eigenvector behavior

- At first sight, approximation is awful (left panel)
- Residual, $\left\|\left(Q+\omega D_Q\right)\tilde{x}(\omega)-\tilde{\lambda}(\omega)\tilde{x}(\omega)\right\|$ different story, typical error is 1%!
- The problem is the occasional degeneracy of the eigenvalue problem.
- If we sample eigenvector at 5 (well-chosen ☺) points, and interpolate linearly, we get an error that is typically larger by a factor of 3-4.
- If the distribution is peaked at the center of the interval, the situation is worse than for uniform distribution.
- **This shows that black-box approaches, even enhanced with AD, may be unadvisable for quantifying eigenvector uncertainty!**



What's next?

- Proof of well-posedness and consistency in the limit; finished (practically ☺)
- Coupled thermo-hydraulics model for one parameter (crossection)
- One group, multiple parameters.
- Multiple groups.
- Note that even if we consider only linear terms, we produce a superior result (since the linear approximation is better in a larger range). So we could produce better "sensitivites"
- Also note that in the large example sensitivities may have been pointless at the center of the interval!

Possible Grand Vision

- Do a principal value component analysis to get the important parameters.
- Compute the orthogonal polynomials with respect to the density function
- Create the SFEM approximation adaptively WITHIN THE LIMIT OF COMPUTATIONAL RESOURCES
- Adaptation is done by projecting the residual on the remaining coefficients.
- Use importance sampling Monte Carlo to validate the approach and capture the remaining randomness.

Possible approach.

Is (SUO(K)) bounded below? Assumptions

A1 Uniformly bounded level sets assumption: There exist a function $\chi(\cdot)$ that is convex, nondecreasing, and that has bounded level sets, and a parameter $\gamma > 0$ such that

$$\chi\left(||x||^{\gamma}\right) \leq f(x, \omega), \quad , \forall \omega \in \Omega. \tag{1}$$

A2 Smoothness assumption: The function $f(x, \omega)$ is twice continuously differentiable in both arguments.

Result The objective function of the problem (SUO(K)) has bounded level sets.

Convergence result

Let $\left(x_0^K, x_1^K, x_2^K, \ldots, x_{k-1}^K\right)$ be a solution of the problem SUO(K). We define $x_k^K = 0_n$, for $k \geq K$. In addition, we defined

$$x_K(\omega)^* = \sum_{i=0}^{k} x_i^K L_k(\omega)$$

Furthermore, assume

$$\sum_{k=0}^{\infty} \left|\left|x_k^K\right|\right|^2 k^2 < M_{1i}$$

Then there exists a function $x^*(\omega)$ continuously differentiable such that, for a subsequence of the sequence we have $x_{K_l}(\omega) \overset{l \to \infty}{\Longrightarrow} x^*(\omega)$ in L_W^2 and any such function $x^*(\omega)$ satisfies the first order conditions if of the optimization problem (SUO).