NUMERICAL METHODS FOR STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS AND THEIR CONTROL

Max Gunzburger

Department of Scientific Computing, Florida State University

gunzburg@scs.fsu.edu

London Mathematical Society Durham Symposium

Computational Linear Algebra for Partial Differential Equations

July 14 – 24, 2008, Durham, UK

All models are wrong, but some models are useful

George Box

Computational results are believed by no one, except for the person who wrote the code

Experimental results are believed by everyone, except for the person who ran the experiment

NUMERICAL METHODS FOR STOCHASTIC PDE'S FOR DUMMIES

WHERE I AM THE DUMMY

INTRODUCTORY REMARKS

Uncertainty is everywhere

- Physical, biological, social, economic, financial, etc. processes always involve uncertainties

- As a result, mathematical models of these processes should account for uncertainty

- Accounting for uncertainty in processes governed by partial differential equations can involve

  – random coefficients in the PDE, boundary condition, and initial condition operators

  – random right-hand sides in the PDE's, boundary conditions, and initial conditions

  – random geometry, i.e., random boundary shapes

- Uncertainty arises because
  - available data are incomplete
    - they are predictable but are too difficult (perhaps impossible) or costly to obtain by measurement
      $\longrightarrow$ media properties in oil reservoirs or aquifers
    - they are unpredictable
      $\longrightarrow$ wind shear, rainfall amounts
  - not all scales in the data and/or solutions can or should be resolved
    - it is too difficult (perhaps impossible) or costly to do so in a computational simulation
      $\longrightarrow$ turbulence, molecular vibrations
    - some scales may not be of interest
      $\longrightarrow$ surface roughness, hourly stock prices

- Of course, it is well known that two experiments run under the "same" conditions will yield different results

Modeling noise

- White noise – input data vary randomly and independently from one point of the physical domain to another and from one time instant to another
  - uncertainty is described in terms of uncorrelated random fields
  - examples: thermal fluctuations; surface roughness; Langevin dynamics

- Colored noise – input data vary randomly from one point of the physical domain to another and from one time instant to another according to a given (spatial/temporal) correlation structure
  - uncertainty is described in terms of correlated random fields
  - examples: rainfall amounts; bone densities; permeabilities within subsurface layers

- Random parameters – input data depend on a finite[†] number of random parameters
  - think of this case as "knobs" in an experiment
  - each parameter may vary independently according to its own given probability density
  - alternately, the parameters may vary according to a given joint probability density
  - examples: homogeneous material properties, e.g., Young's modulus, Poisson's ratio, speed of sound, inflow mass

[†]What we really mean is that the number of parameters is not only finite, but independent of the spatial/temporal discretization; this is not possible for the approximation of white noise for which the number of parameters increases as the grid sizes decrease

- Ultimately, for all three cases, on a computer
  one solves problems involving random parameters

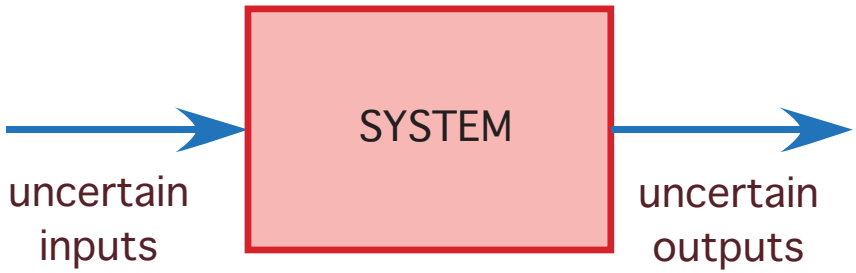
  – in the white noise and colored noise cases, one discretizes the noise so that the discretized noise is determined by a finite number of parameters

    - in the white noise case, the number of parameters has to increase as the spatial and/or temporal resolutions of the numerical scheme used to solve the PDEs increases

    - in the colored noise case, the number of parameters needed to approximate a correlated random field can, in practice, be chosen independently of the spatial/temporal resolutions

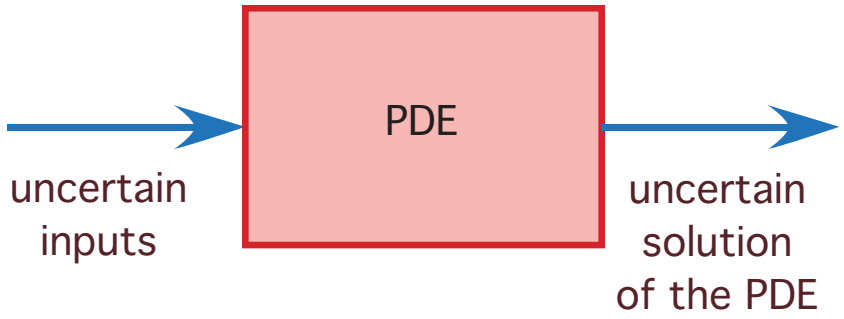
Uncertainty quantification

- Uncertainty quantification is the task of determining statistical information about outputs of a system, given statistical information about the inputs



  - of course, the system may have deterministic inputs as well

- We are interested in systems governed by partial differential equations

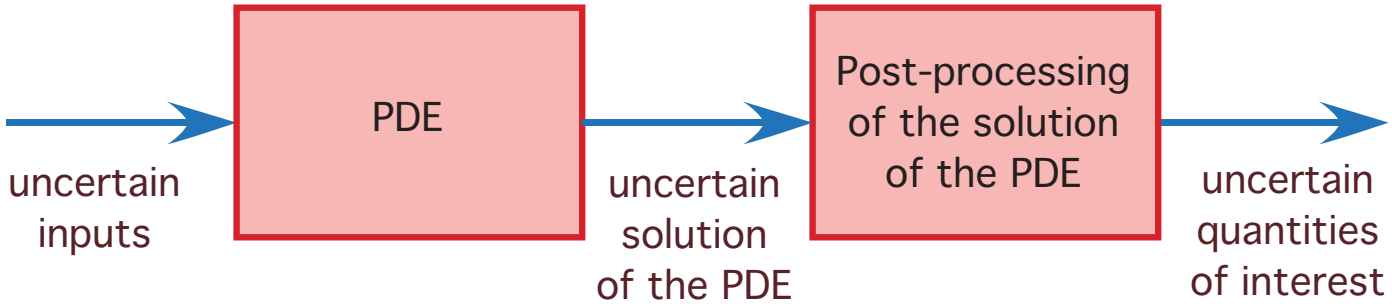


  - the solution of the partial differential equation defines the mapping from the input variables to the output variables

- Often, solutions of the PDE are not the primary output quantity of interest

  – quantities obtained by post-processing solutions of the PDE
    are more often of interest

      - of course, one still has to obtain a solution of
        the PDE to determine the quantity of interest



- A realization of the random system is determined by

  specifying a specific set of input variables

  and then

  using the PDE to determine the corresponding output variables

  – thus, a realization is a solution of a deterministic problem

- One is never interested in individual realizations of solutions of the PDE or of the quantities of interest

  – one is interested in determining statistical information about the quantities of interest, given statistical information about the inputs

Quantities of interest

- Suppose we have N random parameters $\{y_n\}_{n=1}^N$
  - we use the abbreviation $\vec{y} = \{y_1, y_2, \ldots, y_N\}$
  - each y_n could be distributed independently[†] according to its probability density function (PDF) $\rho_n(y_n)$ defined for y_n in a (possibly infinite) interval Γ_n
  - alternately, the parameters could be distributed according to a joint PDF $\rho(y_1, \ldots, y_N)$ that is a mapping from an N-dimensional set Γ into the real numbers
    - independently distributed parameters are the special case for which

$$\rho(y_1, \ldots, y_N) = \prod_{n=1}^{N} \rho_n(y_n) \qquad \text{and} \qquad \Gamma = \Gamma_1 \otimes \Gamma_2 \otimes \cdots \otimes \Gamma_N$$

[†]Without proper justification and sometimes incorrectly, it is almost always assumed that the parameters are independent; based on empirical evidence, sometimes this is a justifiable assumption in the parameters-are-"knobs" case, but for correlated random fields, it is justifiable only for the (spherical) Gaussian case; in general, independence is a simplifying assumption that is invoked for the sake of convenience, e.g., because of a lack of knowledge

- Realization = a solution $u(\mathbf{x}, t; \vec{y})$ of a PDE for a specific choice $\vec{y} = \{y_n\}_{n=1}^{N}$ for the random parameters
  - again, there is no interest in individual realizations

- One may be interested in statistics of solutions of the PDE
  - average or expected value

$$\overline{u}(\mathbf{x}, t) = \mathsf{E}[u(\mathbf{x}, t; \cdot)] = \int_{\Gamma} u(\mathbf{x}, t; \vec{y}) \rho(\vec{y})\, d\vec{y}$$

  - covariance

$$C_u(\mathbf{x}, t; \mathbf{x}', t') = \mathsf{E}\Big[\Big(u(\mathbf{x}, t; \cdot) - \overline{u}(\mathbf{x}, t)\Big)\Big(u(\mathbf{x}', t'; \cdot) - \overline{u}(\mathbf{x}', t')\Big)\Big]$$

$$= \int_{\Gamma} \Big(u(\mathbf{x}, t; \vec{y}) - \overline{u}(\mathbf{x}, t)\Big)\Big(u(\mathbf{x}', t'; \vec{y}) - \overline{u}(\mathbf{x}', t')\Big) \rho(\vec{y})\, d\vec{y}$$

  - variance $\quad C_u(\mathbf{x}, t; \mathbf{x}, t)$
  - higher moments

- One may instead be interested in statistics of spatial/temporal integrals of the solution of the PDE
  - for any fixed $\vec{y}$, we have, e.g.,

$$\mathcal{J}(t;\vec{y}) = \int_{\mathcal{D}} F(u;\vec{y})\, d\mathbf{x} \qquad \text{or} \qquad \mathcal{J}(\mathbf{x};\vec{y}) = \int_{t_0}^{t_1} F(u;\vec{y})\, dt$$

    or

$$\mathcal{J}(\vec{y}) = \int_{t_0}^{t_1} \int_{\mathcal{D}} F(u;\vec{y})\, d\mathbf{x} dt$$

    where $F(\cdot;\cdot)$ is given, $\mathcal{D}$ is a spatial domain, and (t_0, t_1) is a time interval

  - quantities defined with respect to integrals over boundary segments also often occur in practice

– examples

- the space-time average of u

$$\mathcal{J}(\vec{y}) = \int_{t_0}^{t_1} \int_{\mathcal{D}} u(\mathbf{x}, t; \vec{y})\, d\mathbf{x} dt$$

- if $\mathbf{u}$ denotes a velocity field, then

$$\mathcal{J}(t; \vec{y}) = \int_{\mathcal{D}} \mathbf{u}(\mathbf{x}, t; \vec{y}) \cdot \mathbf{u}(\mathbf{x}, t; \vec{y})\, d\mathbf{x}$$

is proportional to the kinetic energy

– again, one is not interested in the values of these quantities for specific choices of the parameters $\vec{y}$

- one is interested in their statistics

– example: expected value of the kinetic energy

$$\mathsf{E}\left[\int_{\mathcal{D}} \mathbf{u}(\mathbf{x},t;\vec{y})\cdot\mathbf{u}(\mathbf{x},t;\vec{y})\,d\mathbf{x}\right]$$

$$= \int_{\Gamma}\int_{\mathcal{D}} \mathbf{u}(\mathbf{x},t;\vec{y})\cdot\mathbf{u}(\mathbf{x},t;\vec{y})\rho(\vec{y})\,d\mathbf{x}\,d\vec{y}$$

- Thus, quantities of interest of this common type involve integrals over the parameter space[†]

  – e.g., for some $G(\cdot)$, integrals of the type

$$\int_{\Gamma} G\big(u(\mathbf{x},t;\vec{y})\big)\rho(\vec{y})\,d\vec{y} \qquad \text{or possibly} \qquad \int_{\Gamma} G\big(u(\mathbf{x},t;\vec{y});\mathbf{x},t,\vec{y}\big)\rho(\vec{y})\,d\vec{y}$$

[†]An important class of quantities of interest that arises in, e.g., reliability studies, but that we do not have time to consider involves integrals over a subset or Γ; in particular, we have

$$\int_{\Gamma} \chi_{u_0} G\big(u(\mathbf{x};\vec{y})\big)\rho(\vec{y})\,d\vec{y} = \int_{\Gamma_{u_0}} G\big(u(\mathbf{x};\vec{y})\big)\rho(\vec{y})\,d\vec{y}$$

where, for some given u_0

$$\chi_{u_0} = \begin{cases} 1 & \text{if } u(\mathbf{x};\vec{y}) \geq u_0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad \Gamma_{u_0} = \{\vec{y}\in\Gamma \text{ such that } u(\mathbf{x};\vec{y}) \geq u_0\}$$

- Ideally, one wants to determine an approximation of the PDF for the quantity of interest,

  i.e., more than just a few statistical moments

  of some output quantity

  - the quantity of interest is a PDF

  - one way (but not the only way) to construct the approximate PDF is to compute many statistical moments of the output quantity

    - so, again, we are faced with evaluating stochastic integrals

Quadrature rules for stochastic integrals

- Integrals of the type

$$\int_{\Gamma} G\big(u(\mathbf{x},t;\vec{y})\big)\rho(\vec{y})\,d\vec{y}$$

cannot, in general, be evaluated exactly

- Thus, these integrals are approximated using a quadrature rule

$$\int_{\Gamma} G\big(u(\mathbf{x},t;\vec{y})\big)\rho(\vec{y})\,d\vec{y} \approx \sum_{q=1}^{Q} w_q G\big(u(\mathbf{x},t;\vec{y}_q)\big)\rho(\vec{y}_q)$$

for some choice of

quadrature weights $\{w_q\}_{q=1}^{Q}$ (real numbers)

and

quadrature points $\{\vec{y}_q\}_{q=1}^{Q}$ (points in the parameter domain Γ)

– Alternately, sometimes the probability density function is used in the determination of the quadrature points and weights so that instead one ends up with the approximation

$$\int_{\Gamma} G\big(u(\mathbf{x},t;\vec{y})\big)\rho(\vec{y})\,d\vec{y} \approx \sum_{q=1}^{Q} w_q G\big(u(\mathbf{x},t;\vec{y}_q)\big)$$

- Monte Carlo integration – the simplest rule $\Longrightarrow$
  - randomly select Q points in Γ according to the PDF $\rho(\vec{y})$
  - evaluate the integrand at each of the sample points
  - average the values so obtained
    - i.e., for all q, $w_q = 1/Q$
      - more on Monte Carlo and other quadrature rules later

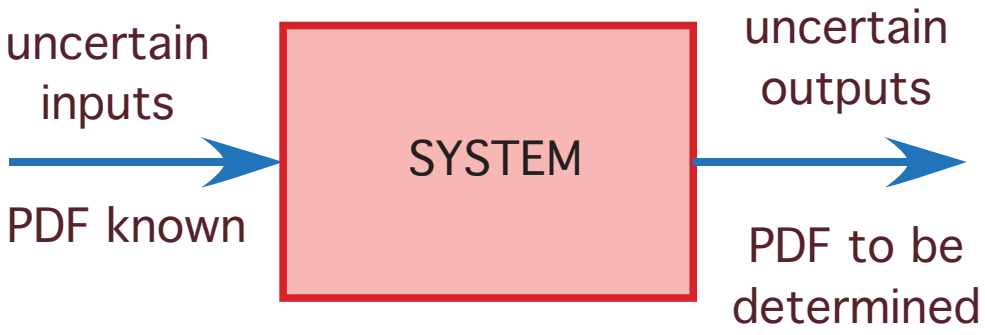
Big problem

- In practice, one usually does not know much about the statistics of the input variables
  - one is lucky if one knows a range of values, e.g., maximum and minimum values, for an input parameter
    - in which case one often assumes that the parameter is uniformly distributed over that range
  - if one is luckier, one knows the mean and variance for the input parameter
    - in which case one often assumes that the parameter is normally distributed
  - of course, one may be completely wrong in assuming such simple probability distributions for a parameter
- This leads to the need to solve stochastic model calibration problems

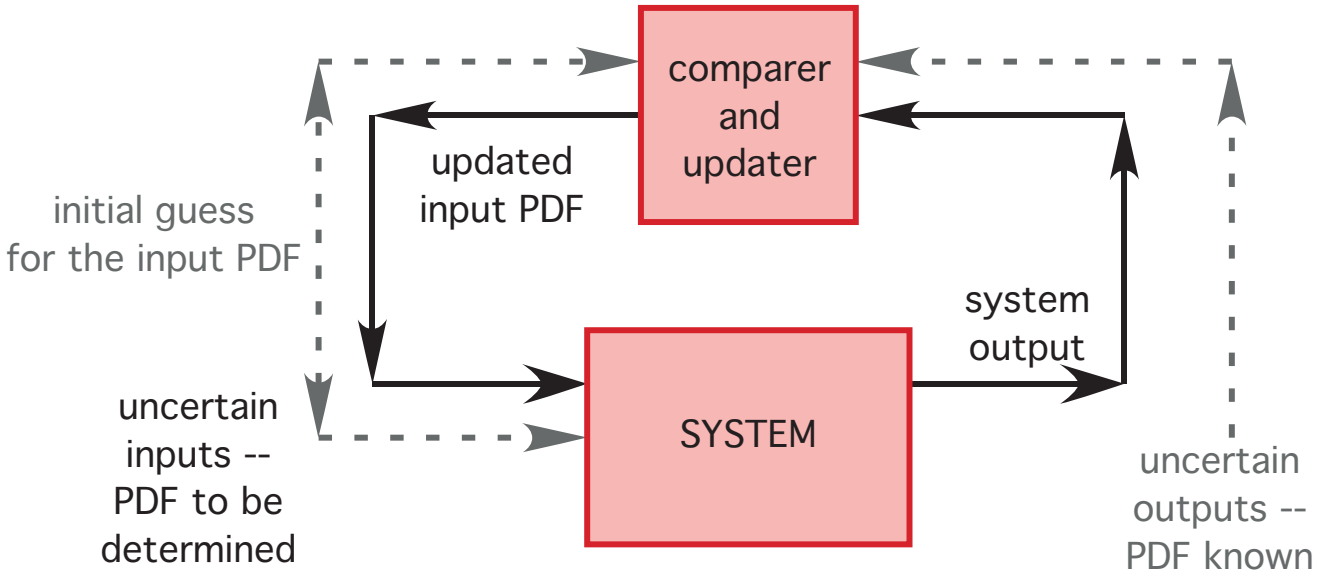
Model calibration

- Model calibration is the task of determining statistical information about the inputs of a system, given statistical information about the outputs
  - e.g., one can use experimental observations to determine the statistical information about the outputs
  - in particular, one wants to identify the probability density functions (PDF) of the input variables

- Of course, the system still maps the inputs to the outputs
  - thus, determining the input PDF is an inverse problem
  - usually involves an iteration in which guesses for the input PDF are updated
  - several ways to do the update, e.g., Baysean, maximum likelyhood, ...



Uncertainty quantification – direct problem

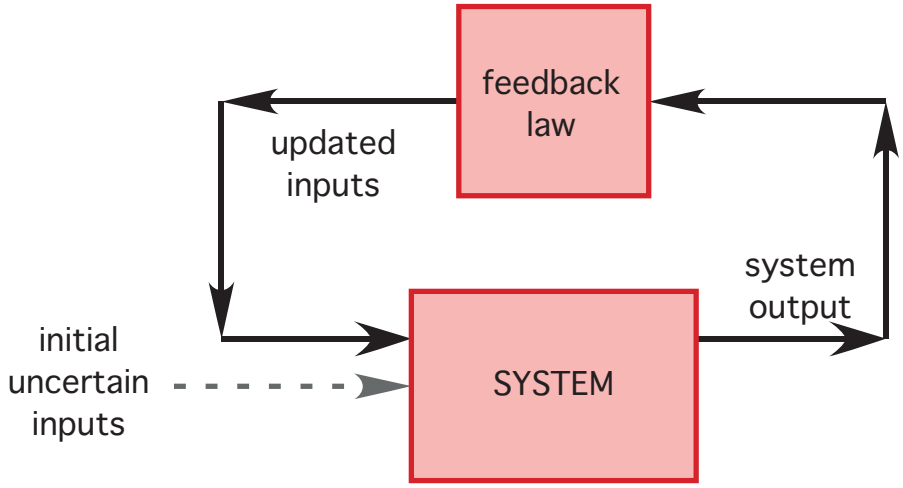


Model calibration – inverse problem

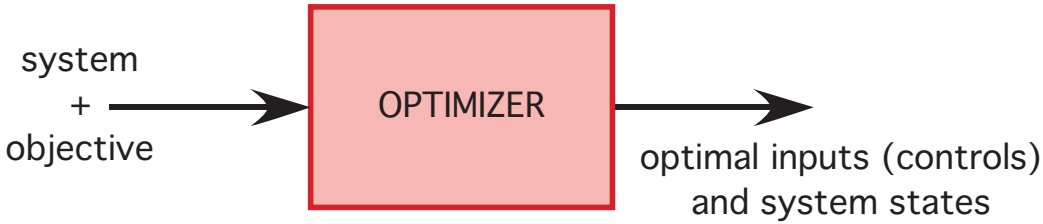
- Model calibration problems are a particular case of more general

  stochastic inverse, or parameter identification, or

  control, or optimization problems



Feedback control



Optimal control

OBSERVATIONS ABOUT THESE LECTURES

- Of greatest interest (to us) are nonlinear problems; however
  - so we focus on methods that are useful in the nonlinear setting
  - however, we do sometimes comment on special features of some methods that only hold for linear problems
- Both time-dependent and steady-state problems are of interest
  - for the sake of simplifying the exposition, we consider mostly steady-state problems
  - however, almost everthing we have to say applies equally well to time-dependent problems

WHITE NOISE

UNCORRELATED RANDOM FIELDS

- White noise refers to the case of uncorrelated random fields $\eta(\mathbf{x}, t; \omega)$ for which we have[†]

$$\mathsf{E}\big(\eta(\mathbf{x}, t; \omega)\big) = 0 \qquad \text{and} \qquad \mathsf{E}\big(\eta(\mathbf{x}, t; \omega)\eta(\mathbf{x}', t'; \omega)\big) = \delta(t - t')\delta(\mathbf{x} - \mathbf{x}')$$

  - at every point in space and at every instant in time, $\eta(\mathbf{x}, t; \omega)$ is independent and identically distributed

    - one determines $\eta(\mathbf{x}, t; \omega)$ at any point in space and any instant in time by sampling according to a given probability distribution

  - the Gaussian case is the one that often arises in practice (or because of a lack of information)

[†]The zero mean and unit variance assumptions are not restrictive

Discretizing white noise

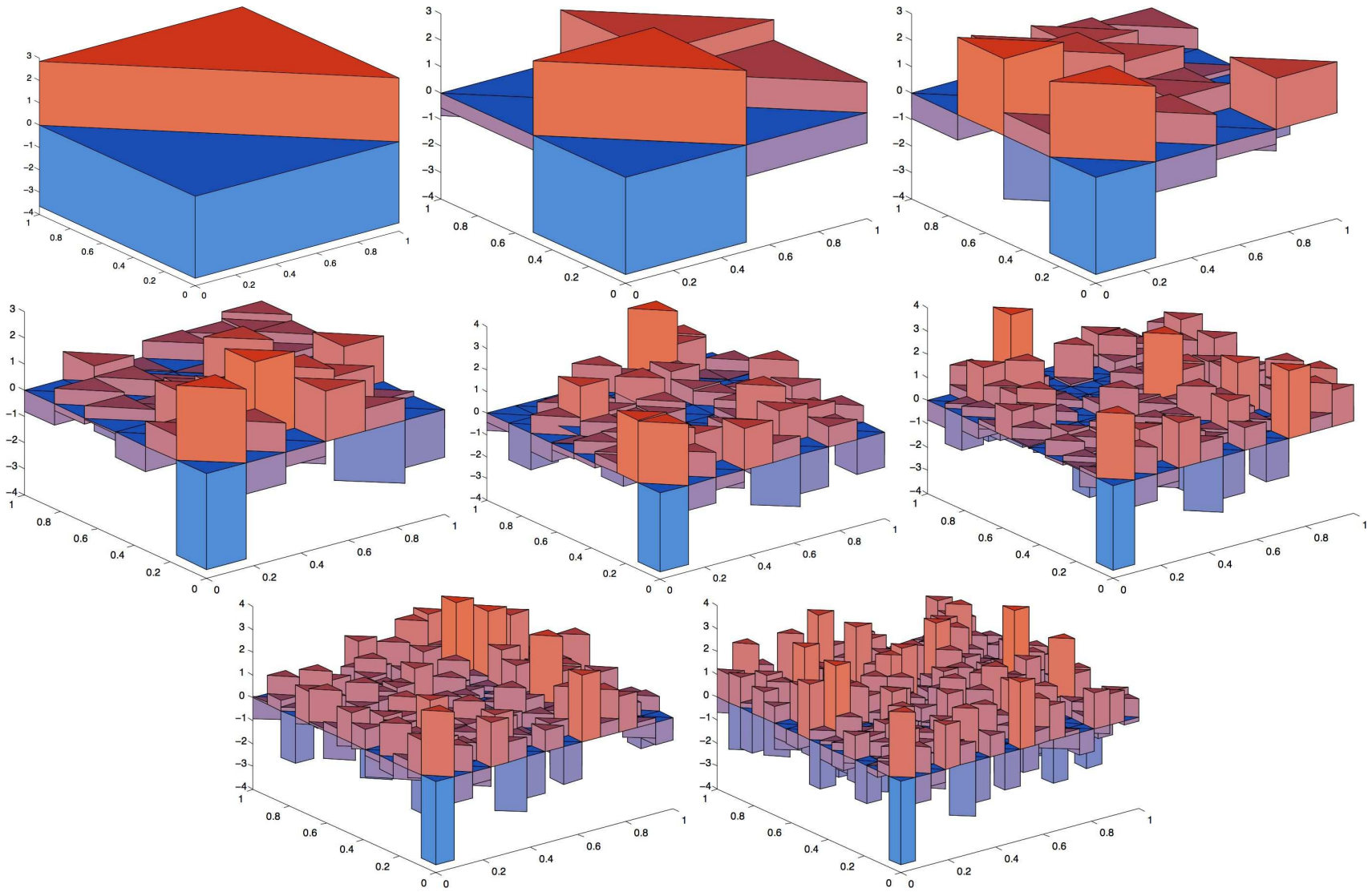
- In computer simulations, one cannot sample the Gaussian distribution at every point of the spatial domain and at every instant of time
  - white noise terms are replaced by discretized white noise terms
    - discretized white noise is more regular that white noise
- Among the means available for discretizing white noise, grid-based methods are the most popular

- To define a single realization of the discretized white noise, we
  - subdivide the spatial domain $\mathcal{D}$ into N_{space} subdomains
  - subdivide the temporal interval $[0, T]$ into N_{time} time subintervals
  - then, in the n_s-th spatial subdomain having volume V_{n_s} and in the n_t-th temporal subinterval having duration Δt_{n_t}, set

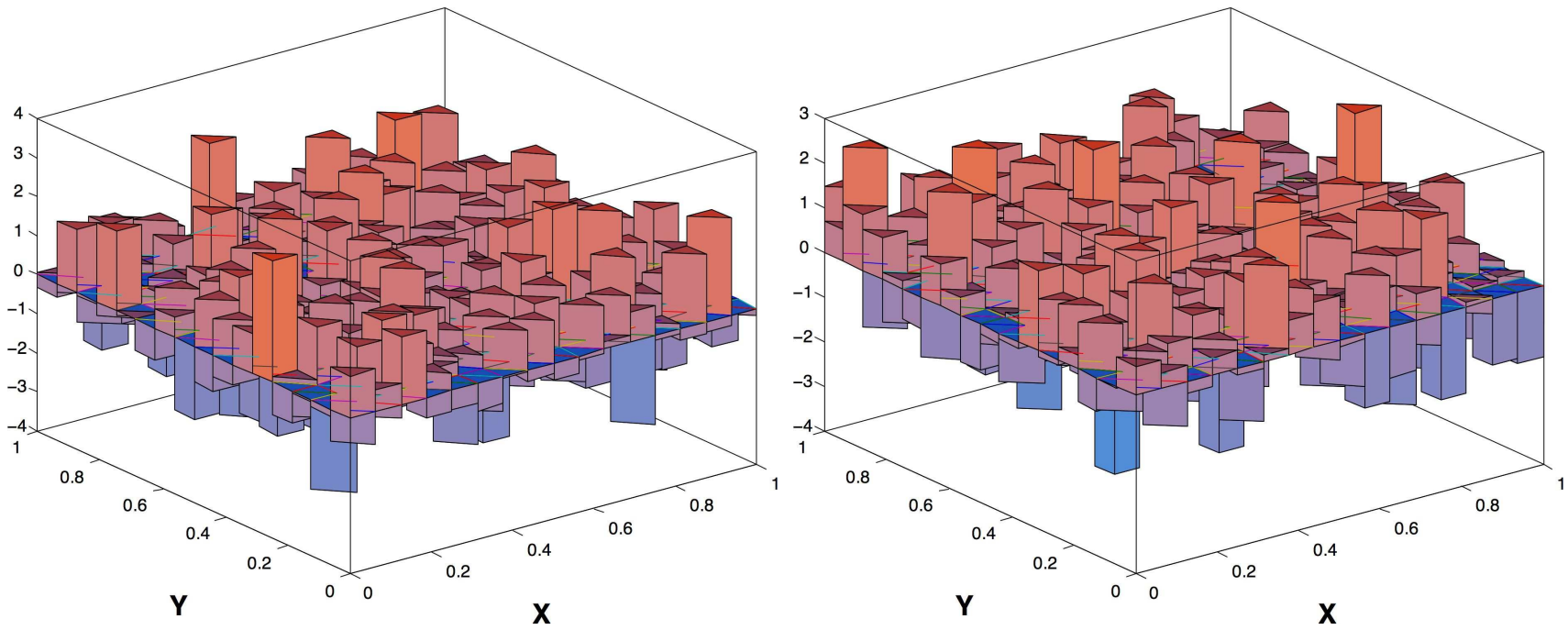
$$\eta_{approximate}(\mathbf{x}, t; \{y_{n_s,n_t}\}) = \frac{1}{\sqrt{\Delta t_{n_t}}\sqrt{V_{n_s}}} y_{n_s,n_t}$$

    where y_{n_s,n_t} are independent Gaussian samples having zero mean and unit variance

- Additional realizations are defined by resampling over the space-time grid



Realizations of discretized white noise at a same time interval in a square subdivided into 2, 8, 32, 72, 238, 242, 338, and 512 triangles



Realizations of discretized white noise at two different time intervals in a square subdivided into the same number of triangles

- Thus, the discretized white noise is piecewise constant in space and time

- Note that the piecewise constant function is much smoother than the random field it approximates

- It can be shown that

$$\lim_{N_{space}\to\infty,\, N_{time}\to\infty} \mathsf{E}\Big(\eta_{approximate}(\mathbf{x},t;\{y_{n_s,n_t}\})\,\eta_{approximate}(\mathbf{x}',t';\{y_{n_s,n_t}\})\Big)$$

$$= \mathsf{E}\big(\eta(\mathbf{x},t)\eta(\mathbf{x}',t')\big) = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t')$$

- The white noise case has been reduced to a case of a large but finite number of parameters

  – we have the

  $N = N_{space}N_{time}$ parameters y_{n_s,n_t}

  where $n_s = 1,\ldots,N_{space}$ and $n_t = 1,\ldots,N_{time}$

  – if we refine the spatial grid and/or reduce the time step, the number of parameters increases

PDE'S FORCED BY WHITE NOISE

- Formally, we can write an evolution equation with white noise forcing as

$$\frac{\partial u}{\partial t} = A(u; \mathbf{x}, t) + f(\mathbf{x}, t) + B(u; \mathbf{x}, t)\eta(\mathbf{x}, t; \omega) \qquad \text{in } \mathcal{D} \times (0, T]$$

where

A is a possibly nonlinear deterministic operator

f is a deterministic forcing function

B is a possibly nonlinear deterministic operator

η is the white noise forcing function

– among many other cases,

A, f, and B can take care of cases with means $\neq 0$ and variances $\neq 1$

- If B is independent of u, we have additive white noise

$$\frac{\partial u}{\partial t} = A(u; \mathbf{x}, t) + f(\mathbf{x}, t) + b(\mathbf{x}, t)\, \eta(\mathbf{x}, t)$$

  - in practice, often b is a constant

- If B depends on u, we have multiplicative white noise

  - of particular interest is the case of B linear in u

$$\frac{\partial u}{\partial t} = A(u; \mathbf{x}, t) + f(\mathbf{x}, t) + b(\mathbf{x}, t) u\, \eta(\mathbf{x}, t)$$

- Some observations
  - solutions are not sufficiently regular for the equations just written to make sense
    - the renowned Ito calculus is introduced to make sense of differential equations with white noise forcing
  - white noise need not be restricted to forcing terms in the PDE
    - in practice, it can also appear
      in the coefficients of the PDEs and boundary and initial conditions
      in the data in boundary and initial conditions
      in the definition of the domain

- Spatial discretization of the PDE can be effected via a finite element method based on a triangulation of the spatial domain $\mathcal{D}$; temporal discretization is effected via a finite difference method, e.g., a backward Euler method

  – it is natural to use the same grids in space and time as are used to discretize the white noise

  – thus, if one refines the finite element grid and the time step, one also refines the grid and time step for the white noise discretization

- Once a realization of the discretized noise is chosen,

  i.e., once one chooses the $N_{space}N_{time}$ Gaussian samples η_{n_s,n_t},

  a realization of the solution of the PDE is determined

  by solving a deterministic problem

- For example, consider the problem

$$\begin{cases} \dfrac{\partial u}{\partial t} = \Delta u + f(\mathbf{x},t) + b(\mathbf{x},t)u\,\eta(\mathbf{x},t;\omega) & \text{in } \mathcal{D}\times(0,T] \\ u = 0 & \text{in } \partial\mathcal{D}\times(0,T] \\ u(\mathbf{x},0) = u_0(\mathbf{x}) & \text{in } \mathcal{D} \end{cases}$$

  - subdivide $[0,T]$ into N_{time} subintervals of duration Δt_{n_t}, $n_t = 1,\ldots,N_{time}$

  - subdivide $\mathcal{D}$ into N_{space} finite elements $\{\mathcal{D}_{n_s}\}_{n_s=1}^{N_{space}}$

  - define a finite element space $S_0^h \subset H_0^1(\mathcal{D})$ with respect to the grid $\{\mathcal{D}_{n_s}\}_{n_s=1}^{N_{space}}$

– choose an approximation $u_{(0,h)}(\mathbf{x})$ to the initial data $u_0(\mathbf{x})$

– sample, from a standard Gaussian distribution,
the $N_{space}N_{time}$ values y_{n_s,n_t}, $n_s = 1, \ldots, N_{space}$ and $n_t = 1, \ldots, N_{time}$

– set $u_h^{(0)}(\mathbf{x}) = u_{(0,h)}(\mathbf{x})$

– then, for $n_t = 1, \ldots, N_{time}$, determine $u_h^{(n_t)}(\mathbf{x}) \in S_0^h$ from

$$\int_{\mathcal{D}} \frac{u_h^{(n_t)} - u_h^{(n_t-1)}}{\Delta t_{n_t}} v_h \, d\mathbf{x} + \int_{\mathcal{D}} \nabla u_h^{(n)} \cdot \nabla v_h \, d\mathbf{x}$$

$$= \int_{\mathcal{D}} f v_h \, d\mathbf{x} + \frac{1}{\sqrt{\Delta t_{n_t}}\sqrt{A_{n_s}}} \sum_{n_s=1}^{N_s} \int_{\mathcal{D}_{n_s}} y_{n_s,n_t} v_h \, d\mathbf{x} \qquad \text{for all } v^h \in S_0^h$$

- note that we have used a backward-Euler time stepping scheme

- This is a standard discrete finite element system for the heat equation, albeit with an unusual right-hand side

- Due to the lack of regularity of solutions of PDE's with white noise,

  the usual notions of convergence

  of the approximate solution to the exact solution

  do not hold,

  even in expectation

  – one has to be satisfied with very weak notions of convergence

COLORED NOISE

CORRELATED RANDOM FIELDS

- We now consider correlated random fields $\eta(\mathbf{x}, t; \omega)$
  - at each point $\mathbf{x}$ in a spatial domain $\overline{\mathcal{D}}$ and at each instant t in an time interval $[t_0, t_1]$, the value of η is determined by a random variable ω whose values are drawn from a given probability distribution
  - however, unlike the white noise case, the covariance function of the random field $\eta(\mathbf{x}, t; \omega)$ does not reduce to delta functions
- In rare cases, a formula for the random field is "known"
  - again, we cannot sample the random field at every spatial and temporal point
  - on the other hand, unlike the white noise case, the fact that the random field is correlated implies that one can find a discrete approximation to the random field for which the number of degrees of freedom can be thought of as fixed, i.e., independent of the spatial and temporal grid sizes

- More often, only the

  mean[†] $\mu_\eta(\mathbf{x}, t)$

  and

  covariance function $\text{cov}_\eta(\mathbf{x}, t; \mathbf{x}', t')$

  are known for points $\mathbf{x}$ and $\mathbf{x}'$ in $\overline{\mathcal{D}}$ and time instants t and t' in $[t_0, t_1]$

  – in this case, what we do not have is a formula for $\eta(\mathbf{x}, t; \omega)$

  – thus, we cannot evaluate $\eta(\mathbf{x}, t; \omega)$ when we need to

  – for example, if $\eta(\mathbf{x}, t; \omega)$ is a coefficient or a forcing function in a PDE, then to determine an approximate realization of the PDE we need to evaluate $\eta(\mathbf{x}, t; \omega)$ for a specific choice of ω and at specific points $\mathbf{x}$ and specific instants of time t used in the discretized PDE

[†] We have that

$$\mu_\eta(\mathbf{x}, t) = \mathsf{E}\big((\eta(\mathbf{x}, t; \cdot)\big)$$

and

$$\text{cov}_\eta(\mathbf{x}, t; \mathbf{x}', t') = \mathsf{E}\Big(\big(\eta(\mathbf{x}, t; \cdot) - \mu_\eta(\mathbf{x}, t)\big)\big(\eta(\mathbf{x}', t'; \cdot) - \mu_\eta(\mathbf{x}', t')\big)\Big)$$

- Examples of covariance functions

$$\text{cov}(\mathbf{x}, t; \mathbf{x}', t') = e^{-|\mathbf{x}-\mathbf{x}'|/L - |t-t'|/T}$$

and

$$\text{cov}(\mathbf{x}, t; \mathbf{x}', t') = e^{-|\mathbf{x}-\mathbf{x}'|^2/L^2 - |t-t'|^2/T^2}$$

where L is the correlation length and T is the correlation time

- large $L, T \Longrightarrow$ long-range order
- small $L, T \Longrightarrow$ short-range order

- Note that covariance functions are symmetric and positive

- So, we have two cases

  – the more common case for which only the mean and covariance function of the random field are known

    - we would like to find a simple formula depending on only a few parameters whose mean and covariance function are approximately the same as the given mean and covariance function

  – the rare case for which the random field is given as a formula but we want to approximate it

    - we would like to approximate it using few random parameters, certainly with a number of parameters that is independent of the spatial and temporal grid sizes

    - of course, this case can be turned into the first case by determining the mean and covariance function of the given random field (this may or may not be a good idea)

- Among the known ways for doing these tasks, we will focus on perhaps the most popular $\Longrightarrow$

  the Karhunen-Loève (KL) expansion of a random field $\eta(\mathbf{x}, t; \omega)$

  – given the mean and covariance of a random field $\eta(\mathbf{x}, t; \omega)$,

  - the KL expansion provides a simple formula that
    can be used whenever one needs a value $\eta(\mathbf{x}, t; \omega)$

  – to keep things simple, we discuss KL expansions
  for the case of spatially-dependent random fields

The Karhunen-Loève expansion

- Given the mean $\mu_\eta(\mathbf{x})$ and covariance $\text{cov}_\eta(\mathbf{x}, \mathbf{x}')$ of a random field $\eta(\mathbf{x}; \omega)$, determine the eigenpairs $\{\lambda_n, b_n(\mathbf{x})\}_{n=1}^{\infty}$ from the eigenvalue problem

$$\int_{\mathcal{D}} \text{cov}_\eta(\mathbf{x}, \mathbf{x}')\, b(\mathbf{x}')\, d\mathbf{x}' = \lambda b(\mathbf{x})$$

  - often in practice, an approximate version of this problem is solved, e.g., using a finite element method

  - due to the symmetry of $\text{cov}_\eta(\cdot;\cdot)$, the eigenvalues λ_n are real and the eigenfunctions $b_n(\mathbf{x})$ can be chosen to be real and orthonormal, i.e.,

$$\int_{\mathcal{D}} b_n(\mathbf{x})\, b_{n'}(\mathbf{x})\, d\mathbf{x} = \delta_{nn'}$$

  - due to the positivity of $\eta(\mathbf{x}; \omega)$, the eigenvalues are all positive

    - without loss of generality, they may be ordered in non-increasing order $\lambda_1 \geq \lambda_2 \geq \cdots$

- Then, the random field $\eta(\mathbf{x};\omega)$ admits the KL expansion[†]

$$\eta(\mathbf{x};\omega) = \mu_\eta(\mathbf{x}) + \sum_{n=1}^{\infty} \sqrt{\lambda_n}\, b_n(\mathbf{x})\, Y_n(\omega)$$

where $\{Y_n(\omega)\}_{n=1}^{\infty}$ are centered and uncorrelated random variables, i.e.,

$$\mathsf{E}\big(Y_n(\omega)\big) = 0 \qquad \mathsf{E}\big(Y_n(\omega) Y_{n'}(\omega)\big) = 0$$

that inherit the probability structure of the random field $\eta(\mathbf{x};\omega)$

  - e.g., if $\eta(\mathbf{x};\omega)$ is a Gaussian random field, then the Y_n's are all Gaussian random variables

[†] To see this, let us make the ansatz

$$\eta(\mathbf{x};\omega) = \mu_\eta(\mathbf{x}) + \sum_{n=1}^{\infty} \alpha_n b_n(\mathbf{x}) y_n(\omega)$$

where

$$\int_{\mathcal{D}} b_n(\mathbf{x}) b_{n'}(\mathbf{x})\, d\mathbf{x} = \delta_{nn'}, \qquad \mathsf{E}(y_n) = 0, \qquad \text{and} \qquad \mathsf{E}(y_n y_{n'}) = \delta_{nn'}$$

i.e., $\{b_n(\cdot)\}_{n=1}^{\infty}$ is a set of orthonormal functions and $\{y_n(\cdot)\}_{n=1}^{\infty}$ is a set of uncorrelated random variables; we then have that

$$\mathsf{E}(\eta) = \mu_\eta(\mathbf{x}) + \sum_{n=1}^{\infty} \alpha_n b_n(\mathbf{x}) \mathsf{E}(y_n) = \mu_\eta(\mathbf{x})$$

and

$$\mathsf{E}\Big(\big(\eta(\mathbf{x};\cdot) - \mu_\eta(\mathbf{x})\big) \big(\eta(\mathbf{x}';\cdot) - \mu_\eta(\mathbf{x}')\big) \Big) = \sum_{n=1}^{\infty} \sum_{n'=1}^{\infty} \alpha_n \alpha_{n'} b_n(\mathbf{x}) b_{n'}(\mathbf{x}') \mathsf{E}(y_n y_{n'}) = \sum_{n=1}^{\infty} \alpha_n^2 b_n(\mathbf{x}) b_n(\mathbf{x}')$$

so that

$$\mathsf{cov}_\eta(\mathbf{x},\mathbf{x}') = \sum_{n=1}^{\infty} \alpha_n^2 b_n(\mathbf{x}) b_n(\mathbf{x}');$$

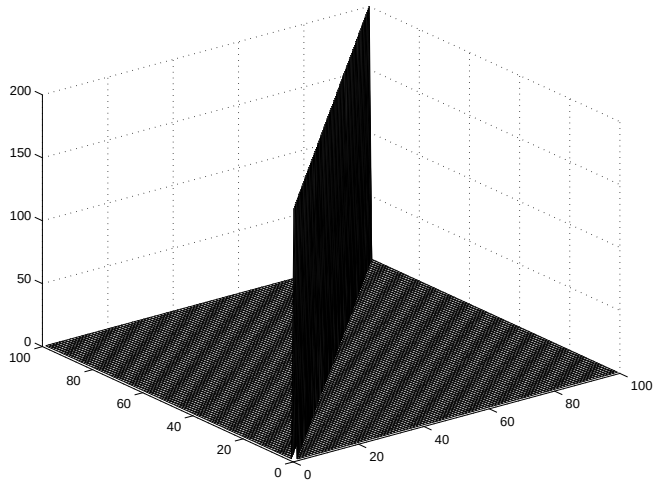
then, we have that

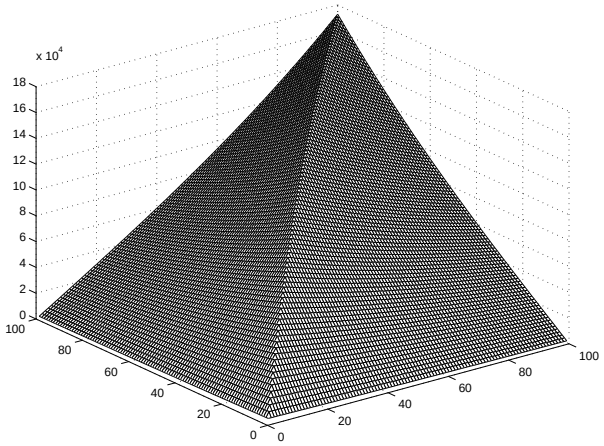
$$\int_{\mathcal{D}} \mathsf{cov}_\eta(\mathbf{x},\mathbf{x}') b_{n'}(\mathbf{x}')\, d\mathbf{x}' = \sum_{n=1}^{\infty} \alpha_n^2 b_n(\mathbf{x}) \int_{\mathcal{D}} b_n(\mathbf{x}') b_{n'}(\mathbf{x}')\, d\mathbf{x}' = \alpha_{n'}^2 b_{n'}(\mathbf{x})$$

so that indeed $\{\alpha_n^2, b_n(\mathbf{x})\}_{n=1}^{\infty}$ are the eigenpairs, i.e., we recover the KL expansion

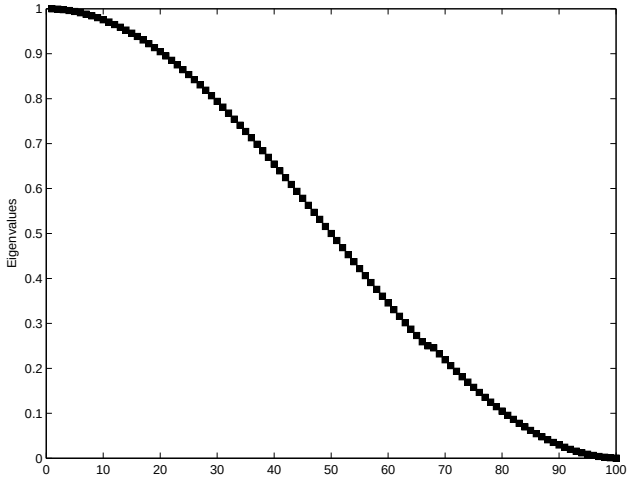
- The usefulness of the KL expansion results from the fact that the eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$ decay as n increases

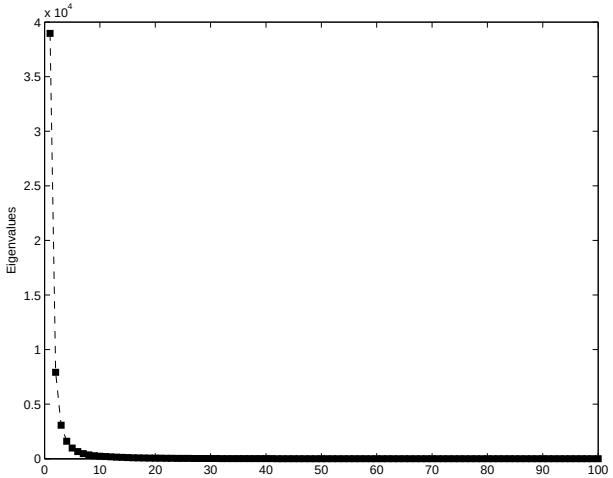
  – how fast they decay depends on the smoothness of the covariance function $\text{cov}_\eta(\mathbf{x}, \mathbf{x}')$ and on the correlation length L





Peaked and smooth covariance functions





Corresponding KL eigenvalues

- The decay of the eigenvalues implies that truncated KL expansions

$$\eta_N(\mathbf{x};\omega) = \mu(\mathbf{x}) + \sum_{n=1}^{N} \sqrt{\lambda_n} b_n(\mathbf{x}) Y_n(\omega)$$

can be accurate approximations to the exact expansions

– if one wishes for the relative error to be less than a prescribed tolerance δ, i.e., if one wants

$$\frac{\|\eta_N - \eta\|^2}{\|\eta\|^2} \leq \delta,$$

one should choose N to be the smallest integer such that

$$\frac{\sum_{n=N+1}^{\infty} \lambda_n}{\sum_{n=1}^{\infty} \lambda_n} \leq \delta \qquad \text{or, equivalently,} \qquad \frac{\sum_{n=1}^{N} \lambda_n}{\sum_{n=1}^{\infty} \lambda_n} \geq 1 - \delta$$

- Although the Y_n's are uncorrelated, in general they are not independent
  - in fact, they are independent if and only if they are (spherical) Gaussian
  - however, every random field can, in principle, be written as a function of a Gaussian random field
    - the inverse of the cumulative probability density of the given field,

    so that, in this way, we only have to deal with Gaussian random variables

- Dealing with independent random variables can have important practical consequences

- One important issue is the well posedness of the PDE when using a KL representation of random fields

  – suppose the coefficient $a(\mathbf{x};\omega)$ of an elliptic PDE is a random field

    - it cannot be a Gaussian random field since then it would admit negative values, which is not allowable

  – one way to get around this is to let, with $a_{min} > 0$,

$$a(\mathbf{x};\omega) = a_{min} + e^{\eta(\mathbf{x};\omega)}$$

  where $\eta(\mathbf{x};\omega)$ is a Gaussian random field with given mean and covariance

  – then, using a truncated KL expansion for $\eta(\mathbf{x};\omega)$, we have that

$$a(\mathbf{x};\omega) = a_{min} + e^{\mu(\mathbf{x}) + \sum_{n=1}^{N} \sqrt{\lambda_n}\, b_n(\mathbf{x})\, Y_n(\omega)}$$

  where $\{Y_n(\omega)\}_{n=1}^{N}$ are Gaussian random variables

Approximating Gaussian random fields

- For Gaussian random fields, we are done: we identify the random variables $\{Y_n(\omega)\}_{n=1}^N$ with Gaussian random parameters $\{y_n\}_{n=1}^N$ such that $\vec{y} \in \Gamma = \mathbb{R}^N$

- Let $\eta(\mathbf{x};\omega)$ be a Gaussian random field
  - we approximate $\eta(\mathbf{x};\omega)$ by its N-term truncated KL expansion

$$\eta_N(\mathbf{x};\omega) = \mu(\mathbf{x}) + \sum_{n=1}^{N} \sqrt{\lambda_n} b_n(\mathbf{x}) y_n$$

    where $\{y_n\}_{n=1}^N$ are Gaussian random parameters

- Thus, we now have a formula for the (approximation to a) random field that involves a finite number of random parameters
  - we can then use any of the methods to be discussed for problems involving a finite number of given random parameters to solve the problems described in terms of Gaussian random fields

Approximating non-Gaussian random fields

- If $\xi(\mathbf{x};\omega)$ is the given correlated random field and if the cumulative density function $F_\xi(\omega)$ is known, then one can write

$$\xi(\mathbf{x};\omega) = F_\xi^{-1}\big(\eta(\mathbf{x};\omega)\big) \qquad \text{where } \eta(\mathbf{x};\omega) \text{ is a Gaussian random field}$$

  - then, one can approximate $\eta(\mathbf{x};\omega)$ using a truncated KL expansion in terms of Gaussian random parameters $\{y_n\}_{i=1}^N$ so that

$$\xi_N(\mathbf{x};\omega) = F_\xi^{-1}\big((\eta_N(\mathbf{x};\omega)\big) = F_\xi^{-1}\Big(\mu(\mathbf{x}) + \sum_{n=1}^{N} \sqrt{\lambda_n} b_n(\mathbf{x}) y_n\Big)$$

  - so, again, we have obtained a formula for an approximation of the general random field $\xi(\mathbf{x};\omega)$ in terms of N random Gaussian parameters so that we can use any of the methods to be discussed for the random parameters case to find approximate solutions of the stochastic PDE

RANDOM PARAMETERS

PDE'S with random inputs depending on random parameters

- One or more

  input functions,

  e.g., coefficients, forcing terms, initial data, etc. in a

  partial differential equation

  depend on a finite number of random parameters

  - the input function could also depend on space and time
  - the random parameters could come from a Karhunen-Loève expansion of a correlated random field
  - the random parameters could appear naturally in the definition of input function

    - e.g., the Young's modulus or a diffusivity coefficient could be random

- Ideally, we would know the probability density function (PDF) for each parameter
  - as has already been mentioned, in practice, we know very little about the statistics of input parameters
  - however, we will assume that we know the PDFs for all the random input parameters

- Example: a nonlinear parabolic equation

$$c(\mathbf{x},t;y_{N_b},\ldots,y_{N_c})\frac{\partial u}{\partial t}-\nabla\cdot\Big(a(\mathbf{x},t;y_1,\ldots,y_{N_a})\nabla u\Big)+b(\mathbf{x},t;y_{N_a+1},\ldots,y_{N_b})u^3$$
$$=f(\mathbf{x},t;y_{N_c+1},\ldots,y_{N_f})\quad\text{on }\mathcal{D}(y_{N_i+1},\ldots,y_{N_g};y_{N_g+1},\ldots,y_{N_h})$$

$$u=f_{dir}(\mathbf{x},t;y_{N_f+1},\ldots,y_{N_d})\quad\text{on }\partial\mathcal{D}_D(y_{N_i+1},\ldots,y_{N_g})$$

$$a(\mathbf{x},t;y_1,\ldots,y_{N_a})\frac{\partial u}{\partial n}=f_{neu}(\mathbf{x},t;y_{N_d+1},\ldots,y_{N_e})\quad\text{on }\partial\mathcal{D}_N(y_{N_g+1},\ldots,y_{N_h})$$

$$u=f_0(\mathbf{x};y_{N_e+1},\ldots,y_{N_i})\quad\text{on }\mathcal{D}(y_{N_i+1},\ldots,y_{N_g};y_{N_g+1},\ldots,y_N)$$

  - the y_n's are random parameters
  - a, b, c, f, f_{dir}, f_{neu}, and f_0 are given functions of $\mathbf{x}$, t, and the random parameters
  - the boundary segments $\partial\mathcal{D}_D$ and $\partial\mathcal{D}_N$ are parametrized by the corresponding random parameters
  - of course, ∇ and $\nabla\cdot$ are operators involving spatial derivatives

- Concrete example: an elliptic PDE for $u(\mathbf{x}; y_1, \ldots, y_5)$

  – consider

$$\nabla \cdot \Big(a(\mathbf{x}; y_1, y_2) \nabla u \Big) = f(\mathbf{x}; y_3, y_4) \quad \text{on } \mathcal{D}(y_5)$$

$$u = 0 \quad \text{on } \partial\mathcal{D}(y_5)$$

where

$$a(\mathbf{x}; y_1, y_2) = 3 + |\mathbf{x}| \big(y_1^2 + \sin(y_2) \big)$$

$$f(\mathbf{x}; y_3, y_4) = y_3 e^{-y_4 |\mathbf{x}|^2}$$

$$\mathcal{D}(y_5) = (0, 1) \times (0, 1 + 0.3 y_5)$$

with

$$\rho_1(y_1) = N(0; 1) \qquad \rho_2(y_2) = U(0; 0.5\pi) \qquad \rho_3(y_3) = N(0; 2)$$

$$\rho_4(y_4) = U(0, 1) \qquad \rho_5(y_5) = U(-1, 1)$$

- The well-posedness of the PDE for all possible values of the parameters is a very important (and sometimes ignored) consideration

  – for the simple elliptic PDE

$$\nabla \cdot \Big(a(\mathbf{x}; y_1, \ldots, y_N) \nabla u \Big) = f(\mathbf{x}) \quad \text{on } \mathcal{D}$$

  we must have, for some $a_{max} \geq a_{min} > 0$,

$$a_{min} \leq a(\mathbf{x}; y_1, \ldots, y_{\widetilde{N}}) \leq a_{max} \qquad \text{for all } \mathbf{x} \in \mathcal{D} \text{ and all } \vec{y} \in \Gamma$$

  – this could place a constraint on how one chooses the PDF for the parameters

  – for example, if we have

$$a(\mathbf{x}; y) = a_0 + y$$

  where $a_0 > 0$, we cannot choose y to be a Gaussian random parameter

A brief taxonomy of methods for stochastic PDEs with random input parameters

- Stochastic finite element methods (SFEMs)

  $\Longrightarrow$ methods for which spatial discretization is effected using finite element methods[†]

- One particular class of SFEMs is known as stochastic Galerkin methods (SGMs)

  $\Longrightarrow$ methods for which probabilistic discretization is also effected using a Galerkin method

  – polynomial chaos and generalized polynomial chaos methods are SGMs

  – we will also consider other SGMs

[†] Throughout, we assume that spatial discretization is effected using finite element methods; most of what we say also holds for other spatial discretization approaches, e.g., finite differences, finite volumes, spectral, etc.

- Another class of SFEMs are stochastic sampling methods (SSMs)

  $\Longrightarrow$ points in the parameter domain Γ are sampled,
  then used as inputs for the PDE, and then
  ensemble averages of output quantities of interest are computed

  – Monte-Carlo finite element methods are the simplest SSMs

  – stochastic collocation methods (SCMs) are also SSMs

  - the sampling points are the quadrature points corresponding to some quadrature rule

Example used to describe numerical methods for SPDEs

- Let $\mathcal{D} \subset \mathbb{R}^d$ denote a spatial domain[†] with boundary $\partial\mathcal{D}$
  - $d = 1, 2$, or 3 denotes the spatial dimension
  - $\mathbf{x} \in \mathcal{D}$ denotes the spatial variable

- Let $\Gamma \in \mathbb{R}^N$ denote a parameter domain
  - N denotes the number of parameters
  - $\vec{y} = (y_1, y_2, \ldots, y_N) \in \Gamma$ denotes the random parameter vector
  - note that we have a finite number of parameters $\{y_n\}_{n=1}^N$
    but they can take on values anywhere in the Euclidean domain Γ

[†]For the sake of simplicity, we now consider stationary problems; all we have to say holds equally well for time-dependent problems

- Let $u(\mathbf{x};\vec{y}) \in X \times Z$ denote the solution of the SPDE[†‡]
  - generally, $Z = L^q_\rho(\Gamma)$, the space of functions of N variables whose q-th power is integrable with respect to the joint PDF (the weight function) $\rho(\cdot)$, i.e., those functions $g(\vec{y})$ for which

$$\int_\Gamma |g(\vec{y})|^q \rho(\vec{y})\, d\vec{y} < \infty$$

    - q is chosen according to how many statistical moments one wants to have well defined
    - the most common choice is $q = 2$ so that up to the second moments are well defined
    - if $\{y_1, \ldots, y_N\}$ are independent and if $L^q_{\rho_n}(\Gamma_n)$ denotes the space of functions that have integrable q-th powers with respect to the PDF $\rho_n(y_n)$, we have that

$$L^q_\rho(\Gamma) = L^q_{\rho_1}(\Gamma_1) \otimes L^q_{\rho_2}(\Gamma_2) \otimes \cdots \otimes L^q_{\rho_N}(\Gamma_N)$$

[†]Often, X is a Sobolev space such as $H^1_0(\mathcal{D})$

[‡]It is not always convenient to use a product space $X \times Z$; for example, it may make more sense to have $u \in L^q_\rho(\Gamma; X)$

- It is entirely natural to then treat a function $u(\mathbf{x};\vec{y})$ of d spatial variables and of N random parameters as a function of $d+N$ variables

- This leads one to consider a Galerkin weak formulation in physical and parameter space: seek $u(\mathbf{x};\vec{y}) \in X \times Z$

$$\int_\Gamma \int_\mathcal{D} S(u;\vec{y})T(v)\rho(\vec{y})\,d\mathbf{x}d\vec{y} = \int_\Gamma \int_\mathcal{D} vf(\vec{y})\rho(\vec{y})\,d\mathbf{x}d\vec{y} \qquad \forall\, v \in X \times Z$$

where[†]

  - $S(\cdot;\cdot)$ is, in general, a nonlinear operator[‡]
  - $T(\cdot)$ is a linear operator

[†]Of course, if $\mathrm{E}(\cdot)$ denotes the expected value, this may be expressed in the form

$$\mathrm{E}\left(\int_\mathcal{D} S(u;\vec{y})T(v)\rho(\vec{y})\,d\mathbf{x} - \int_\mathcal{D} vf(\vec{y})\rho(\vec{y})\,d\mathbf{x}\right) = 0$$

[‡]S, T, and f could also depend on $\mathbf{x}$, but we do not explicitly keep track of such dependences

- In general, we would have a sum of such terms, i.e., we would have that

$$\sum_{m=1}^{M} \int_{\Gamma} \int_{\mathcal{D}} S_m(u;\vec{y}) T_m(v) \rho(\vec{y})\, d\mathbf{x} d\vec{y}$$

$$= \int_{\Gamma} \int_{\mathcal{D}} v f(\vec{y}) \rho(\vec{y})\, d\mathbf{x} d\vec{y} \qquad \forall\, v \in X \times Z$$

  - however, without loss of generality, it suffices for our purposes to consider the simpler single-term form

$$\int_{\Gamma} \int_{\mathcal{D}} S(u;\vec{y}) T(v) \rho(\vec{y})\, d\mathbf{x} d\vec{y} = \int_{\Gamma} \int_{\mathcal{D}} v f(\vec{y}) \rho(\vec{y})\, d\mathbf{x} d\vec{y} \qquad \forall\, v \in X \times Z$$

- In general,
  - both S and T could involve derivatives with respect to $\mathbf{x}$
  - but S does not involve derivatives with respect to $\vec{y}$

- Example

  – suppose our SPDE problem is given by

$$-\nabla \cdot \Big(a(\vec{y})\nabla u\Big) + c(\vec{y})u^3 = f(\vec{y}) \qquad \text{in } \mathcal{D} \qquad \text{and} \qquad u = 0 \qquad \text{in } \partial\mathcal{D}$$

    - of course, a, c, and f could also depend on $\mathbf{x}$

  – we then have that $X = H_0^1(\mathcal{D})$ and $Z = L_\rho^2(\Gamma)$ and the weak formulation:

    - seek $u(\mathbf{x}; \vec{y}) \in H_0^1(\mathcal{D}) \times L_\rho^2(\Gamma)$ such that

$$\int_{\mathcal{D}} \int_{\Gamma} \Big(a(\vec{y})\nabla u\Big) \cdot \nabla v \rho(\vec{y})\, d\vec{y} d\mathbf{x} + \int_{\mathcal{D}} \int_{\Gamma} \Big(c(\vec{y})u^3\Big) v \rho(\vec{y})\, d\vec{y} d\mathbf{x}$$
$$= \int_{\mathcal{D}} \int_{\Gamma} f(\vec{y}) v \rho(\vec{y})\, d\vec{y} d\mathbf{x} \qquad \forall\, v \in H_0^1(\mathcal{D}) \times L_\rho^2(\Gamma)$$

  – in the first term, we have that $S(u, \vec{y}) = a(\vec{y})\nabla u$ and $T = \nabla v$

  – in the second term, we have that $S(u, \vec{y}) = c(\vec{y})u^3$ and $T = v$

- We assume that all methods considered use the same approach to effect discretization with respect to the spatial variables

  – we focus on finite element methods,

    i.e., on stochastic finite element methods

  – throughout, $\{\phi_j(\mathbf{x})\}_{j=1}^{J}$ denotes a basis for the finite element space $X_J \subset X$ used to effect spatial discretization

    - note that J denotes the dimension of the finite element space

- We assume that Γ is a parameter box

  - without loss of generality, it can be taken to be a hypercube in $\mathbb{R}^N$

  - for parameters with unbounded PDFs, Γ can be of infinite extent

  - if the parameters are constrained, Γ need not be so simple

    e.g., if y_1 and y_2 are independent except that we require that $y_1^2 + y_2^2 \leq 1$, then Γ would be the unit circle