OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Uncertainty quantification & approximation theory for parameterized (stochastic) PDEs

Part I: Background, motivation, and "tools of the trade"

Clayton G. Webster†⋆

Hoang Tran⋆, Guannan Zhang⋆

†Department of Mathematics, University of Tennessee

⋆Department of Computational & Applied Mathematics (CAM)
Oak Ridge National Laboratory

Supporting agencies: **DOE** (ASCR, BES), **DOD** (AFOSR, DARPA), **NSF** (CM)

`equinox.ornl.gov` `formulate.ornl.gov` `tasmanian.ornl.gov`

Summary of the course

Uncertainty quantification & approximation theory for parameterized (stochastic) PDEs

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

DAY 1

1. Background, motivation, and "tools of the trade": Clayton Webster
2. Methods and algorithms: Clayton Webster

DAY 2

3. Orthogonal polynomials and best approximation: Hoang Tran
4. Discrete least squares and compressed sensing techniques: Hoang Tran

DAY 3

5. Sparse grid interpolation via **global** Lagrange polynomials: Clayton Webster
6. Sparse grids interpolation via **local** hierarchical polynomials: Guannan Zhang

DAY 4

7. Multilevel sampling and interpolation methods: Guannan Zhang
8. Other topics, opportunities, and open discussion: Tran, Webster, and Zhang

Part I Outline

Background, motivation, and "tools of the trade"

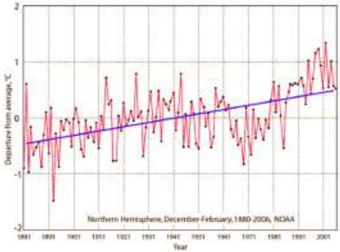
1. Why parameterized (stochastic) models?
2. An overview of uncertainty quantification (UQ)
3. Stochastic models
4. Tools of the trade: random variables
5. Tools of the trade: random processes and random fields
6. SVD: Discrete version of the Karhunen-Loève expansion
7. The Karhunen-Loève expansion
8. Summary of Part I

Why parameterized (stochastic) models?

A transition to non-deterministic simulations

Many applications (especially those predicting future events) are affected by a relatively large amount of uncertainty in the input data such as model coefficients, forcing terms, boundary conditions, geometry, etc.

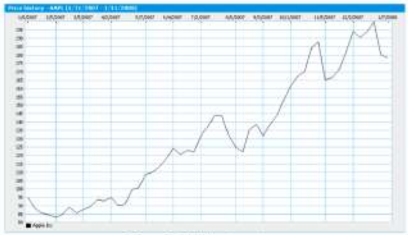
1. Predicting future climate changes



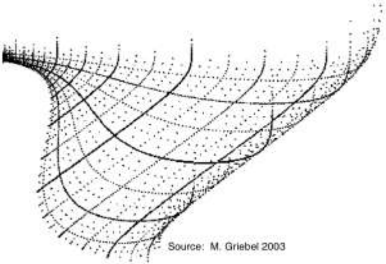
Global warming?

Examples

2. Forecasting financial markets



Source: Fidelity Investments



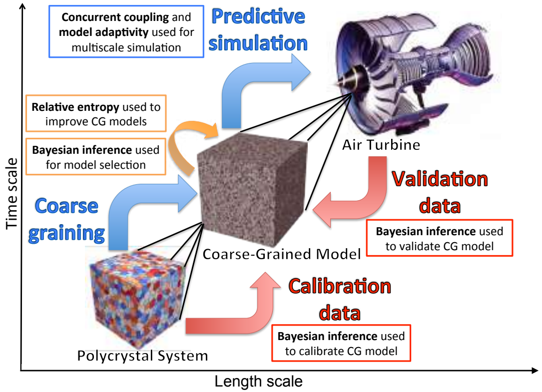
Source: M. Griebel 2003

The amount of uncertainty may depend on the number of: economic factors, underlying assets, or the number of time points/time steps, as well as human behaviors, etc.

Examples

formulate.ornl.gov

❸ Modeling and predicting the behavior of large-scale engineered systems



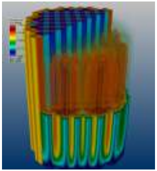
- The model itself may contain an incomplete description of parameters, processes or fields (not possible or too costly to measure).
- There may be small, unresolved scales in the model that act as a kind of background noise (i.e. macro behavior from micro structure).

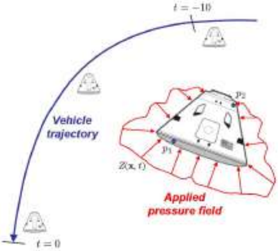
Additional examples

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

UQ examples of international importance:

- Enhancement of reliability of smart energy grids.
- Development of renewable energy technologies.
- Vulnerability analysis of water and power supplies.
- Understanding complex biological networks.
- Modeling unwanted vibrations re-entry vehicles experience when pierce the earths atmosphere.
- Design and licensing of current and future nuclear energy reactors.





Types of uncertainties

Stochastic models give quantitative information about uncertainty. In practice it is necessary to address the following types of uncertainties:

1. Uncertainty may be aleatoric which means random and is due to the intrinsic variability in the system.

   Remark: by variability we mean a type of uncertainty that is inherent and irreducible, e.g., turbulent fluctuations of a flow field around an airplane wing, permeability in an aquifer, etc.

   OR

2. Uncertainty may be epistemic which means due to incomplete knowledge.

   Remark: can be reduced by additional experimentation, improvements in measuring devices, etc., e.g., mechanic properties of many bio-materials, polymeric fluids, highly heterogeneous or composite materials, the action of wind or seismic vibrations on civil structures, etc.

Uncertainty quantification (UQ) attempts to quantitatively access the impact of uncertain data on simulation outputs.

Types of uncertainties

Stochastic models give quantitative information about uncertainty. In practice it is necessary to address the following types of uncertainties:

1. Uncertainty may be aleatoric which means random and is due to the intrinsic variability in the system.

   Remark: by variability we mean a type of uncertainty that is inherent and irreducible, e.g., turbulent fluctuations of a flow field around an airplane wing, permeability in an aquifer, etc.

   OR

2. Uncertainty may be epistemic which means due to incomplete knowledge.

   Remark: can be reduced by additional experimentation, improvements in measuring devices, etc., e.g., mechanic properties of many bio-materials, polymeric fluids, highly heterogeneous or composite materials, the action of wind or seismic vibrations on civil structures, etc.

Uncertainty quantification (UQ) attempts to quantitatively access the impact of uncertain data on simulation outputs.

Types of uncertainties

Stochastic models give quantitative information about uncertainty. In practice it is necessary to address the following types of uncertainties:

1. Uncertainty may be aleatoric which means random and is due to the intrinsic variability in the system.

   Remark: by variability we mean a type of uncertainty that is inherent and irreducible, e.g., turbulent fluctuations of a flow field around an airplane wing, permeability in an aquifer, etc.

   OR

2. Uncertainty may be epistemic which means due to incomplete knowledge.

   Remark: can be reduced by additional experimentation, improvements in measuring devices, etc., e.g., mechanic properties of many bio-materials, polymeric fluids, highly heterogeneous or composite materials, the action of wind or seismic vibrations on civil structures, etc.

Uncertainty quantification (UQ) attempts to quantitatively access the impact of uncertain data on simulation outputs.

Uncertainty quantification (UQ)

Worst scenario approaches

Let A denote the input data and B the output set such that $u : A \mapsto B$. There are various ways to describe the uncertainty in A with the **goal** of describing the uncertainty in some quantity of interest (QoI) denoted $Q(u)$:

Worst scenario approaches: Typically A is an ϵ-ball around some nominal input data and the goal is to determine the worst case associated with the set relation $B = u(A)$

The range of the uncertainty of $Q(u)$ is then defined by the interval I

$$I := [\underline{Q}(u), \overline{Q}(u)] = \left[\inf_{a \in A} Q(u(a)), \sup_{a \in A} Q(u(a)) \right]$$

- the choice of the input set A is, in a large way, subjective and should be regarded as a working assumption.

"Assumption has many advantages. Chiefly these are the same as those of theft over honest toil." – Bertrand Russel

Uncertainty quantification (UQ)

Worst scenario approaches

Let A denote the input data and B the output set such that $u : A \mapsto B$. There are various ways to describe the uncertainty in A with the **goal** of describing the uncertainty in some quantity of interest (QoI) denoted $Q(u)$:

Worst scenario approaches: Typically A is an ϵ-ball around some nominal input data and the goal is to determine the worst case associated with the set relation $B = u(A)$

The range of the uncertainty of $Q(u)$ is then defined by the interval I

$$I := [\underline{Q}(u), \overline{Q}(u)] = \left[\inf_{a \in A} Q(u(a)), \sup_{a \in A} Q(u(a)) \right]$$

- the choice of the input set A is, in a large way, subjective and should be regarded as a working assumption.

  "Assumption has many advantages. Chiefly these are the same as those of theft over honest toil." – Bertrand Russel

Uncertainty quantification (UQ)

Knowledge-based methods

Fuzzy sets and possibility theory: deterministic approach to UQ which generalizes classical set theory. Let $C \subset A$:

- for each $x \in A$ set membership is defined by $\mu_C : A \to [0,1]$, expressing the degree of truth of the statement "x belongs to C."
- define the α-cut of C by $C_\alpha \stackrel{\text{def}}{=} \{x \in A : \mu_C \geq \alpha\}$ which gives a set characterization of uncertainty
- the operator u then propagates the fuzziness in A into the fuzziness in B

Evidence theory (Dempster-Shafer Theory): generalizing the probabilistic approach by defining the Belief $Bel(C)$ (lower bound) and Plausibility $Pl(C)$ (upper bound) functions, for the likelihood of an event C.

- define the likelihood $m : \Phi \to [0,1]$ of a countable family of events Ψ

$$\sum_{\varphi \in \Phi} m(\varphi) = 1, \quad m(\emptyset) = 0, \quad \text{however } \varphi_1 \subset \varphi_2 \not\Rightarrow m(\varphi_1) \leq m(\varphi_2)$$

$$Bel(C) = \sum_{\varphi \in \Phi, \varphi \subset C} m(\varphi), \quad Pl(C) = \sum_{\varphi \in \Phi, \varphi \cap C \neq \emptyset} m(\varphi)$$

Uncertainty quantification (UQ)

Knowledge-based methods

Fuzzy sets and possibility theory: deterministic approach to UQ which generalizes classical set theory. Let $C \subset A$:

- for each $x \in A$ set membership is defined by $\mu_C : A \to [0,1]$, expressing the degree of truth of the statement "x belongs to C."
- define the α-cut of C by $C_\alpha \stackrel{\text{def}}{=} \{x \in A : \mu_C \geq \alpha\}$ which gives a set characterization of uncertainty
- the operator u then propagates the fuzziness in A into the fuzziness in B

Evidence theory (Dempster-Shafer Theory): generalizing the probabilistic approach by defining the Belief $Bel(C)$ (lower bound) and Plausibility $Pl(C)$ (upper bound) functions, for the likelihood of an event C.

- define the likelihood $m : \Phi \to [0,1]$ of a countable family of events Ψ

$$\sum_{\varphi\in\Phi} m(\varphi) = 1, \quad m(\emptyset) = 0, \quad \text{however } \varphi_1 \subset \varphi_2 \not\Rightarrow m(\varphi_1) \leq m(\varphi_2)$$

$$Bel(C) = \sum_{\varphi\in\Phi, \varphi\subset C} m(\varphi), \quad Pl(C) = \sum_{\varphi\in\Phi, \varphi\cap C\neq\emptyset} m(\varphi)$$

Uncertainty quantification (UQ)

Stochastic/Probabilistic methods

What is probability theory? Understood as a mathematical theory of a finite measure.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ denote a (complete) probability space: Ω is the event space, $\mathscr{F} \subset 2^{\Omega}$ is the σ-algebra and $\mathbb{P}$ is the probability measure, satisfying:

1. $0 \leq \mathbb{P}(A)$, if $A \in \mathscr{F}$ and $\mathbb{P}(\Omega) = 1$
2. A **positive measure** $\mathbb{P} : \mathscr{F} \to [0,1]$ which is countably additive, i.e.,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i), \text{ for } \{A_i\}_{i=1}^{\infty} \in \mathscr{F} \text{ disjoint}$$

Stochastic / probabilistic methods: given a probability measure on the input data A the mapping u induces a probability measure on the output set $B \implies$ SODEs/SPDEs (Doob-Dynkin Lemma)

- applies to aleatoric phenomena, i.e., frequencies of occurance.
- applies to epistemic concepts in the realm of Bayesian and maximum entropy methods.

Uncertainty quantification (UQ)

Stochastic/Probabilistic methods

What is probability theory? Understood as a mathematical theory of a finite measure.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ denote a (complete) probability space: Ω is the event space, $\mathscr{F} \subset 2^{\Omega}$ is the σ-algebra and $\mathbb{P}$ is the probability measure, satisfying:

1. $0 \leq \mathbb{P}(A)$, if $A \in \mathscr{F}$ and $\mathbb{P}(\Omega) = 1$
2. A **positive measure** $\mathbb{P} : \mathscr{F} \to [0, 1]$ which is countably additive, i.e.,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i), \text{ for } \{A_i\}_{i=1}^{\infty} \in \mathscr{F} \text{ disjoint}$$

Stochastic / probabilistic methods: given a probability measure on the input data A the mapping u induces a probability measure on the output set $B \Longrightarrow$ SODEs/SPDEs (Doob-Dynkin Lemma)

- applies to aleatoric phenomena, i.e., frequencies of occurance.
- applies to epistemic concepts in the realm of Bayesian and maximum entropy methods.

Uncertainty quantification (UQ)

Stochastic/Probabilistic methods

What is probability theory? Understood as a mathematical theory of a finite measure.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ denote a (complete) probability space: Ω is the event space, $\mathscr{F} \subset 2^\Omega$ is the σ-algebra and $\mathbb{P}$ is the probability measure, satisfying:

1. $0 \leq \mathbb{P}(A)$, if $A \in \mathscr{F}$ and $\mathbb{P}(\Omega) = 1$
2. A **positive measure** $\mathbb{P} : \mathscr{F} \to [0, 1]$ which is countably additive, i.e.,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i), \text{ for } \{A_i\}_{i=1}^{\infty} \in \mathscr{F} \text{ disjoint}$$

Stochastic / probabilistic methods: given a probability measure on the input data A the mapping u induces a probability measure on the output set $B \Longrightarrow$ SODEs/SPDEs (Doob-Dynkin Lemma)

- applies to aleatoric phenomena, i.e., frequencies of occurance.
- applies to epistemic concepts in the realm of Bayesian and maximum entropy methods.

Uncertainty quantification (UQ)

Stochastic/Probabilistic methods

What is probability theory? Understood as a mathematical theory of a finite measure.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ denote a (complete) probability space: Ω is the event space, $\mathscr{F} \subset 2^\Omega$ is the σ-algebra and $\mathbb{P}$ is the probability measure, satisfying:

1. $0 \leq \mathbb{P}(A)$, if $A \in \mathscr{F}$ and $\mathbb{P}(\Omega) = 1$
2. A **positive measure** $\mathbb{P} : \mathscr{F} \to [0, 1]$ which is countably additive, i.e.,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i), \text{ for } \{A_i\}_{i=1}^{\infty} \in \mathscr{F} \text{ disjoint}$$

Stochastic / probabilistic methods: given a probability measure on the input data A the mapping u induces a probability measure on the output set $B \Longrightarrow$ SODEs/SPDEs (Doob-Dynkin Lemma)

- applies to aleatoric phenomena, i.e., frequencies of occurance.
- applies to epistemic concepts in the realm of Bayesian and maximum entropy methods.

Stochastic models

Abstract setting

Models (linear or nonlinear) for a system $\mathcal{F}$ may be stationary with state u, exterior loading or forcing f and random model description (realization) $\omega \in \Omega$, with probability measure $\mathbb{P}$:

$$\mathcal{F}(\omega)[u(x)] = f(x, \omega) \quad \text{for a.e. } x \in D \subset \mathbb{R}^n.$$

Evolution in time may be:

- discrete (e.g. Markov chain), driven by discrete random process:

$$u_{n+1} = \mathcal{F}(\omega)[u_n]$$

- continuous (e.g. Markov process ≡ Stochastic Differential Equation), driven by random processes:

$$du = (\mathcal{F}(\omega)[u] - f(\omega, x, t))\, dt + \mathcal{B}(\omega)[u]\, dW(\omega, t) + \mathcal{P}(\omega)[u]\, dQ(\omega, t).$$

In this Itô evolution equation, $W(\omega, t)$ is a Wiener process and $Q(\omega, t)$ is a Poisson process.

Stochastic models

Abstract setting

Models (linear or nonlinear) for a system $\mathcal{F}$ may be stationary with state u, exterior loading or forcing f and random model description (realization) $\omega \in \Omega$, with probability measure $\mathbb{P}$:

$$\mathcal{F}(\omega)[u(x)] = f(x, \omega) \quad \text{for a.e. } x \in D \subset \mathbb{R}^n.$$

Evolution in time may be:

- discrete (e.g. Markov chain), driven by discrete random process:

$$u_{n+1} = \mathcal{F}(\omega)[u_n]$$

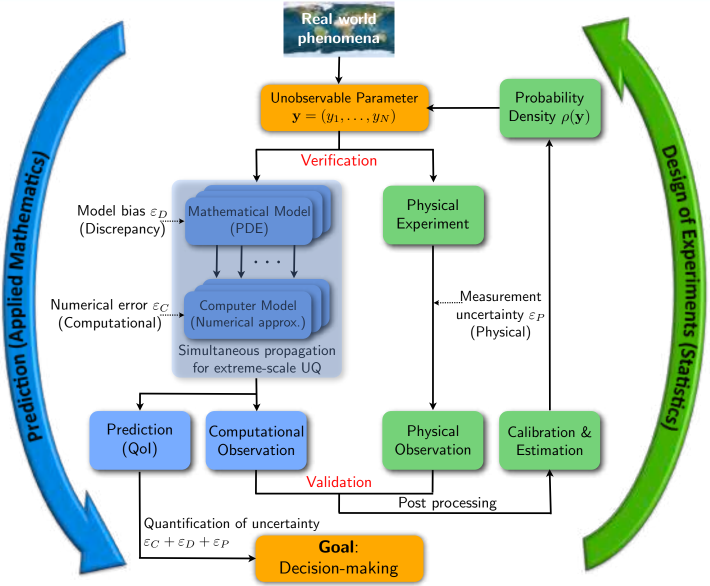
- continuous (e.g. Markov process $\equiv$ Stochastic Differential Equation), driven by random processes:

$$du = (\mathcal{F}(\omega)[u] - f(\omega, x, t))\, dt + \mathcal{B}(\omega)[u]\, dW(\omega, t) + \mathcal{P}(\omega)[u]\, dQ(\omega, t).$$

  In this Itô evolution equation, $W(\omega, t)$ is a Wiener process and $Q(\omega, t)$ is a Poisson process.

Predictions and Decision-making (equinox.ornl.gov)

Uncertainties are ubiquitous in all modeling efforts and our understanding of phenomena



An example stochastic partial differential equation (SPDE)

Pressure distribution in a porous medium

OAK RIDGE National Laboratory COMPUTATIONAL & APPLIED MATHEMATICS

Consider the following simple equation describing the pressure distribution in a porous medium:

$$\begin{cases} -\nabla \cdot (a(x)\,\nabla u(x)) &= f(x) \quad \text{in } D \subset \mathbb{R}^n, \\ u(x) &= 0 \quad \text{on } \partial D, \end{cases}$$

where $n = 1, 2, 3$, $f(x)$ is the source, $a(x)$ describes the permeability and $u(x)$ is the pressure distribution.

Q: What if the input data is random?

$$\begin{cases} -\nabla \cdot (a(x,\omega)\,\nabla u(x,\omega)) &= f(x,\omega) \quad \text{a.e. in } D, \\ u(x,\omega) &= 0 \quad \text{a.e. on } \partial D, \end{cases}$$

where $\omega \in \Omega$, $f(x,\omega)$ is the stochastic source, $a(x,\omega)$ describes the stochastic permeability, and $u(x,\omega)$ is the stochastic pressure distribution.

An example stochastic partial differential equation (SPDE)

Pressure distribution in a porous medium

Consider the following simple equation describing the pressure distribution in a porous medium:

$$\begin{cases} -\nabla \cdot (a(x)\,\nabla u(x)) &= f(x) \quad \text{in } D \subset \mathbb{R}^n, \\ u(x) &= 0 \quad \text{on } \partial D, \end{cases}$$

where $n = 1, 2, 3$, $f(x)$ is the source, $a(x)$ describes the permeability and $u(x)$ is the pressure distribution.

Q: What if the input data is random?

$$\begin{cases} -\nabla \cdot (a(x,\omega)\,\nabla u(x,\omega)) &= f(x,\omega) \quad \text{a.e. in } D, \\ u(x,\omega) &= 0 \quad \text{a.e. on } \partial D, \end{cases}$$

where $\omega \in \Omega$, $f(x,\omega)$ is the stochastic source, $a(x,\omega)$ describes the stochastic permeability, and $u(x,\omega)$ is the stochastic pressure distribution.

An example stochastic partial differential equation (SPDE)

Pressure distribution in a porous medium

Consider the following simple equation describing the pressure distribution in a porous medium:

$$\begin{cases} -\nabla \cdot (a(x)\,\nabla u(x)) = f(x) & \text{in } D \subset \mathbb{R}^n, \\ u(x) = 0 & \text{on } \partial D, \end{cases}$$

where $n = 1, 2, 3$, $f(x)$ is the source, $a(x)$ describes the permeability and $u(x)$ is the pressure distribution.

Q: What if the input data is random?

$$\begin{cases} -\nabla \cdot (a(x,\omega)\,\nabla u(x,\omega)) = f(x,\omega) & \text{a.e. in } D, \\ u(x,\omega) = 0 & \text{a.e. on } \partial D, \end{cases}$$

where $\omega \in \Omega$, $f(x,\omega)$ is the stochastic source, $a(x,\omega)$ describes the stochastic permeability, and $u(x,\omega)$ is the stochastic pressure distribution.

Assumptions and well-posedness

1. the solution has realizations in the Banach space $\mathcal{V} \equiv H_0^1(D)$, i.e., $u(\cdot,\omega) \in \mathcal{V}$ almost surely
$$\|u(\cdot,\omega)\|_{\mathcal{V}} \leq C\|f(\cdot,\omega)\|_{\mathcal{V}^*},$$
where $\mathcal{V}^* = H^{-1}(D)$.

2. the forcing term $f \in L^2_{\mathbb{P}}(\Omega;\mathcal{V}^*)$ is such that the solution u is unique and bounded in $L^2_{\mathbb{P}}(\Omega;\mathcal{V})$, i.e., Banach-valued functions that have finite second moments.

   For $q \in \mathbb{N}_+$
$$L^q_{\mathbb{P}}(\Omega;\mathcal{V}) = \left\{ v : \Omega \to \mathcal{V} \mid v \text{ is measurable and } \int_{\Omega} \|v(\omega,\cdot)\|^q_{\mathcal{V}}\, dP(\omega) < +\infty \right\}.$$

3. $a(\cdot,\omega)$ uniformly bounded and coercive, i.e., there exists $a_{\min}, a_{\max} \in (0,+\infty)$ such that
$$\mathbb{P}\left[\omega \in \Omega : \; a(\omega,x) \in (a_{\min},a_{\max})\, \forall x \in \overline{D}\right] = 1.$$

Challenges and goals

Approximate u or some statistical QoI depending on u

Difficulties arise since instead of just asking for $u(x)$, we instead want to approximate the entire stochastic solution $u(x, \omega)$ or some statistical QoI depending on u:

$$\phi_u := \mathbb{E}\left[\phi(u)\right] = \int_\Omega \int_D \psi(u(x,\omega), x, \omega)\, dx\, dP(\omega)$$

- **Moments:** $\overline{u} = \mathbb{E}[u](x)$ or $\mathbb{V}ar[u](x) = \mathbb{E}[\widetilde{u}^2](x)$, where $\widetilde{u} = u - \overline{u}$
- **Probabilities:** $\mathbb{P}\left\{u \geq u_0\right\} = \mathbb{P}\left[\{\omega \in \Omega \,:\, u(\omega) \geq u_0\}\right] = \mathbb{E}\left[\chi_{\{u \geq u_0\}}\right]$
- **Statistics of functionals of u:**

$$\phi(u) = \int_{\Sigma \subset D} u(x, \cdot\,)\, dx,$$

where Σ is a subdomain of interest.

Our goal: to develop highly efficient, robust, and scalable techniques that include uncertainty in the models, and allows us to quantify uncertainty in the outputs while providing reliable and verifiable predictions.

Challenges and goals

Approximate u or some statistical QoI depending on u

Difficulties arise since instead of just asking for $u(x)$, we instead want to approximate the entire stochastic solution $u(x, \omega)$ or some statistical QoI depending on u:

$$\phi_u := \mathbb{E}\left[\phi(u)\right] = \int_\Omega \int_D \psi(u(x,\omega), x, \omega)\, dx\, dP(\omega)$$

- **Moments:** $\overline{u} = \mathbb{E}[u](x)$ or $\mathbb{V}ar[u](x) = \mathbb{E}[\widetilde{u}^2](x)$, where $\widetilde{u} = u - \overline{u}$
- **Probabilities:** $\mathbb{P}\left\{u \geq u_0\right\} = \mathbb{P}\left[\{\omega \in \Omega \,:\, u(\omega) \geq u_0\}\right] = \mathbb{E}\left[\chi_{\{u \geq u_0\}}\right]$
- **Statistics of functionals of u:**

  $$\phi(u) = \int_{\Sigma \subset D} u(x, \cdot\,)\, dx,$$

  where Σ is a subdomain of interest.

Our goal: to develop highly efficient, robust, and scalable techniques that include uncertainty in the models, and allows us to quantify uncertainty in the outputs while providing reliable and verifiable predictions.

Plan of attack

1. Determine an accurate representation for the input stochastic (random) fields can be both simulated and analyzed, e.g.,a Karhunen-Loève expansion.

2. Transform the stochastic problem into a **deterministic** parametric version in $\mathbb{R}^n \times$ an $\infty-$ dimensional space.

3. Design an adaptive **dimensional reduction** procedure.

4. Design an adaptive **discretization** procedure using sampling methods, polynomial methods, or a combination of both.

5. Develop theory which **justifies** the approach.

6. Compare the convergence and **complexity** of the approach with existing methods.

Plan of attack

1. Determine an accurate representation for the input stochastic (random) fields can be both simulated and analyzed, e.g.,a Karhunen-Loève expansion.
2. Transform the stochastic problem into a **deterministic** parametric version in $\mathbb{R}^n \times$ an $\infty-$ dimensional space.
3. Design an adaptive **dimensional reduction** procedure.
4. Design an adaptive **discretization** procedure using sampling methods, polynomial methods, or a combination of both.
5. Develop theory which **justifies** the approach.
6. Compare the convergence and **complexity** of the approach with existing methods.

Plan of attack

1. Determine an accurate representation for the input stochastic (random) fields can be both simulated and analyzed, e.g.,a Karhunen-Loève expansion.

2. Transform the stochastic problem into a **deterministic** parametric version in $\mathbb{R}^n \times$ an $\infty-$ dimensional space.

3. Design an adaptive **dimensional reduction** procedure.

4. Design an adaptive **discretization** procedure using sampling methods, polynomial methods, or a combination of both.

5. Develop theory which **justifies** the approach.

6. Compare the convergence and **complexity** of the approach with existing methods.

Plan of attack

Oak Ridge National Laboratory — Computational & Applied Mathematics

1. Determine an accurate representation for the input stochastic (random) fields can be both simulated and analyzed, e.g.,a Karhunen-Loève expansion.
2. Transform the stochastic problem into a **deterministic** parametric version in $\mathbb{R}^n \times$ an $\infty-$ dimensional space.
3. Design an adaptive **dimensional reduction** procedure.
4. Design an adaptive **discretization** procedure using sampling methods, polynomial methods, or a combination of both.
5. Develop theory which **justifies** the approach.
6. Compare the convergence and **complexity** of the approach with existing methods.

Plan of attack

1. Determine an accurate representation for the input stochastic (random) fields can be both simulated and analyzed, e.g.,a Karhunen-Loève expansion.
2. Transform the stochastic problem into a **deterministic** parametric version in $\mathbb{R}^n \times$ an $\infty-$ dimensional space.
3. Design an adaptive **dimensional reduction** procedure.
4. Design an adaptive **discretization** procedure using sampling methods, polynomial methods, or a combination of both.
5. Develop theory which **justifies** the approach.
6. Compare the convergence and **complexity** of the approach with existing methods.

Plan of attack

1. Determine an accurate representation for the input stochastic (random) fields can be both simulated and analyzed, e.g.,a Karhunen-Loève expansion.
2. Transform the stochastic problem into a **deterministic** parametric version in $\mathbb{R}^n \times$ an $\infty-$ dimensional space.
3. Design an adaptive **dimensional reduction** procedure.
4. Design an adaptive **discretization** procedure using sampling methods, polynomial methods, or a combination of both.
5. Develop theory which **justifies** the approach.
6. Compare the convergence and **complexity** of the approach with existing methods.

Tools of the trade

Random variables and random fields

Input data $a(x,\omega)$, $f(x,\omega)$, and the solution $u(x,\omega)$ of the SPDE will (more likely) be a random field defined by a set of random variables $\boldsymbol{y}(\omega) = (y_1(\omega), \ldots, y_d(\omega))$, s.t.

$$\boldsymbol{y}(\omega) : \Omega \to \mathcal{U} = \prod_{i=1}^{d} \mathcal{U}_i \subset \mathbb{R}^d,$$

where $d \in \mathbb{N}$ that can be very large or even infinite.

WLOG assume $a(x,\omega) = a(x,\boldsymbol{y})$, $f(x,\omega) = f(x,\boldsymbol{y})$, and the solution $u(x,\omega) = u(x,\boldsymbol{y})$

We need to answer the following questions:

1. How to deal with random variables $y_k(\omega)$, $k = 1, 2, \ldots$?
2. How to represent random fields, $a(\cdot, \boldsymbol{y})$, $f(\cdot, \boldsymbol{y})$, and $u(\cdot, \boldsymbol{y})$?

Tools of the trade

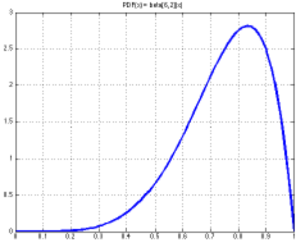
Random variables

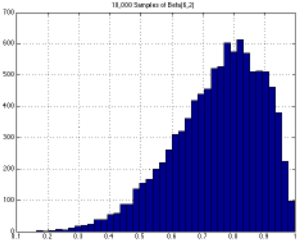
An $\mathbb{R}$-valued random variable (RV) $y : (\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{R}$ is completely specified by a probability density function (pdf) ϱ_y and a cumulative distribution function (cdf) F_y s.t. $\forall t \in \mathbb{R}$:

$$F_y(\lambda) := \mathbb{P}\left[\{y(\omega) \leq \lambda\}\right] = \int_{\{y(\omega)\leq\lambda\}} d\mathbb{P}(\omega) = \mathbb{E}\left[\chi_{\{y(\omega)\leq\lambda\}}\right]$$

$$= \int_{-\infty}^{\lambda} \varrho_y(t)\, dt,$$

where $\int_{\mathbb{R}} \varrho_y(t)\, dt = 1$





Random variables

Covariance, correlation and independence

For $y \in L^1_{\mathbb{P}}(\Omega)$ define the expected (mean) by

$$\overline{y} = \mathbb{E}\left[y\right] = \int_{\Omega} y(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} t\varrho_y(t)\, dt$$

and fluctuating part by $\widetilde{y} = y(\omega) - \overline{y}(\omega)$, with $\mathbb{E}[\widetilde{y}] = 0$.

- The variance $\mathbb{V}ar[y] = \mathbb{E}\left[\widetilde{y} \otimes \widetilde{y}\right] = \mathbb{E}\left[(\widetilde{y})^2\right] = \mathbb{C}ov[y, y]$
- Let $\boldsymbol{y}(\omega) = (y_k(\omega))_{k=1}^d, d \in \mathbb{N}_+$ be a random vector, then the covariance and correlation of two RVs:

$$\mathbb{C}ov[y_i, y_j] := \mathbb{E}\left[\widetilde{y_i} \otimes \widetilde{y_j}\right], \quad \mathbb{C}orr = \frac{\mathbb{C}ov[y_i, y_j]}{\sqrt{\mathbb{V}ar[y_i]}\sqrt{\mathbb{V}ar[y_j]}}$$

  - uncorrelated if $\mathbb{C}ov[y_i, y_j] = 0$ (orthogonal), perfectly correlated if $\mathbb{C}orr = 1$ and perfectly anti-correlated if $\mathbb{C}orr = -1$
  - independent if $\forall \phi_1, \phi_2$, $\mathbb{E}[\phi_1(y_1)\phi_2(y_2)] = \mathbb{E}[\phi_1(y_1)]\mathbb{E}[\phi_2(y_2)]$

Random variables

Covariance, correlation and independence

For $y \in L^1_{\mathbb{P}}(\Omega)$ define the expected (mean) by

$$\overline{y} = \mathbb{E}\left[y\right] = \int_{\Omega} y(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} t \varrho_y(t)\, dt$$

and fluctuating part by $\widetilde{y} = y(\omega) - \overline{y}(\omega)$, with $\mathbb{E}[\widetilde{y}] = 0$.

- The variance $\mathbb{V}ar[y] = \mathbb{E}\left[\widetilde{y} \otimes \widetilde{y}\right] = \mathbb{E}\left[(\widetilde{y})^2\right] = \mathbb{C}ov[y, y]$
- Let $\boldsymbol{y}(\omega) = (y_k(\omega))_{k=1}^d, d \in \mathbb{N}_+$ be a random vector, then the covariance and correlation of two RVs:

$$\mathbb{C}ov[y_i, y_j] := \mathbb{E}\left[\widetilde{y_i} \otimes \widetilde{y_j}\right], \quad \mathbb{C}orr = \frac{\mathbb{C}ov[y_i, y_j]}{\sqrt{\mathbb{V}ar[y_i]}\sqrt{\mathbb{V}ar[y_j]}}$$

  - uncorrelated if $\mathbb{C}ov[y_i, y_j] = 0$ (orthogonal), perfectly correlated if $\mathbb{C}orr = 1$ and perfectly anti-correlated if $\mathbb{C}orr = -1$
  - independent if $\forall \phi_1, \phi_2$, $\mathbb{E}[\phi_1(y_1)\phi_2(y_2)] = \mathbb{E}[\phi_1(y_1)]\mathbb{E}[\phi_2(y_2)]$

Random variables

Covariance matrix for vectors

We can now define a **covariance matrix** $\mathbb{C}ov[\boldsymbol{y}]$ whose (i,j) entry is the covariance between $y_i(\omega)$ and $y_j(\omega)$:

$$\mathbb{C}ov[\boldsymbol{y}]_{ij} = \mathbb{C}ov[y_i, y_j]$$

- $\mathbb{C}ov[\boldsymbol{y}]$ is symmetric, nonnegative definite, and has diagonal elements $\mathbb{C}ov[\boldsymbol{y}]_{ii} = \mathbb{V}ar[y_i]$

As before, the **correlation matrix** can be defined from the covariance matrix. Form a diagonal matrix Σ from the square roots of the variances, then compute the correlation matrix by:

$$\mathbb{C}orr[\boldsymbol{y}] = \Sigma^{-1}\mathbb{C}ov[\boldsymbol{y}]\Sigma$$

- The diagonal entries of $\mathbb{C}orr[\boldsymbol{y}]$ are 1.
- The Cauchy-Schwarz inequality guarantees that the off-diagonal elements lie between -1 and $+1$.
- Value of each covariance entry indicates the strength and direction of the correlation between the corresponding components.

Random variables

Covariance matrix for vectors

We can now define a **covariance matrix** $\mathbb{C}ov[\boldsymbol{y}]$ whose (i, j) entry is the covariance between $y_i(\omega)$ and $y_j(\omega)$:

$$\mathbb{C}ov[\boldsymbol{y}]_{ij} = \mathbb{C}ov[y_i, y_j]$$

- $\mathbb{C}ov[\boldsymbol{y}]$ is symmetric, nonnegative definite, and has diagonal elements $\mathbb{C}ov[\boldsymbol{y}]_{ii} = \mathbb{V}ar[y_i]$

As before, the **correlation matrix** can be defined from the covariance matrix. Form a diagonal matrix Σ from the square roots of the variances, then compute the correlation matrix by:

$$\mathbb{C}orr[\boldsymbol{y}] = \Sigma^{-1}\mathbb{C}ov[\boldsymbol{y}]\Sigma$$

- The diagonal entries of $\mathbb{C}orr[\boldsymbol{y}]$ are 1.
- The Cauchy-Schwarz inequality guarantees that the off-diagonal elements lie between -1 and $+1$.
- Value of each covariance entry indicates the strength and direction of the correlation between the corresponding components.

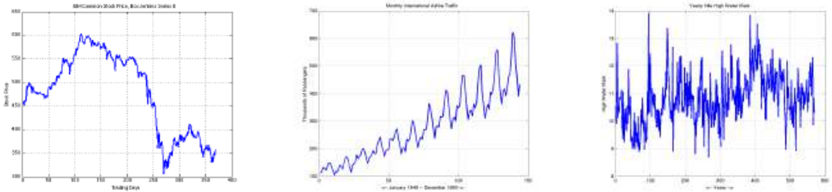
Tools of the trade

Random processes and fields: generalizations of random variables

A random process/field (RP/RF) $a(x, \boldsymbol{y}(\omega))$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and indexed by a deterministic domain $D \subset \mathbb{R}^n$, returns a real value

1. a set of RVs indexed by $x \in D$. For every $x \in D$, $a(x, \cdot)$ is a RV on Ω
2. a function-valued RV. For every $\omega \in \Omega$, $a(\cdot, \boldsymbol{y}(\omega))$ is a random function - a realization - of x in the domain D

Often only second order information - mean and covariance are known



- Mean $\overline{a}(x) = \mathbb{E}[a](x) = \int_\Omega a(\cdot, \boldsymbol{y}(\omega))d\mathbb{P}(\omega)$ and $\mathbb{V}ar[a](x) = \mathbb{E}[(\widetilde{a})^2](x)$ as a function of x with fluctuation part $\widetilde{a}(x, \boldsymbol{y}(\omega)) = a - \overline{a}$
- $\mathbb{P}\left[a \geq a_0\right] = \mathbb{P}\left[\{\omega \in \Omega \,:\, a(x, \boldsymbol{y}(\omega)) \geq a_0\}\right] = \mathbb{E}\left[\chi_{\{a \geq a_0\}}\right]$

Random fields (RF)

White vs. colored noise

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

The covariance covariance function describes the interaction between points in $D \subset \mathbb{R}^d$:

$$\mathbb{C}ov[a](x_1, x_2) := \mathbb{E}\left[\widetilde{a}(\,\cdot\,, x_1)\widetilde{a}(\,\cdot\,, x_2)\right], \text{ for } x_1, x_2 \in D \times D$$

- if $\overline{a}(x) \equiv \overline{a}$ and $\mathbb{C}ov[a](x_1, x_2) = C_a(x_1 - x_2)$ then the process is homogeneous. Here representation through the spectrum using Fourier expansion if **well known**
- we **will** consider colored noise approximations using correlated second-order RFs: $a(x, \boldsymbol{y}(\omega)) \in L^2_{\mathbb{P}}(\Omega; \mathcal{V})$.
- we will **not** focus on white noise approximations which refers to uncorrelated RFs for which:

$$\overline{a}(x) = 0 \text{ and } \mathbb{C}ov[a](x_1, x_2) = \delta(x_1 - x_2)$$

Random fields (RF)

White vs. colored noise

The covariance covariance function describes the interaction between points in $D \subset \mathbb{R}^d$:

$$\mathbb{C}ov[a](x_1, x_2) := \mathbb{E}\left[\widetilde{a}(\cdot, x_1)\widetilde{a}(\cdot, x_2)\right], \text{ for } x_1, x_2 \in D \times D$$

- if $\overline{a}(x) \equiv \overline{a}$ and $\mathbb{C}ov[a](x_1, x_2) = C_a(x_1 - x_2)$ then the process is homogeneous. Here representation through the spectrum using Fourier expansion if **well known**
- we **will** consider colored noise approximations using correlated second-order RFs: $a(x, \boldsymbol{y}(\omega)) \in L^2_{\mathbb{P}}(\Omega; \mathcal{V})$.
- we will **not** focus on white noise approximations which refers to uncorrelated RFs for which:

$$\overline{a}(x) = 0 \text{ and } \mathbb{C}ov[a](x_1, x_2) = \delta(x_1 - x_2)$$

Random fields (RF)

White vs. colored noise

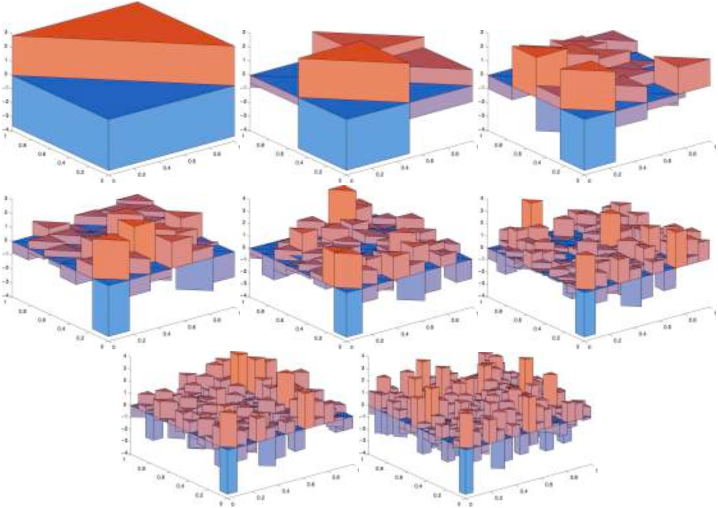
The covariance covariance function describes the interaction between points in $D \subset \mathbb{R}^d$:

$$\mathbb{C}ov[a](x_1, x_2) := \mathbb{E}\left[\widetilde{a}(\,\cdot\,, x_1)\widetilde{a}(\,\cdot\,, x_2)\right], \text{ for } x_1, x_2 \in D \times D$$

- if $\overline{a}(x) \equiv \overline{a}$ and $\mathbb{C}ov[a](x_1, x_2) = C_a(x_1 - x_2)$ then the process is homogeneous. Here representation through the spectrum using Fourier expansion if **well known**
- we **will** consider colored noise approximations using correlated second-order RFs: $a(x, \boldsymbol{y}(\omega)) \in L^2_{\mathbb{P}}(\Omega; \mathcal{V})$.
- we will **not** focus on white noise approximations which refers to uncorrelated RFs for which:
$$\overline{a}(x) = 0 \text{ and } \mathbb{C}ov[a](x_1, x_2) = \delta(x_1 - x_2)$$

Discretized white noise

Piecewise constant approximations



Discretized white noise over a square subdivided into $2, \ldots, 512$ triangles

Random fields (RF)

Finite dimensional noise assumption

Given the mean and covariance of the RF we would like to construct a simple representation which captures this information and used for simulations.

- Ideally, this involves a combination of countably many independent RVs
- Let $\mathcal{U}_n \equiv y_n(\Omega) \subset \mathbb{R}$ be the image of the RV, i.e., $\mathcal{U}_n = [-1, 1]$, and $\mathcal{U} = \prod_{n=1}^{d} \mathcal{U}_n \subset \mathbb{R}^d$, the image of the random vector $\boldsymbol{y}(\omega) : \Omega \to \mathcal{U}$.
- Let $\varrho : \mathcal{U} \to \mathbb{R}_+$, with $\varrho \in L^\infty(\mathcal{U})$ be the joint probability density function (PDF) of $\boldsymbol{y} = (y_1, \ldots, y_d)$, then we **want** that:

$$\varrho(\boldsymbol{y}) = \prod_{n=1}^{d} \varrho_n(y_n), \text{ where } \boldsymbol{y} \in \mathcal{U} \text{ and } y_n \in \mathcal{U}_n, \forall n$$

- The independence of the d RVs allows to see each of them as the axis of a coordinate system (Doob-Dynkin Lemma)
- The most popular approach: Karhunen-Loève (KL) expansion - involves an ∞-dimensional expansion of the random field suitably truncated
- Challenge: this truncation in d RVs can be high-dimensional
- In the discrete case - KL is similar to ROM and the SVD of a matrix

Random fields (RF)

Finite dimensional noise assumption

Given the mean and covariance of the RF we would like to construct a simple representation which captures this information and used for simulations.

- Ideally, this involves a combination of countably many independent RVs
- Let $\mathcal{U}_n \equiv y_n(\Omega) \subset \mathbb{R}$ be the image of the RV, i.e., $\mathcal{U}_n = [-1, 1]$, and $\mathcal{U} = \prod_{n=1}^{d} \mathcal{U}_n \subset \mathbb{R}^d$, the image of the random vector $\boldsymbol{y}(\omega) : \Omega \to \mathcal{U}$.
- Let $\varrho : \mathcal{U} \to \mathbb{R}_+$, with $\varrho \in L^\infty(\mathcal{U})$ be the joint probability density function (PDF) of $\boldsymbol{y} = (y_1, \ldots, y_d)$, then we **want** that:

$$\varrho(\boldsymbol{y}) = \prod_{n=1}^{d} \varrho_n(y_n), \text{ where } \boldsymbol{y} \in \mathcal{U} \text{ and } y_n \in \mathcal{U}_n, \ \forall n$$

- The independence of the d RVs allows to see each of them as the axis of a coordinate system (Doob-Dynkin Lemma)
- The most popular approach: Karhunen-Loève (KL) expansion - involves an ∞-dimensional expansion of the random field suitably truncated
- Challenge: this truncation in d RVs can be high-dimensional
- In the discrete case - KL is similar to ROM and the SVD of a matrix

Motivate the Karhunen-Loève expansion

The Singular Value Decomposition (SVD)

Every (real) m by n matrix A has a singular value decomposition:

$$A = U\, S\, V^T$$

where

- U is an m by m orthogonal matrix ($U^T\, U = I$);
- S is an m by n diagonal matrix with nonnegative entries;
- V is an n by n orthogonal matrix;

The diagonal entries of S, called the singular values of A, are chosen to appear in descending order, and are equal to the square roots of the nonzero eigenvalues of AA^T or $A^T A$

SVD

Facts about the SVD I

$$A = U\,S\,V^T$$

- r, the number of nonzero diagonal elements in S, is the rank of A.
  - very small non-zeros may indicate numeric singularities

- the i-th diagonal element of S is the i-th largest eigenvalue of AA^T (and also of AA^T). Hence, we may write this value as $\sqrt{\lambda_i}$.

- Let u_i and v_i^T be the i-th columns of U and V^T. Then A maps the i-th column of V^T to the i-th column of U.

- The columns of U and V provide a singular value expansion of A:

$$A = \sum_{i=1}^{r} \sqrt{\lambda_i}\, u_i\, v_i^T$$

SVD

Facts about the SVD II

If we use all r terms, the singular value expansion is **exact**.

Let A^k represent the sum of just the first k terms of the expansion:

- A^k is a matrix of rank k, the sum of k rank-1 outer products
- Of all rank k matrices, A^k is the best approximation to A in two senses:

1. Minimum L^2 norm:

$$\left\|A - A^k\right\|_{L^2} \equiv \text{square root of maximum eigenvalue of } (A - A^k)^T (A - A^k)$$

$$\left\|A - A^k\right\|_{L^2}^2 = s_{k+i}^2 = \lambda_{k+1}$$

2. Minimum Frobenius (sum of squares) norm:

$$\left\|A - A^k\right\|_F \equiv \sqrt{\sum_{i,j} (A_{i,j} - A_{i,j}^k)^2}$$

$$\left\|A - A^k\right\|_F^2 = \sum_{k+1}^{r} s_i^2 = \sum_{k+1}^{r} \lambda_i$$

SVD

Using the facts I

U and V are natural bases for the input and output of A.

In the natural bases, the SVD shows that multiplying by A is simply stretching the i-th component by s_i:

$$x = \sum_{i=1}^{r} v_i^T * c_i \implies y = A * x = \sum_{i=1}^{r} u_i * (s_i * c_i)$$

- The relative size of the singular values indicates the importance of each column.
- The singular value expansion produces an optimal, indexed family of reduced order models of A.

SVD

Using the facts II

- SVD is the discrete version of the KL expansion that is typically applied to RF that produce: for any time t, a field of values varying spatially with x.
- Since it's easier to understand discrete problems, let's prepare for the KL expansion by looking at how the SVD is used with a set of data.
- Let us re-imagine the columns of our discrete data as being n snapshots in discrete time indexed by j. Each snapshot will record m values in a "space" indexed by i.

SVD

- If we pack our data into a single matrix A, then $A_{i,j}$ means the measurement at position i and time j.
- It is reasonable to expect correlation in this data; the "neighbors" of $A_{i,j}$, in either space or time, might tend to have similar values.
- Moreover, the overall "shape" of the data for one time or one spatial coordinate might be approximately repeated elsewhere in the data.
- This is exactly the kind of behavior the SVD can detect.

Space

Time								
1890	1	12	12	33	29	22	3	0
1891	0	31	23	44	18	13	1	0
1892	0	23	44	25	17	17	13	1
1893	1	30	49	37	15	23	10	1
1894	0	30	18	74	9	5	0	2

SVD example

Snowfall at Michigan Tech

We have a data file of the monthly snowfall in inches, over 121 winters at Michigan Tech. We'll think of the months as the "space" dimension.

Year	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Tot
1890	1	12	12	33	29	22	3	0	112
1891	0	31	23	44	18	13	1	0	130
1892	0	23	44	25	17	17	13	1	140
1893	1	30	49	37	15	23	10	1	166
1894	0	30	18	74	9	5	0	2	138
….	….	….	….	….	….	….	….	….	….
2006	6	6	27	38	37	20	31	0	165
2007	0	21	40	55	32	24	14	0	186
2008	0	17	70	85	27	5	15	0	219
2009	3	4	87	39	19	0	0	0	152
2010	0	26	33	72	18	13	18	0	180

http://www.mtu.edu/alumni/favorites/snowfall/snowfall.html

SVD example

Snowfall at Michigan Tech

To analyze our data, we consider each of the 121 snowfall records, starting with x^{1890}, as a column of 8 numbers, and form the m=8 by n=121 matrix A:

$$A = \left[x^{1890} | x^{1891} | \ldots | x^{2010}\right]$$

- Now we determine the SVD decomposition $A = USV^T$.
- The columns of U are an orthogonal set of "spatial" behaviors or modes (typical behavior in a fixed year over a span of months).
- The columns of V are typical behaviors or modes in a fixed month over a span of years.
    - In both cases, the most important behaviors are listed first.
- The diagonal matrix S contains the "importance" or "energy" or signal strength associated with each behavior.

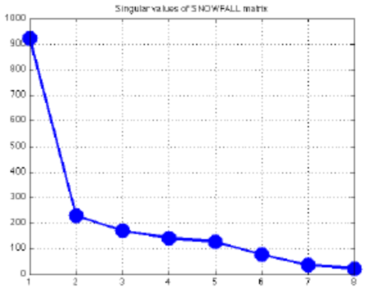
SVD example

The first 9 singular values

The S data shows the relative importance of the first two modes is:

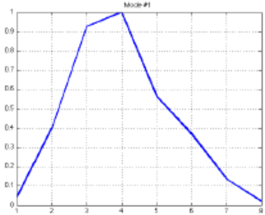
$$\frac{s_1^2}{\sqrt{\sum_{i=1}^{8} s_i^2}} = 0.87 \qquad \frac{s_2^2}{\sqrt{\sum_{i=1}^{8} s_i^2}} = 0.05$$

The first pair of modes, u_1 and v_1, by itself, can approximate the entire dataset with a relative accuracy of 87%.

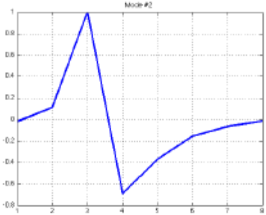


SVD example

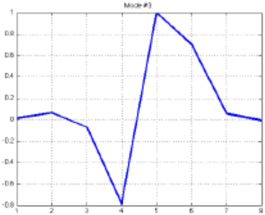
Four strongest snowfall modes for a year



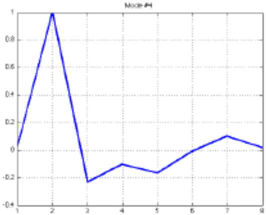
1 December/January High (DOMINANT)



2 More December, less later



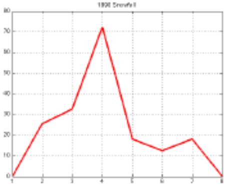
3 February High, less January



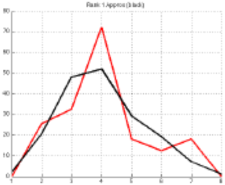
4 More November snow

SVD example

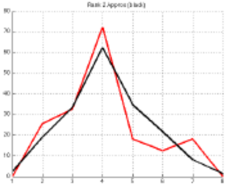
Approximating 2010-2011 snowfall



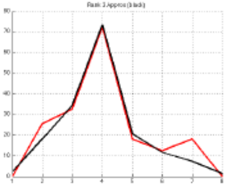
2010 − 2011 Data



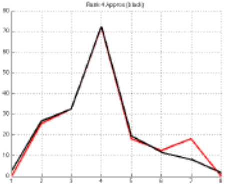
1 Mode



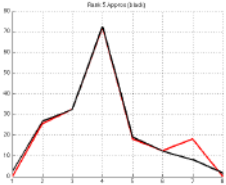
2 Modes



3 Modes



4 Modes



5 Modes

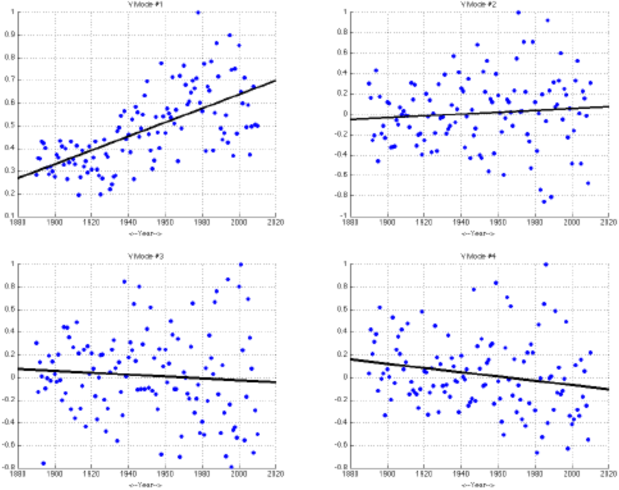
The same kind of approximating is occurring for all 121 sets of data!

SVD example

Four strongest "time" Modes

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

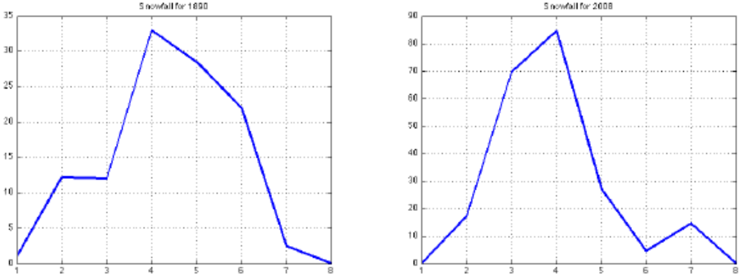
The linear regression line suggests the "December/January High" pattern (upper left) is steadily gaining importance over the years.



SVD

Typical old and new snowfall patterns

To see how heaviest snowfall is coming earlier, compare the 1890 January/February style snowfall with the 2008 December/January style:



SVD

Conclusions

1. Data gathered at discrete places and times is easier to understand than the corresponding continuous cases.

2. The SVD shows how underlying patterns and correlations can be detected, and represented as a sum of the form:

$$A = \sum_{i=1}^{r} \sqrt{\lambda_i}\, u_i\, v_i^T,$$

where the λ values represent a strength, the u's represent variation in space, and v variation in time.

3. The structure of the u and v vectors suggests something about the preferred modes of the system, and the size of the λ coefficients allows us to understand the relative important of different modes, and to construct reduced order models if we wish.

SVD

Conclusions

1. Data gathered at discrete places and times is easier to understand than the corresponding continuous cases.

2. The SVD shows how underlying patterns and correlations can be detected, and represented as a sum of the form:

$$A = \sum_{i=1}^{r} \sqrt{\lambda_i}\, u_i\, v_i^T,$$

where the λ values represent a strength, the u's represent variation in space, and v variation in time.

3. The structure of the u and v vectors suggests something about the preferred modes of the system, and the size of the λ coefficients allows us to understand the relative important of different modes, and to construct reduced order models if we wish.

SVD

Conclusions

1. Data gathered at discrete places and times is easier to understand than the corresponding continuous cases.

2. The SVD shows how underlying patterns and correlations can be detected, and represented as a sum of the form:

$$A = \sum_{i=1}^{r} \sqrt{\lambda_i}\, u_i\, v_i^T,$$

   where the λ values represent a strength, the u's represent variation in space, and v variation in time.

3. The structure of the u and v vectors suggests something about the preferred modes of the system, and the size of the λ coefficients allows us to understand the relative important of different modes, and to construct reduced order models if we wish.

SVD

Conclusions II

- Given that our data was stored in A, we may think of the matrices AA^T and $A^T A$ as a form of a covariance matrix.
- The singular values $\sqrt{\lambda_i}$ are the square roots of eigenvalues of both these matrices.
- . U contains eigenvectors of the "spatial" covariance matrix AA^T.
- V contains eigenvectors of the "temporal" covariance matrix $A^T A$.
- **Very similar statements will hold for the continuous case**.

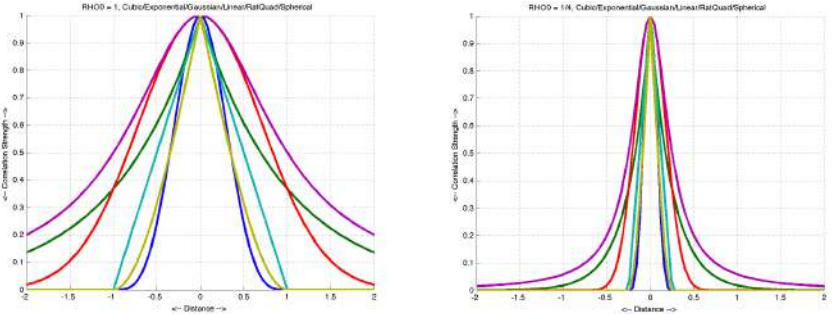
  It will be helpful to keep the discrete case in mind as we briefly discuss the continuous case next.

Stochastic representation of a RF

Karhunen-Loève expansion of the input data

The Karhunen-Loève expansion of a RF is a Fourier-type series based on the spectral expansion of its covariance function - other names Proper Orthogonal Decomposition (POD), Principle Component Analysis (PCA)

Let $a(x, \boldsymbol{y}(\omega)) \in L^2_{\mathbb{P}}$ be a RF with continuous covariance $\mathbb{C}_a : D \times D \to \mathbb{R}$



Examples of 1d covariance kernels for correlation lengths $L_c = 1$ and $L_c = 1/4$

Karhunen-Loève expansion
Properties of the covariance function

1. $\mathbb{C}_a$ **is symmetric** if $\mathbb{C}_a(x_1, x_2) = \mathbb{C}_a(x_2, x_1)$, $\forall x_1, x_2 \in D$.

2. $\mathbb{C}_a$ **is non-negative definite** if for any $n = 1, \ldots$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}_a(x_i, x_j) v_i v_j \geq 0, \forall (x_1, \ldots, x_n) \in D^n \text{ and } (v_1, \ldots, v_n) \in \mathbb{R}^n.$$

In matrix notation: $v^T \mathbb{C}_a(x, x) v \geq 0,\ \forall v, x.$

Define the associated linear covariance operator $T_{\mathbb{C}_a} : L^2(D) \to L^2(D)$ s.t.:

$$[T_{\mathbb{C}_a} f](x_1) = \int_D \mathbb{C}_a(x_1, x_2) f(x_2) dx_2, \ \ \forall f \in L^2(D).$$

Observation: If $T_{\mathbb{C}_a} f \in C^0(D)$ $\forall f \in L^2(D)$, $\mathbb{C}_a \mapsto T_{\mathbb{C}_a}$ is injective, and $T_{\mathbb{C}_a}$ is compact, symmetric and non-negative definite, then:

- it has a countable sequence of real eigenvalues $\{\lambda_n\} \subset \mathbb{R}_+$, $\lambda_n \to 0$.
- corresponding eigenfunctions $\{b_n(x)\}$ are $L^2(D)$-orthonormal.

Karhunen-Loève expansion

Properties of the covariance function

1. $\mathbb{C}_a$ **is symmetric** if $\mathbb{C}_a(x_1, x_2) = \mathbb{C}_a(x_2, x_1)$, $\forall x_1, x_2 \in D$.
2. $\mathbb{C}_a$ **is non-negative definite** if for any $n = 1, \ldots$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}_a(x_i, x_j) v_i v_j \geq 0, \forall (x_1, \ldots, x_n) \in D^n \text{ and } (v_1, \ldots, v_n) \in \mathbb{R}^n.$$

In matrix notation: $v^T \mathbb{C}_a(x, x) v \geq 0,\ \forall v, x.$

Define the associated linear covariance operator $T_{\mathbb{C}_a} : L^2(D) \to L^2(D)$ s.t.:

$$[T_{\mathbb{C}_a} f](x_1) = \int_D \mathbb{C}_a(x_1, x_2) f(x_2) dx_2, \quad \forall f \in L^2(D).$$

Observation: If $T_{\mathbb{C}_a} f \in C^0(D)\ \forall f \in L^2(D)$, $\mathbb{C}_a \mapsto T_{\mathbb{C}_a}$ is injective, and $T_{\mathbb{C}_a}$ is compact, symmetric and non-negative definite, then:

- it has a countable sequence of real eigenvalues $\{\lambda_n\} \subset \mathbb{R}_+$, $\lambda_n \to 0$.
- corresponding eigenfunctions $\{b_n(x)\}$ are $L^2(D)$-orthonormal.

Karhunen-Loève expansion

Spectral representation of the kernel

Mercer's (1909) spectral representation of the kernel:

$$\mathbb{C}_a(x_1, x_2) = \sum_{n=1}^{\infty} \lambda_n b_n(x_1) b_n(x_2)$$

Eigenvalues/eigenfunctions constructed from a 2nd-order Fredholm equation:

$$[T_{\mathbb{C}_a} b_n](x_1) = \int_D \mathbb{C}_a(x_1, x_2) b_n(x_2) dx_2 = \lambda_n b_n(x_1), \quad n = 1, \ldots$$

$$\text{with} \quad \int_D b_n(x_1) b_m(x_1) dx_1 = \delta_{nm}$$

Theorem. [Mercer, 1909].

Given $\mathbb{C}_a$ continuous, symmetric, non-negative definite, then:

$$\lim_{d \to \infty} \max_{(x_1, x_2) \in D \times D} \left| \mathbb{C}_a(x_1, x_2) - \sum_{n=1}^{d} \lambda_n b_n(x_1) b_n(x_2) \right| = 0$$

Karhunen-Loève expansion

Spectral representation of the kernel

Mercer's (1909) spectral representation of the kernel:

$$\mathbb{C}_a(x_1, x_2) = \sum_{n=1}^{\infty} \lambda_n b_n(x_1) b_n(x_2)$$

Eigenvalues/eigenfunctions constructed from a 2nd-order Fredholm equation:

$$[T_{\mathbb{C}_a} b_n](x_1) = \int_D \mathbb{C}_a(x_1, x_2) b_n(x_2) dx_2 = \lambda_n b_n(x_1), \quad n = 1, \ldots$$

$$\text{with} \quad \int_D b_n(x_1) b_m(x_1) dx_1 = \delta_{nm}$$

Theorem. [Mercer, 1909].

Given $\mathbb{C}_a$ continuous, symmetric, non-negative definite, then:

$$\lim_{d\to\infty} \max_{(x_1,x_2)\in D\times D} \left| \mathbb{C}_a(x_1, x_2) - \sum_{n=1}^{d} \lambda_n b_n(x_1) b_n(x_2) \right| = 0$$

Karhunen-Loève expansion

Approximating an ∞-dimensional RF

$$a(x,\omega) = \overline{a}(x) + \sum_{n=1}^{+\infty} b_n(x)\, y_n(\omega),$$

$$\text{with } y_n(\omega) = \int_D (a(\omega,x) - \overline{a}(x)) b_n(x)\, dx$$

- $(\lambda_n, b_n(x))$ are eigenpairs of $T_{\mathbb{C}_a}$; $y_n(\omega)$ are centered, uncorrelated RVs, i.e.,

$$\mathbb{E}[y_n] = 0, \; \mathbb{C}ov[y_n, y_m] = \mathbb{E}[y_n y_m] = \delta_{nm}$$

but not necessarily independent, with $\mathbb{V}ar[y_n] = \lambda_n$.

If the basis $\{b_n\}$ has spectral approx. properties and the realizations of a are smooth, then $\lambda_n = \mathbb{V}ar[y_n] \to 0$ sufficiently fast as $n \to \infty$ and we can truncate the series

$$a(\omega,x) \approx a_d(\omega,x) = \overline{a}(x) + \sum_{n=1}^{d} b_n(x)\, y_n(\omega),$$

Rate of decay depends on the smoothness of $\mathbb{C}_a$ and the correlation length L_c.

Karhunen-Loève expansion

Approximating an ∞-dimensional RF

$$a(x,\omega) = \overline{a}(x) + \sum_{n=1}^{+\infty} b_n(x)\, y_n(\omega),$$

$$\text{with } y_n(\omega) = \int_D (a(\omega,x) - \overline{a}(x)) b_n(x)\, dx$$

- $(\lambda_n, b_n(x))$ are eigenpairs of $T_{\mathbb{C}_a}$; $y_n(\omega)$ are centered, uncorrelated RVs, i.e.,

$$\mathbb{E}[y_n] = 0,\ \mathbb{C}ov[y_n, y_m] = \mathbb{E}[y_n y_m] = \delta_{nm}$$

but not necessarily independent, with $\mathbb{V}ar[y_n] = \lambda_n$.

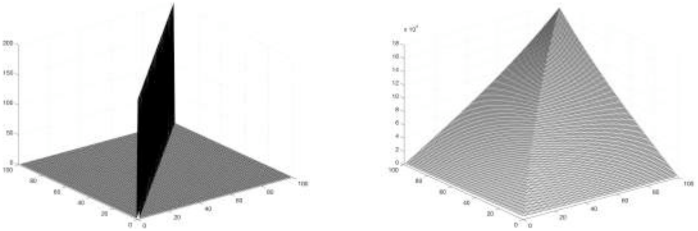
If the basis $\{b_n\}$ has spectral approx. properties and the realizations of a are smooth, then $\lambda_n = \mathbb{V}ar[y_n] \to 0$ sufficiently fast as $n \to \infty$ and we can truncate the series

$$a(\omega,x) \approx a_d(\omega,x) = \overline{a}(x) + \sum_{n=1}^{d} b_n(x)\, y_n(\omega),$$

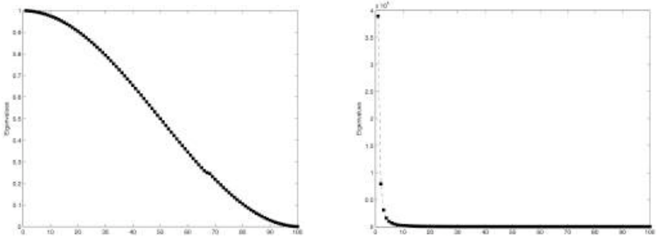
Rate of decay depends on the smoothness of $\mathbb{C}_a$ and the correlation length L_c.

Karhunen-Loève expansion

Example: Ornstein-Uhlenbeck process



Peaked and smooth covariance functions



Corresponding KL eigenvalues

Karhunen-Loève expansion

Truncation error minimization

This truncated expansion corresponds to the Best d-term approximation:

$$\min_{\substack{(y_n, b_n) \\ \int_D b_n b_m = \delta_{nm}}} \mathbb{E}\left[\int_D \left(a - \overline{a}(x) - \sum_{n=1}^{d} b_n y_n\right)^2\right].$$

- If we truncate using the d largest eigenvalues, we have an optimal - in variance - expansion in d random varibles.
- i.e., with $\mathbb{C}_a$ continuous, a_d converges uniformly to a (Mercer's Theorem)

$$\sup_{x \in D} \mathbb{E}[(a - a_d)^2](x) = \sup_{x \in D} \left\{ \mathbb{C}_a(x) - \sum_{n=1}^{d} \lambda_n b_n^2(x) \right\} \to 0, \text{ as } d \to \infty.$$

- "Karhunen-Loève expansion is the SVD" of the map $A : L^2(D) \to L^2_{\mathbb{P}}(\Omega)$, where $\mathbb{C}_a := A^* A$, i.e., truncate at the d largest eigenvalues of $A^* A$:

  $\Rightarrow$ Model redcution - find a d-sparse representation.

Karhunen-Loève expansion

Truncation error minimization

This truncated expansion corresponds to the Best d-term approximation:

$$\min_{\substack{(y_n, b_n) \\ \int_D b_n b_m = \delta_{nm}}} \mathbb{E}\left[\int_D \left(a - \overline{a}(x) - \sum_{n=1}^{d} b_n y_n\right)^2\right].$$

- If we truncate using the d largest eigenvalues, we have an optimal - in variance - expansion in d random varibles.
- i.e., with $\mathbb{C}_a$ continuous, a_d converges uniformly to a (Mercer's Theorem)

$$\sup_{x \in D} \mathbb{E}[(a - a_d)^2](x) = \sup_{x \in D} \left\{ \mathbb{C}_a(x) - \sum_{n=1}^{d} \lambda_n b_n^2(x) \right\} \to 0, \text{ as } d \to \infty.$$

- "Karhunen-Loève expansion is the SVD" of the map $A : L^2(D) \to L^2_{\mathbb{P}}(\Omega)$, where $\mathbb{C}_a := A^* A$, i.e., truncate at the d largest eigenvalues of $A^* A$:

  $\Rightarrow$ Model redcution - find a d-sparse representation.

Karhunen-Loève expansion

Truncation error minimization

This truncated expansion corresponds to the Best d-term approximation:

$$\min_{\substack{(y_n, b_n) \\ \int_D b_n b_m = \delta_{nm}}} \mathbb{E}\left[\int_D \left(a - \overline{a}(x) - \sum_{n=1}^{d} b_n y_n\right)^2\right].$$

- If we truncate using the d largest eigenvalues, we have an optimal - in variance - expansion in d random varibles.
- i.e., with $\mathbb{C}_a$ continuous, a_d converges uniformly to a (Mercer's Theorem)

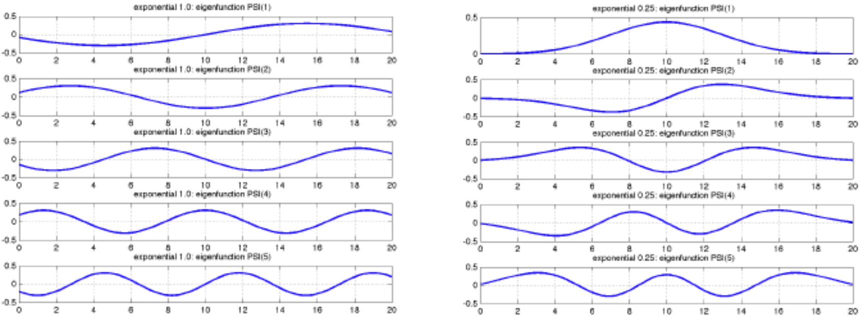
$$\sup_{x \in D} \mathbb{E}[(a - a_d)^2](x) = \sup_{x \in D} \left\{\mathbb{C}_a(x) - \sum_{n=1}^{d} \lambda_n b_n^2(x)\right\} \to 0, \text{ as } d \to \infty.$$

- "Karhunen-Loève expansion is the SVD" of the map $A : L^2(D) \to L^2_{\mathbb{P}}(\Omega)$, where $\mathbb{C}_a := A^* A$, i.e., truncate at the d largest eigenvalues of $A^* A$:

  $\Rightarrow$ Model redcution - find a d-sparse representation.

Karhunen-Loève expansion

Example eigenfunctions

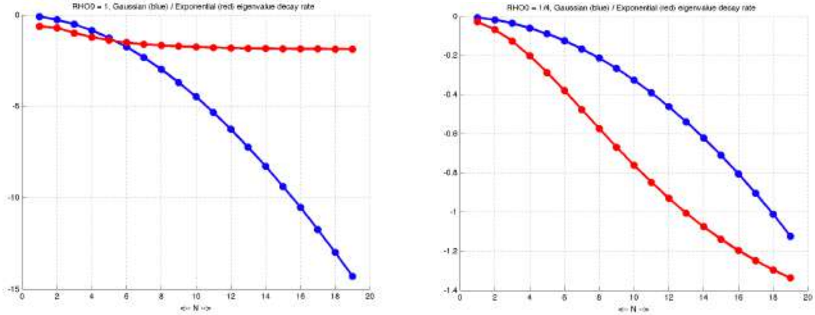


Distributions of the eigenfunctions of 1d exponential kernels, $\mathbb{C}_a(x_1, x_2) = \sigma^2 e^{-\frac{|x_1 - x_2|}{L_c}}$, for correlation lengths $L_c = 1$ and $L_c = 1/4$

Karhunen-Loève expansion

Examples of eigenvalues

exp.: $\mathbb{C}_a(x_1, x_2) = \sigma^2 e^{-\frac{\|x_1 - x_2\|_1}{L_c^2}}$ Gaussian: $\mathbb{C}_a(x_1, x_2) = \sigma^2 e^{-\frac{\|x_1 - x_2\|^2}{L_c^2}}$



Eigenvalues values of the kernels for correlation lengths $L_c = 1$ and $L_c = 1/4$

- More modes required as the correlation decreases (noise level increases).
- In the asymptotic limit of white noise $\Rightarrow$ infinity number of modes.
- For a given L_c, the smoothness of the covariance kernel $\mathbb{C}_a$ dictates the convergence rate of the eigenvalues.

Karhunen-Loève expansion

Convergence of the spectrum

- the truncation error decreases monotonically with the number of terms in the expansion.
- the convergence is inversely proportional to the correlation length and depends on the regularity of the covariance kernel.

Theorem. [Schwab et al., 2005].

- If $\mathbb{C}_a$ is piecewise analytic on $D \times D$ with $D \subset \mathbb{R}^d$ then:

$$0 \leq \lambda_n \leq c_1 \exp(-c_2\, n^{1/d}), \quad \forall c_1, c_2 > 0 \text{ independent of } n.$$

- If $\mathbb{C}_a$ is piecewise $H^k \otimes L^2$ with $k > 0$ then:

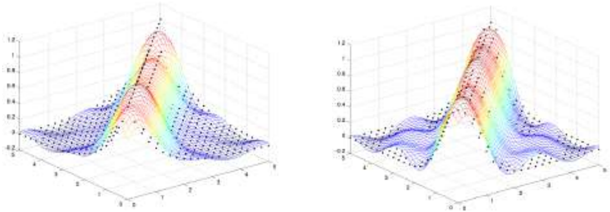
$$0 \leq \lambda_n \leq c_3\, n^{-k/d}, \quad \forall c_3 > 0.$$

Remark: similar to SVD - if one wants the relative error (in the variance) less than some tolerance δ, i.e., $\|a - a_d\|^2 \leq \delta\, \|a\|$, then choose d s.t.

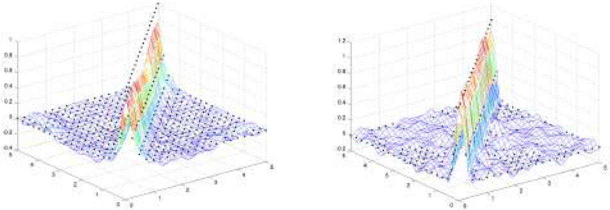
$$\sum_{n=d+1}^{\infty} \lambda_n \leq \delta \sum_{n=1}^{\infty} \lambda_n.$$

Karhunen-Loève expansion

Approximating 90% variance with Gaussian kernel



Exponential and Gaussian kernels with $L_c = 1$ and $d = 5$ modes



Exponential and Gaussian kernels with $L_c = 1/4$ and $d = 20$ modes

Karhunen-Loève expansion

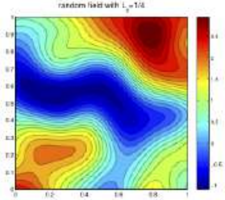
Special case: Gaussian RFs

Although the y_n's are uncorrelated, in general they are not independent:

- Gaussian RVs are uncorrelated $\iff$ independent, i.e., $\varrho(\boldsymbol{y}) = \prod_{n=1}^{d} \varrho_n(y_n)$.

Gaussian random fields: For every $x \in D$, $a(x, \cdot) \sim N(\mu(x), \varrho(x, x'))$

- Characterized by mean $\mu(x)$ and covariance $\varrho(x, x') = \mathbb{E}[\hat{a}(x, \cdot)\hat{a}(x', \cdot)]$, where $\hat{a}(x, \cdot) = a(x, \cdot) - \mu(x)$
- e.g. exponential $\varrho(x, x') = \sigma^2 e^{-\frac{|x-x'|}{L_c}}$, Gaussian $\varrho(x, x') = \sigma^2 e^{-\frac{(x-x')^2}{L_c}}$, etc.



Remark: The diffusion coefficient can not be a Gaussian field (finite probability becomes negative). Use nonlinear transformations, e.g., lognormal model

$$a(x, \omega) = a_{min} + e^{\gamma(x,\omega)}, \qquad \gamma \sim N(\mu(\cdot), \varrho(\cdot, \cdot)),$$

i.e., a truncated Gaussian random field.

Karhunen-Loève expansion

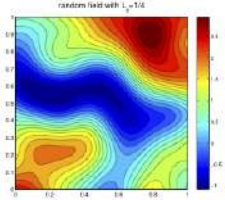
Special case: Gaussian RFs

Although the y_n's are uncorrelated, in general they are not independent:

- Gaussian RVs are uncorrelated $\Longleftrightarrow$ independent, i.e., $\varrho(\boldsymbol{y}) = \prod_{n=1}^{d} \varrho_n(y_n)$.

Gaussian random fields: For every $x \in D$, $a(x,\cdot) \sim N(\mu(x), \varrho(x,x'))$

- Characterized by mean $\mu(x)$ and covariance $\varrho(x,x') = \mathbb{E}[\hat{a}(x,\cdot)\hat{a}(x',\cdot)]$, where $\hat{a}(x,\cdot) = a(x,\cdot) - \mu(x)$
- e.g. exponential $\varrho(x,x') = \sigma^2 e^{-\frac{|x-x'|}{L_c}}$, Gaussian $\varrho(x,x') = \sigma^2 e^{-\frac{(x-x')^2}{L_c}}$, etc.



Remark: The diffusion coefficient can not be a Gaussian field (finite probability becomes negative). Use nonlinear transformations, e.g., lognormal model

$$a(x,\omega) = a_{min} + e^{\gamma(x,\omega)}, \qquad \gamma \sim N(\mu(\cdot), \varrho(\cdot,\cdot)),$$

i.e., a truncated Gaussian random field.

Karhunen-Loève expansion

Stationary non-Gaussian RFs [Grigoriu, 2007]

Let $a(x,\omega)$ be a given stationary non-Gaussian random field with a given (or approximated) marginal CDF F_a, then:

1. one can translation process, i.e., a nonlinear transformation of a stationary Gaussian field $\gamma(x,\omega)$ with zero mean and unit variance:

$$a(x,\omega) = F_a^{-1} \circ \Phi(\gamma(x,\omega)),$$

   where Φ is the CDF of $N(0,1)$.

2. we can approximate $\gamma(x,\omega)$ using a truncated Karhunen-Loève expansion in terms of Gaussian random parameters $\{y_n(\omega)\}_{n=1}^d$ s.t.

$$a_d(x,\omega) = F_a^{-1} \circ \Phi(\gamma_d(x,\omega)) = F_a^{-1} \circ \Phi\left(b_0(x) + \sum_{n=1}^{d} b_n(x) y_n(\omega)\right).$$

Summary of Part I

What have we done and what's coming next?

- Motivation, probabilistic, and stochastic models, aleatoric and epistemic uncertainty
- Formulating a "Plan of attack" for solving stochastic problems
- Random variables and random fields
- Stochastic representation of a random field:
  - Discrete case, Singular value decomposition
  - Spectral expansion, Karhunen-Loève expansion
- What's next? Well-posed (stochastic) parameterized PDEs, regularity, and numerical approximations:
- How to compute a numerical solution $u(x, \boldsymbol{y})$, where $\boldsymbol{y} \in \mathcal{U} \in \mathbb{R}^d$?
  - Monte Carlo FEMs
  - stochastic Galerkin FEMs
  - stochastic collocation FEMs
  - convergence and complexity analysis
- What happens when d becomes large?
  - curse of dimensionality and sparse representations