

Chapter 3

Random Inputs parametrization

This section is devoted to some general aspects related with *input parametrization*. This task is strictly related to the generation of random variables, and the chance to reduce the dimensionality of inputs that are expressed in the form of random fields. Indeed, given an established deterministic model for a physical system, how to properly set up a stochastic model to study the effect of uncertainty in the inputs to the system is a key objective (we will address it in the following).

Prior to any simulation, the key step is to properly characterize the random inputs. More specifically, the goal is to reduce the infinite-dimensional probability space to a finite-dimensional space that is amenable to computing. This is accomplished by parameterizing the probability space by a set of a finite number of random variables. More importantly, it is desirable to require the set of random variables to be mutually independent – the independence requirement is very much a concern from a practical point of view: although there exist some techniques to loosen it, we shall continue to employ this widely adopted requirement. To summarize:

The critical step in formulating a stochastic system is to properly characterize the probability space defined by the random inputs by a set of a finite number of mutually independent random variables. Often, this cannot be done exactly, inducing approximation errors.

In the following subsections we will see how to describe random inputs of a general system, considering first the simple case of a set of parameters, then facing the more general (and complex) case of random fields, that is, spatially-distributed random functions.

there's an issue of dimensionality reduction behind: to treat, e.g., a random field we have to reduce its dimension

3.1 Random parameters

How many parameters?
It may be the case in which the parametrization is redundant (because we haven't discovered yet that among some of the parameters that affect system there's maybe a relationship)

When the random inputs to a system are the system parameters, the parameterization procedure is straightforward, for the inputs are already in the form of parameters.

In the simplest case, the mathematical model depends on a set of p (say, physical) uncertain, mutually independent parameters $\mathbf{Y} = (Y_1, \dots, Y_p) \in \mathbb{R}^p$ that may be represented as a random vector $\mathbf{Y}(\omega) = (Y_1(\omega), \dots, Y_p(\omega)) \in \mathbb{R}^p$ for any outcome $\omega \in \Omega$, with a given joint probability distribution; here Ω denotes the outcome set and $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^p$ denotes a generic P -dimensional random vector. For ease of notation, we will often denote by $\mathbf{y} = (y_1, \dots, y_p) \in \mathbb{R}^p$ the random input, treating it as a variable which the solution depends on.

We denote the range of Y_i by $\Gamma_i = Y_i(\Omega) \subset \mathbb{R}$ and take $\Gamma = \prod_{i=1}^p \Gamma_i$. Since the parameters are assumed to be mutually independent, the joint density $f_{\mathbf{Y}} : \Gamma \rightarrow \mathbb{R}$ factorizes as follows:

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^p f_i(y_i)$$

where each parameter $Y_i(\omega) : \Omega \rightarrow \mathbb{R}$ has an associate density $\rho_i : \Gamma_i \rightarrow \mathbb{R}$. Hence, any Monte Carlo method will exploit the sampling techniques seen in the previous sections.

The more important issue is then to identify the independent parameters in the set. The assumption of mutually independent parameters is very often required, unless the joint density can be directly constructed. This assumption also underlies many sampling methods. Unfortunately, parameters are often correlated and hence dependent; in this case, suitable transformations must be sought to obtain a new set of independent parameters.

The problem can be stated as follows: let $\mathbf{Y} = (Y_1, \dots, Y_n)$, $n > 1$, be the system parameters with a prescribed distribution function $F_{\mathbf{Y}}(\mathbf{y})$, and find a set of mutually independent random variables $\mathbf{Z} = (Z_1, \dots, Z_p) \in \mathbb{R}^p$, where $1 \leq p \leq n$, such that $\mathbf{Y} = T(\mathbf{Z})$ for a suitable transformation function T .

- **Example 3.1.1.** Consider an ordinary differential equation with two random parameters,

$$\begin{cases} u'(t, \omega) = -\alpha(\omega)u, & t \in (0, T) \\ u(0, \omega) = \beta(\omega) \end{cases}$$

where the rate constant α and the initial condition β are assumed to be random. Thus, the input random variables are $\mathbf{Y}(\omega) = (\alpha, \beta) \in \mathbb{R}^2$. If α and β are mutually independent, then we simply let $\mathbf{Z}(\omega) = \mathbf{Y}(\omega)$. The solution $u(t, \omega) : [0, T] \times \Omega \rightarrow \mathbb{R}$ can now be expressed as $u(t, \mathbf{Z}) : [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}$, which has one time dimension and two random dimensions. If α and β are not independent of each other, it implies that there exists a function f such that $f(\alpha, \beta) = 0$. Then it is possible to find a random variable $Z(\omega)$ to parameterize the relation such that

$$\alpha(\omega) = a(Z(\omega)), \quad \beta(\omega) = b(Z(\omega)),$$

and $f(a, b) = 0$. Or, equivalently, the dependence between α and β implies that there exists a function g such that $\beta = g(\alpha)$. Then we can let $Z(\omega) = \alpha(\omega)$ and $\beta(\omega) = g(Z(\omega))$. Which approach is more convenient in practice is problem-dependent. Nevertheless, the random inputs via α and β can now be expressed via a single random variable Z , and the solution becomes $u(t, Z) : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$, which now has only one random dimension.

In practice, when there are many parameters in a system, finding the exact form of the functional dependence among all the parameters can be a challenging (and unnecessary) task. This is especially true when the only available information is the (joint) probability distributions of the parameters. The goal now is to transform the parameters to a set of independent random parameters by using their distribution functions.

Remark 3.1.1. For instance, if parameters are normally distributed (and dependent), the Cholesky decomposition of the covariance matrix provides such a strategy. Indeed, let $\mathbf{Y} = (Y_1, \dots, Y_n) \sim N(\mathbf{0}, C)$, where $C \in \mathbb{R}^{n \times n}$ is the covariance matrix and the expectation is assumed to be zero (without loss of generality). Let $\mathbf{Z} \sim N(\mathbf{0}, I)$, where $I \in \mathbb{R}^{n \times n}$ is the $n \times n$ identity matrix, be a uncorrelated Gaussian vector of size n . Thus, the components of \mathbf{Z} are mutually independent. If $A \in \mathbb{R}^{n \times n}$, then $A\mathbf{Z} \sim N(\mathbf{0}, AA^\top)$. Therefore, if one finds a matrix A such that $AA^\top = C$, then $\mathbf{Y} = A\mathbf{Z}$ will have the given distribution $N(\mathbf{0}, C)$. Since C is real and symmetric, solving the problem $AA^\top = C$ can be readily done via, for example, Cholesky's decomposition.

3.2 Random fields and dimensionality reduction

In many cases, the random inputs are stochastic processes. For example, the inputs could be a time-dependent random forcing term that is a stochastic process in time, or an uncertain material property, e.g., conductivity, that is a stochastic process in space.

The parameterization problem can be stated as follows: let $\{Y_t, t \in I\}$ be a stochastic process that models the random inputs, where t is the index belonging to an index set I , and find a suitable transformation function R such that $Y_t = R(\mathbf{Z})$, where $\mathbf{Z} = (Z_1, \dots, Z_d)$, $d \geq 1$, are mutually independent. Note that the index set I can be in either the space domain or the time domain and is usually an infinite-dimensional object.

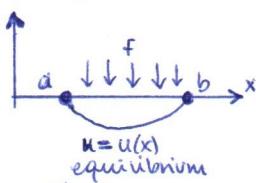
Since we require d to be a finite integer, the transformation cannot be exact. Therefore, $Y_t \approx R(\mathbf{Z})$ in a proper norm or metric, and the accuracy of the approximation will be problem-dependent.

Remark 3.2.1. A straightforward approach is to consider the finite-dimensional version of Y_t instead of Y_t directly. This requires one to first discretize the index domain I into a set of finite indices and then study the process $(Y_{t_1}, \dots, Y_{t_n})$, $t_1, \dots, t_n \in I$, which is now a finite-dimensional random vector. The discretization of Y_t into its finite-dimensional version is obviously an approximation: the finer the discretization, the better the approximation. However, this is not desired because a finer discretization leads to a larger dimension of n and can significantly increase the computational burden. Some kinds of dimension reduction techniques are required to keep the dimension as low as possible while maintaining a satisfactory approximation accuracy.

3.2.1 A more formal definition of random field

(general random fields, not only gaussians)

for example, 1D structure of an elastic cable (no mass) with the two extremities fixed. There are some forces \mathbf{f} on the cable and it all forms an equilibrium position $u = u(x)$:



equations:

$$\begin{cases} -(\alpha(x; \omega) u'(x; \omega))' = f(x; \omega) \\ u(a, \omega) = 0 \\ u(b, \omega) = 0 \end{cases} \quad \text{boundary}$$

with $x \in (a, b)$.

$\alpha(x; \omega)$ are properties of the material and they're not constant.

$\mathbf{f}(x; \omega)$ is a load (by unit of mass) that is not completely known (e.g. a bridge, how many cars are passing now?). As soon as we fix ω we have all functions of x , however if we fix x we have random

variables (one random variable for each x) (since their values depend on the outcome ω). Hence both $\mathbf{f}(\cdot; \omega)$ and $\alpha(\cdot; \omega)$ are random fields.

Whenever random inputs are spatially-dependent (or time-dependent) random quantities – so that they may vary randomly from one point of the physical domain D to another (or from one time instant to another) – input uncertainty has to be described in terms of random fields. Since we will more often use random fields to represent physical properties defined over a spatial domain, here we explicitly denote by $x \in D \subset \mathbb{R}^d$, $d \geq 1$, the (space) variable used to index the random field – we used $t \in I$ in Section 2.6, but the conclusions are indeed equivalent.

Focusing on the spatially-dependent, scalar case (think, for instance, to the conductivity coefficient of a given material), a random field $a = a(x; \omega)$ is a stochastic process taking values in \mathbb{R} for any $x \in D$, for any outcome $\omega \in \Omega$, meaning that:

1. for a fixed point $x \in D$, $a(x; \cdot)$ is a random variable over Ω ;
2. for a fixed $\omega \in \Omega$, $a(\cdot; \omega)$ is a realization of the random field over the physical domain D .

Let $D \subset \mathbb{R}^d$ be a physical domain. A random field $a(x, \omega) : D \times \Omega \rightarrow \mathbb{R}$ is a collection of infinite random variables: $a(x, \omega)$ is a random variable for each $x \in D$.

We denote by:

- $\bar{a}(x) = \mathbb{E}[a(x, \cdot)] : D \rightarrow \mathbb{R}$ its expected value;
- $\text{Cov}_a(x_1, x_2) : D \times D \rightarrow \mathbb{R}$ its covariance function, defined as

$$\text{Cov}_a(x_1, x_2) = \mathbb{E}[(a(x_1, \cdot) - \bar{a}(x_1))(a(x_2, \cdot) - \bar{a}(x_2))];$$

the covariance is a two-point statistics that models the spatial correlation of the stochastic field between two points $x_1, x_2 \in D$;

- $\text{Var}_a(x) = \text{Cov}_a(x, x) : D \rightarrow \mathbb{R}$ its variance. We say that $a(x, \omega)$ is a second order random field if $\text{Var}_a(x) < \infty$, for all $x \in D$.

Usually we'll give the expected value and the covariance function to fully describe a random field

Definition 3.2.1. A random field $a(x, \omega) : \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$ is said to be (weakly) stationary if its law is invariant under translation, that is, $a(x + h, \omega) \sim a(x, \omega) \forall h \in \mathbb{R}$. For a (weakly) stationary field, $\mathbb{E}[a] = \mu \in \mathbb{R}$ (independent of x) and

$$\text{Cov}_a(x_1, x_2) = \text{Cov}_a(0, x_2 - x_1) = \widetilde{\text{Cov}}_a(x_2 - x_1).$$

Note that while Cov_a is defined on $D \times D$, $\widetilde{\text{Cov}}_a$ is defined on D only.

Definition 3.2.2. A weakly stationary random field is said isotropic if the covariance function depends only on $\|x_1 - x_2\|$, that is,

$$\text{Cov}_a(x_1, x_2) = \widetilde{\text{Cov}}_a(\|x_1 - x_2\|).$$

The covariance function is indeed the most relevant feature of a random field. For a second order random field the covariance function is:

1. *bounded* if $\text{Cov}_a(x_1, x_2) \leq \sqrt{\text{Var}_a(x_1)} \sqrt{\text{Var}_a(x_2)}$;
2. *symmetric* if $\text{Cov}_a(x_1, x_2) = \text{Cov}_a(x_2, x_1)$;
3. *semi-positive definite* if for any $\xi_1, \dots, \xi_N \in D$, the matrix $(C)_{ij} = \text{Cov}_a(\xi_i, \xi_j)$ is semi-positive definite.

In particular, continuity and differentiability of a random field can be deduced by the same properties of the covariance function:

- a random field $a : \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$ is said to be *mean square continuous* at $\bar{x} \in D$ if

$$\lim_{x \rightarrow \bar{x}} \mathbb{E}[(a(x) - a(\bar{x}))^2] = 0.$$

A centered (=zero mean) random field $a : \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$ is mean square continuous at $\bar{x} \in D$ if and only if its covariance function $\text{Cov}_a(x_1, x_2)$ is continuous at $x_1 = x_2 = \bar{x}$.

- a random field $a : \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$ is said to be *mean square differentiable* at $\bar{x} \in D$ if

$$\frac{\partial}{\partial x} a(\bar{x}, \cdot) = \lim_{h \rightarrow 0} \frac{a(\bar{x} + h, \cdot) - a(\bar{x}, \cdot)}{h}$$

exists in the mean square sense. A centered random field $a : \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$ is mean square differentiable at $\bar{x} \in D$ if and only if $\partial^2_{x_1 x_2} \text{Cov}_a(x_1, x_2)$ exists and is finite at $x_1 = x_2 = \bar{x}$.

3.2.2 Karhunen-Loeve expansion

basically this is the Principal Component Analysis but on stochastic processes (the idea is exactly the same)

takes a function and gives another function (of variable x)

$$(T\psi)(x) = \int_D \text{Cov}_a(x, y)\psi(y)dy.$$

Then, the Mercer's theorem¹ (see below) ensures that

There exists a sequence of values $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq \dots \geq 0$ with $\lim_{k \rightarrow \infty} \lambda_k = 0$ and a corresponding sequence of functions $b_i(x) : D \rightarrow \mathbb{R}$ such that

$$(Tb_i)(x) = \int_D \text{Cov}_a(x, y)b_i(y)dy = \lambda_i b_i(x) \quad \text{and} \quad \int_D b_i(x)b_j(x)dx = \delta_{ij}. \quad (3.1)$$

¹ Mercer's Theorem is actually an analogue of an Eigenvalue or Singular Value Decomposition. The only difference here is that the kernel is now an object contained in an infinite-dimensional space.

$\text{Cov}_a(x, y)$

Copyright © Andrea Manzoni, 2020

the eigenvectors are orthogonal w.r.t. the scalar product

somewhat an eigenvalue/vector decomposition
 $(\lambda_i = \text{eigenvalues},$
 $b_i(x) = \text{eigenfunctions}$
 $(\text{extremes of eigenvectors}))$

SVD - Singular Value Decomposition

$A \in \mathbb{R}^{m \times n}$ real (approximate)

$\Rightarrow \exists U, Z$ orthogonal matrices : $U = [\vec{u}_1 | \dots | \vec{u}_m] \in \mathbb{R}^{m \times m}$

$\boxed{\begin{aligned} A &\text{ orthogonal} \\ \Leftrightarrow A^T A &= A A^T = I \\ \Leftrightarrow A^{-1} &= A^T \end{aligned}}$

$([A]_{ij} \in \mathbb{R} \forall i, j)$

$Z = [\vec{z}_1 | \dots | \vec{z}_n] \in \mathbb{R}^{n \times n}$

s.t.
$$A = U \Sigma Z^T \quad \text{SVD of } A$$

with $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n}$

$$p = \min\{m, n\}$$

and: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$

\downarrow
singular values of A

Then $\{\vec{u}_1, \dots, \vec{u}_m\}$ are left singular vectors,
 $\{\vec{z}_1, \dots, \vec{z}_n\}$ are right singular vectors

$$\boxed{A} \underset{m \times n}{=} \boxed{U} \underset{m \times m}{=} \boxed{\begin{matrix} 0 & 0 \\ 0 & \Sigma \\ 0 & 0 \end{matrix}} \underset{m \times n}{=} \boxed{Z^T} \underset{n \times n}{=}$$

* where the FROBENIUS norm of a matrix:
 $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$

$$\boxed{A} \underset{m \times n}{=} \boxed{U} \underset{m \times m}{=} \boxed{\begin{matrix} 0 & & \\ & \Sigma & \\ & & 0 \end{matrix}} \underset{m \times n}{=} \boxed{Z^T} \underset{n \times n}{=}$$

$A A^T$ and $A^T A$ are symmetric (and squared)

$\Rightarrow \sigma_i(A) = \sqrt{\lambda_i(A^T A)}$
 eigenvalues of $(A^T A)$

$$\|A\|_2 = \sigma_{\max} \quad \forall A \in \mathbb{R}^{m \times n}$$

$$\|A\|_F = \sqrt{\sum_{i=1}^p \sigma_i^2}$$

*

Low-rank approximations :

We can compute a low-rank apprx of a given matrix ~~A~~ by using the SVD :
(lower-rank)

If $A \in \mathbb{R}^{m \times n}$ has rank r ($r \leq \min\{m, n\}$), thanks to SVD:

$$A = \sum_{i=1}^r \sigma_i \vec{u}_i \vec{z}_i^T \quad (*)$$

\vec{u}_i : left singular vector
 \vec{z}_i : right singular vector

EXACT

A can be written as the sum of rank 1 matrices.

If we cut before ~~$\vec{u}_{k+1}, \dots, \vec{u}_n$~~ r we get a low-rank approximation.
(This is what a jpeg algorithm does to images).

The sum of $k \leq r$ terms in $(*)$ capture as much as the "energy" of A as possible:

$$A_k = \sum_{i=1}^k \sigma_i \vec{u}_i \vec{z}_i^T \quad \text{among all matrices of rank } k \text{ it's the best one.}$$

("best" in the sense of SCHMIDT - ECKART - YOUNG theorem:

given $A \in \mathbb{R}^{m \times n}$ of rank $r \geq A_k = \sum_{i=1}^k \sigma_i \vec{u}_i \vec{z}_i^T$ if s.t.

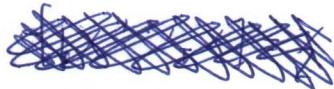
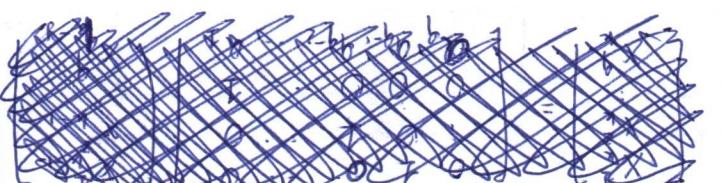
$$\|A - A_k\|_F = \min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B) \leq k \\ (\text{with rank at most } k)}} \|A - B\|_F = \sqrt{\sum_{i=k+1}^r \sigma_i^2}$$

we can quantify the error (in the approx.)



$\|\cdot\|_2$ very similar:

$$\|A - A_k\|_2 = \min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B) \leq k}} \|A - B\|_2 = \sigma_{k+1}$$



KARHUNEN-LOEVE EXPANSION is based on SVD
(it is actually the ∞ -dimensional counterpart of SVD)

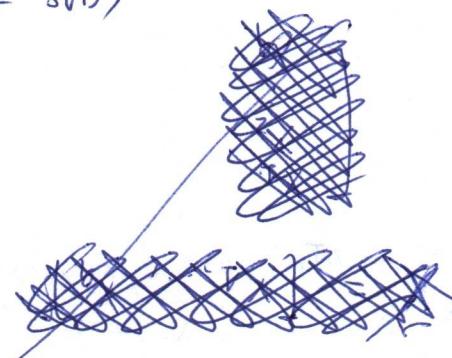


Image compression

- Given a grayscale image having $m \times n$ pixels, each pixel is an integer number $a_{ij} \in [0, 255]$ representing the intensity varying from black (255) to white (0).
- Performing SVD of the intensity matrix $A = (a_{ij})$ allows to compress the original image.
- Indeed, a compressed image (at level k) corresponds to the best approximation of rank k to the original matrix, which is represented by matrix A_k of Schmidt-Eckart-Young Theorem.
- The rank k compressed image requires storage of only $(m+n+1)r$ integers instead of mn integers.

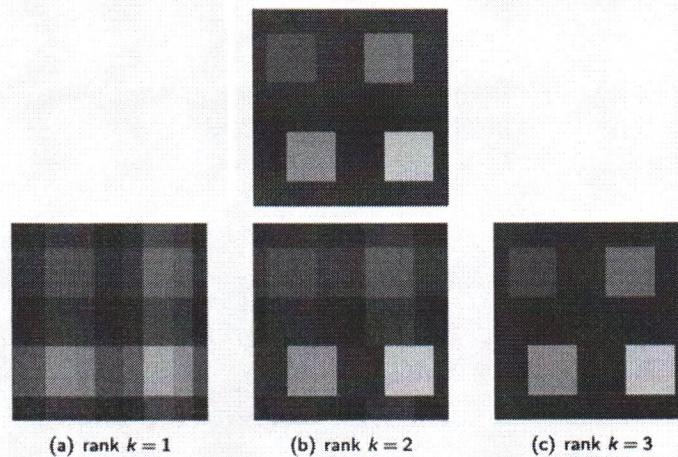


Figure: Example of a grayscale image and its low-rank approximations of rank $r = 1, 2, 3$

A has rank $k = 3$: the low-rank approximation of rank $k = 3$ of the matrix yields an image which is exactly equal to the original picture

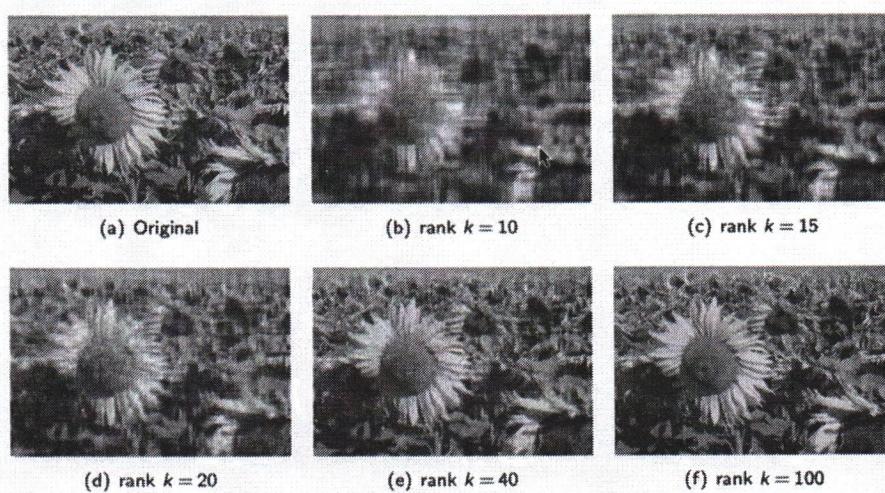


Figure: Compression of a grayscale image by low-rank approximation

Here the compression is more difficult: singular values decrease much slower than before.

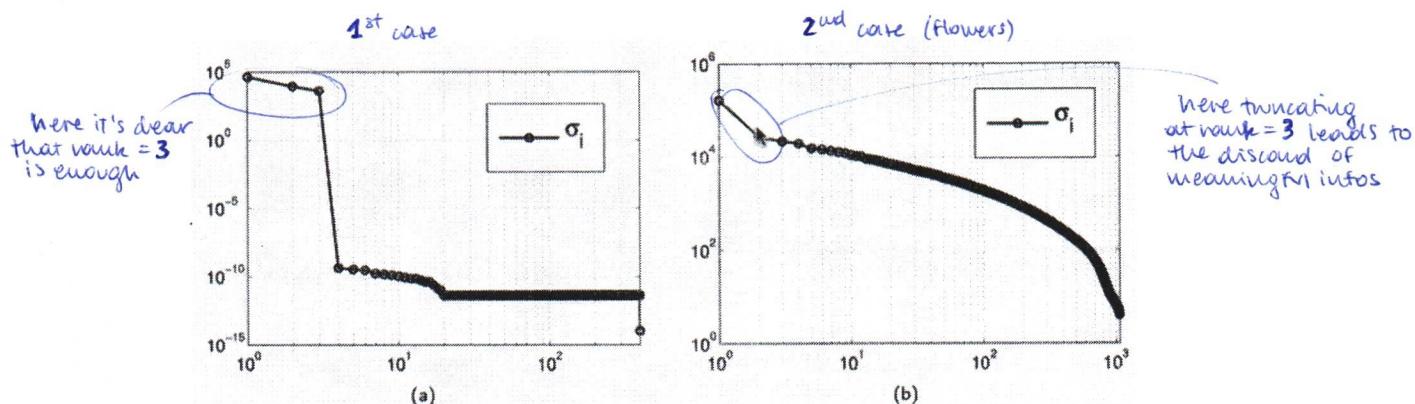


Figure: Singular values of the two images

Hence, b_i are the orthogonal eigenfunctions and λ_i are the corresponding eigenvalues of the eigenvalue problem (3.1). If we now define the sequence of random variables $Y_i(\omega)$, $i = 1, 2, \dots$

$$Y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (a(x, \omega) - \mathbb{E}[a](x)) b_i(x) dx$$

which are uncorrelated with zero mean and unit variance, that is,

$$\mathbb{E}[Y_i] = 0, \quad \mathbb{E}[Y_i Y_j] = \delta_{ij},$$

random objects since $a(x, \omega)$ is a random field

we're projecting the field (for which we subtract the mean) on each direction b_i

then the random field $a(x, \omega)$ can be represented as the infinite series, also called

Karhunen-Loeve expansion

$$a(x, \omega) = \mathbb{E}[a](x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} b_i(x) Y_i(\omega). \quad (3.2)$$

if we start from a gaussian random field
uncorrelated means
independent
→ if $a(x, \omega)$ is a gaussian random field
we can sample from each $Y_i(\omega)$ independently

Remark 3.2.2. The same approach will work if D is replaced by a finite set of points. Then the integral in (3.1) becomes a finite sum, as does the expansion (3.2), and eigenfunctions become eigenvectors. Thus, even if no analytic solution to problem (3.1) is possible, we now have an eigenvector problem that will be easy to solve numerically, as long as the number of points is not too large. Whether we are in the finite (or infinite) scenario, in order to simulate $a(x, \omega)$ via the Karhunen-Loeve expansion, one needs only once to solve a deterministic eigenvector (or eigenfunction) problem, and then simulate nothing more difficult than a set of i.i.d. random variables.

The Karhunen-Loeve expansion is a consequence of the following Mercer's theorem:

Theorem 3.2.1. (Mercer) Let $D \subset \mathbb{R}^d$ be a compact domain and $K : D \times D \rightarrow \mathbb{R}$ a Mercer's kernel, that is:

1. K is symmetric: $K(x, y) = K(y, x)$ for all $x, y \in D$;
2. K is continuous;
3. K is semi-positive definite.

Define, moreover, the so-called Hilbert-Schmidt (compact) operator $T_K : L^2(D) \rightarrow L^2(D)$ as

$$T_K f(x) = \int_D K(x, y) f(y) dy \quad \forall f \in L^2(D).$$

Then, there exists an orthonormal basis $\{b_i\}_i$ of $L^2(D)$ consisting of eigenfunctions of T_K such that the corresponding sequence of eigenvalues $\{\lambda_i\}_i$ is non-negative. The eigenfunctions corresponding to non-zero eigenvalues are continuous in D and

$$K(x, y) = \sum_{i=1}^{\infty} \lambda_i b_i(x) b_i(y)$$

where the convergence is absolute and uniform, that is

$$\lim_{n \rightarrow \infty} \sup_{x, y \in D} \left| K(x, y) - \sum_{i=1}^n \lambda_i b_i(x) b_i(y) \right| = 0.$$

Moreover, under the above assumptions (D compact and Cov_a continuous) it holds

$$\lim_{N \rightarrow \infty} \sup_{x \in D} \mathbb{E} \left[\left(a(x, \cdot) - \mathbb{E}[a](x) - \sum_{i=1}^N \sqrt{\lambda_i} b_i(x) Y_i(\cdot) \right)^2 \right] = 0.$$

Copyright © Andrea Manzoni, 2020

$a_N(x, \omega)$

N-terms truncated KL expansion

$$\lim_{N \rightarrow \infty} \sup_{x \in D} \mathbb{E}[(a(x, \cdot) - a_N(x, \cdot))^2] = 0$$

→ in terms of variance this means that the KL expansion is the best representation of the random field consisting of N terms

The KL expansion is the best N -terms approximation in terms of variance, that is,

$$\{Y_n, b_n\}_{n=1}^N = \arg \min_{\xi_n, \psi_n : \int_D \psi_n \psi_m = \delta_{mn}} \mathbb{E} \left[\int_D \left(a(x, \cdot) - \mathbb{E}[a](x) - \sum_{i=1}^N \xi_i(\cdot) \psi_i(x) \right)^2 dx \right].$$

Moreover, the convergence rate of the N -term truncation

$$a_N(x, \omega) = \mathbb{E}[a](x) + \sum_{n=1}^N \sqrt{\lambda_n} b_n(x) Y_n(\omega) \quad (3.3)$$

depends on the decay of the eigenvalues λ_n which, in turns, depends on the smoothness of the covariance function (and the correlation length of the process). One typically chooses N so that the sum of the neglected terms is sufficiently small compared with the sum of the first N terms. **Hands-On Problems 3.1-3.2** focus on a possible Matlab implementation.

- **Example 3.2.1.** Let $\text{Cov}_a(x, y) = \exp(-|y - x|/a)$, where $a > 0$ is the correlation length, and let $x \in D = (-b, b)$ be in a bounded domain with length $2b$. Then the eigenvalue problem can be solved analytically, and the eigenvalues are

$$\lambda_i = \begin{cases} \frac{2a}{1+a^2 w_i^2} & \text{if } i \text{ is even} \\ \frac{2a}{1+a^2 v_i^2} & \text{if } i \text{ is odd} \end{cases}$$

and the corresponding eigenfunctions are

$$b_i(x) = \begin{cases} \frac{\sin(w_i x)}{\sqrt{b - \frac{\sin(2w_i b)}{2w_i}}} & \text{if } i \text{ is even} \\ \frac{\cos(w_i x)}{\sqrt{b + \frac{\sin(2v_i b)}{2v_i}}} & \text{if } i \text{ is odd} \end{cases}$$

where w_i and v_i are the solutions of the equations

$$\begin{cases} aw + \tan(wb) = 0 & \text{if } i \text{ is even} \\ 1 - av \tan(vb) = 0 & \text{if } i \text{ is odd.} \end{cases}$$

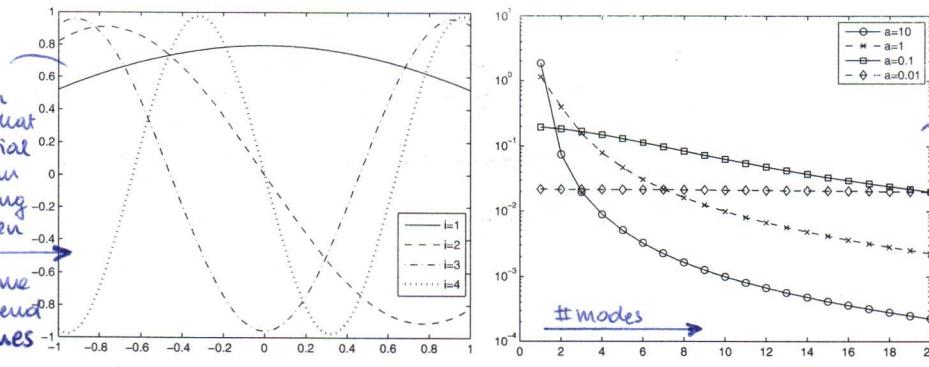
In Figure 3.1, left, the first four eigenfunctions are shown for the exponential covariance function in $(-1, 1)$. It is obvious that the higher modes (the eigenfunctions with larger index i) have a finer structure compared to the lower modes. The eigenvalues are shown in Figure 3.1, right for several different correlation lengths a . It can be seen that the eigenvalues decay, and the decay rate is larger when the correlation length is longer. When the correlation length is very small, e.g., $a = 0.01$, the decay of the eigenvalues is barely visible.

- **Example 3.2.2.** The limit of diminishing correlation length is the zero correlation case, when the covariance function takes the form $\text{Cov}_a(x, y) = \delta(x - y)$. Now any orthogonal functions can be the eigenfunctions and the eigenvalues are constant, i.e. $\lambda_n = 1$ for all n . Hence, the eigenvalues do not decay.

we can't predict what happens in any point, no matter how close we are

The other limit is when $\text{Cov}_a(x, y) = 1$, which implies an infinite correlation length and that the field a is fully correlated. This is the rather trivial case where the process depends on just one random variable. In this case, it is possible to show that there exists one nonzero eigenvalue corresponding to a constant eigenfunction and that the rest of the eigenvalues are zero, that is, $\lambda_n = 0$ for $n > 1$.

Remark 3.2.3. The aforementioned examples illustrate an important property of the KL expansion: for a given covariance function, the decay rate of the eigenvalues depends inversely on the correlation length. Long correlation length implies that the process is strongly correlated and results in a fast decay of the eigenvalues. The limit of this, infinitely long correlation length, is the fully



This are the modes. since we have a random field, which is an object that depends on both the spatial variable and the random input, we'll have something like a linear combination of spatial function (function of x) times some coefficient that will depend on the eigenvalues times some coefficients that will depend on random inputs. At the end of the day we'll only need to sample from those random inputs to get the expression of the random field. (At least the low-rank approximation or the finite dimension counterpart)

Figure 3.1: Left: the first four eigenfunctions of the exponential covariance function. Right: The first 20 eigenvalues of the exponential covariance function with different correlation lengths a .

correlated case where the eigenvalues decay to zero immediately. Conversely, a weakly correlated process has short correlation length and results in a slow decay of the eigenvalues. The limit of this, the uncorrelated process with zero correlation length, has no eigenvalue decay.

Proof. (KL expansion). Let $D \subset \mathbb{R}^d$ be compact and $a : D \times \Omega \rightarrow \mathbb{R}$ a random field with continuous covariance Cov_a . Then, Cov_a is a Mercer's kernel (symmetric, continuous and semi-positive definite) and by Mercer's theorem

$$\text{Cov}_a(x, y) = \sum_{i=1}^{\infty} \lambda_i b_i(x) b_i(y) \quad \Rightarrow \quad \text{Var}_a(x) = \text{Cov}_a(x, x) = \sum_{i=1}^{\infty} \lambda_i b_i(x)^2.$$

Let us now define the truncated KL expansion

$$a_N(x, \omega) = \mathbb{E}[a](x) + \sum_{i=1}^N \sqrt{\lambda_i} b_i(x) Y_i(\omega)$$

with

$$Y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (a(x, \omega) - \mathbb{E}[a](x)) b_i(x) dx.$$

We remark that

$$\mathbb{E}[Y_i(\cdot)] = \frac{1}{\sqrt{\lambda_i}} \mathbb{E} \left[\int_D (a(x, \cdot) - \mathbb{E}[a](x)) b_i(x) dx \right] = \frac{1}{\sqrt{\lambda_i}} \int_D \mathbb{E}[a(x, \cdot) - \mathbb{E}[a](x)] b_i(x) dx = 0$$

and that

$$\begin{aligned} \mathbb{E}[Y_i(\cdot) Y_j(\cdot)] &= \frac{1}{\sqrt{\lambda_i \lambda_j}} \mathbb{E} \left[\int_D \int_D (a(x, \cdot) - \mathbb{E}[a](x))(a(y, \cdot) - \mathbb{E}[a](y)) b_i(x) b_j(y) dx dy \right] \\ &= \frac{1}{\sqrt{\lambda_i \lambda_j}} \int_D \int_D \mathbb{E}[(a(x, \cdot) - \mathbb{E}[a](x))(a(y, \cdot) - \mathbb{E}[a](y))] b_i(x) b_j(y) dx dy \\ &= \frac{1}{\sqrt{\lambda_i \lambda_j}} \int_D b_i(x) \left(\int_D \text{Cov}_a(x, y) b_j(y) dy \right) dx = \frac{1}{\sqrt{\lambda_i \lambda_j}} \int_D b_i(x) \lambda_j b_j(x) dx = \delta_{ij}. \end{aligned}$$

Moreover,

$$\begin{aligned} \text{Var}_{a_N}(x) &= \text{Cov}_{a_N}(x, x) = \mathbb{E}[(a_N(x, \omega) - \mathbb{E}[a](x))(a_N(y, \omega) - \mathbb{E}[a](y))] \\ &= \mathbb{E} \left[\sum_{i=1}^N \sqrt{\lambda_i} b_i(x) Y_i(\cdot) \sum_{j=1}^N \sqrt{\lambda_j} b_j(x) Y_j(\cdot) \right] = \sum_{i,j=1}^N \sqrt{\lambda_i \lambda_j} \mathbb{E}[Y_i(\cdot) Y_j(\cdot)] b_i(x) b_j(x) \\ &= \sum_{i,j=1}^N \sqrt{\lambda_i \lambda_j} \delta_{ij} b_i(x) b_j(x) = \sum_{i=1}^N \lambda_i b_i(x)^2. \end{aligned}$$

Decay of the eigenvalues (correlated to how accurate will be the low-dimensional representation) depending on the "correlation length" (a). If a is small we can recover information on what happens in a point as soon as we know what happen close to that point. If a is large, even if we have informations far from a point we can still recover information on the point. Very large correlation lengths means that the random field will be very regular (we know what happens in points from which we are far from), very small correlation length lead to non-reducibility in terms of dimensionality and we need to consider very high number of modes.

Let us now set $a'(x, \omega) = a(x, \omega) - \mathbb{E}[a](x)$ and $a'_N(x, \omega) = a_N(x, \omega) - \mathbb{E}[a](x)$ observe that, from the definition of $Y_i(\omega)$,

$$\sqrt{\lambda_i} Y_i(\omega) = \int_D a'(x, \omega) b_i(x) dx,$$

so that

$$\begin{aligned}\mathbb{E}[a'(x, \cdot) a'_N(x, \cdot)] &= \sum_{i=1}^N \sqrt{\lambda_i} \mathbb{E}[a'(x, \cdot) Y_i(\cdot)] b_i(x) = \sum_{i=1}^N \mathbb{E}[\int_D a'(x, \cdot) a'(z, \cdot) b_i(z) dz] b_i(x) \\ &= \sum_{i=1}^N b_i(x) \int_D \text{Cov}_a(x, z) b_i(z) dz = \sum_{i=1}^N \lambda_i b_i(x)^2 = \text{Var}_{a_N}(x).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}[(a'(x, \cdot) - a'_N(x, \cdot))^2] &= \mathbb{E}[(a'(x, \cdot))^2] + \mathbb{E}[(a'_N(x, \cdot))^2] - 2\mathbb{E}[a'(x, \cdot) a'_N(x, \cdot)] \\ &= \text{Var}_a(x) + \text{Var}_{a_N}(x) - 2\text{Var}_{a_N}(x) = \text{Var}_a(x) - \text{Var}_{a_N}(x) \\ &= \sum_{i=N+1}^{\infty} \lambda_i b_i(x)^2 \xrightarrow{N \rightarrow \infty} 0, \text{ uniformly in } x \in D.\end{aligned}$$

□

Remark 3.2.4. The total variance of the field/process, defined as the integral of $\text{Var}_a(x)$ over the index set D , is

since the conv. is uniform we can exchange ∫ and Σ

$$\int_D \text{Var}_a(x) dx = \int_D \sum_{i=1}^{\infty} \lambda_i b_i(x)^2 dx = \sum_{i=1}^{\infty} \int_D \lambda_i b_i(x)^2 dx = \sum_{i=1}^{\infty} \lambda_i \int_D b_i(x)^2 dx = \sum_{i=1}^{\infty} \lambda_i$$

since the convergence of the sum is uniform by Mercer's theorem. As a result, if the KL expansion gets truncated, we have that the N -truncated expansion explains the $\sum_{k=1}^N \lambda_k / \sum_{k=1}^{\infty} \lambda_k$ of the total variance of the field/process.

The truncated KL series (3.3) provides a way to approximate a random process by a function (a series) involving a finite number of random variables. The set of random variables $Y_i(\omega)$ are uncorrelated. For Gaussian random variables, uncorrelation and independence are equivalent. Furthermore, linear combinations of Gaussian random variables remain Gaussian-distributed. Therefore, if $a(x, \omega)$ is a Gaussian process, then the random variables Y_i in (3.3) are independent Gaussian random variables. Hence (3.3) provides a natural way to parameterize a Gaussian process by a finite number of independent Gaussian random variables.

When the input random processes are non-Gaussian, their parameterization and dimension reduction are significantly more challenging. The main problem is that, for non-Gaussian distributions, uncorrelation of the random variables Y_i does not imply independence. Hence the KL expansion does not provide a way of parameterization with independent variables. In many practical computations, one often still uses the KL expansion for an input process and then further assumes that the Y_i are independent – though this is not a rigorous approach.

- KL expansions represent a random field for which only the mean and covariance functions are known in terms of parameters; having a parametric representation means that we can evaluate a realization of the field by choosing values for the random parameters;
- the random parameters in KL expansions are uncorrelated; in the case of a Gaussian random field, they are also independent, hence, we can parameterize a Gaussian process by a finite number of independent Gaussian random variables.

the parameters are
 $Y_i(\omega) \quad i=1, \dots, N$

3.2.3 Back to the case of Gaussian random fields

Recall that a random field $a : D \times \Omega \rightarrow R$ is Gaussian if all finite dimensional distributions are Gaussian. i.e. for any $x_1, \dots, x_n \in D \subset \mathbb{R}^d$, $d \geq 1$, the random vector $\mathbf{Y}(\omega) = (a(x_1, \omega), \dots, a(x_n, \omega))$ has a multivariate Gaussian distribution. If $d = 1$, a Gaussian random field is also called a Gaussian process (GP). Gaussian random fields are extremely regular random fields². If we consider the KL expansion

$$a(x, \omega) = E[a](x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} b_i(x) Y_i(\omega).$$

we have that Y_i are independent $N(0, 1)$ random variables. In other words, the KL expansion of a Gaussian field involves random variables $Y_i(\omega)$ which are not only uncorrelated, but also independent. The latter is a useful property in practical applications, since it makes the KL expansions very useful for generating samples of Gaussian stochastic processes; for instance, it is extensively used in the stochastic Galerkin and collocation methods for PDEs with random inputs.

We now provide some examples of covariance models, denoting by σ^2 the variance of the field, and by l_c the correlation length, and some information about the regularity of the resulting Gaussian random field:

- *exponential*

$$\text{Cov}_a(\|x - y\|) = \sigma^2 e^{-\frac{\|x-y\|}{l_c}}$$

smoothness of the resulting Gaussian field: almost surely $C^{0,\alpha}$, $\alpha < 1/2$ (the same regularity as a Brownian motion);

- *squared exponential (Gaussian)*

$$\text{Cov}_a(\|x - y\|) = \sigma^2 e^{-\frac{\|x-y\|^2}{2l_c^2}}$$

smoothness of the resulting Gaussian field: analytic almost surely.

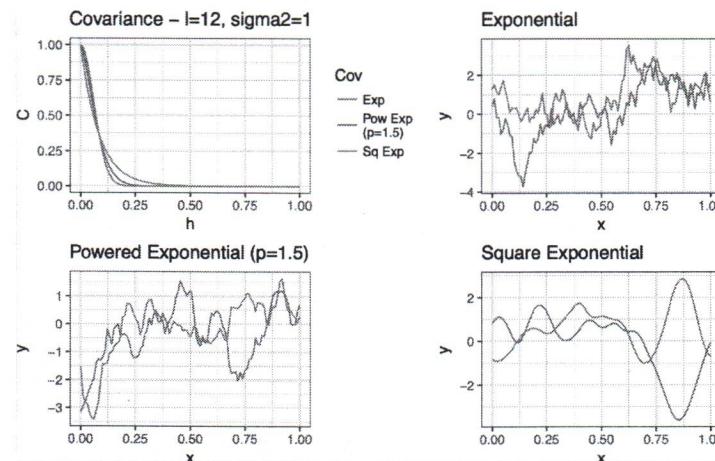


Figure 3.2: Exponential and squared exponential kernel functions, with realizations of resulting Gaussian fields ($d = 1$)

²For a centered Gaussian random field, mean square integrability implies integrability in L^p , for any $p > 0$. Not only, for Gaussian fields mean square differentiability implies almost sure sample path continuity.

- Matérn

$$\text{Cov}_a(\|x - y\|) = \sigma^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu} \frac{\|x - y\|}{l_c} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{\|x - y\|}{l_c} \right)$$

where ν is a smoothness parameter, Γ is the gamma function, K_ν is the modified Bessel function of the second kind.

A Gaussian field with Matérn covariance has sample functions that are $\lceil \nu - 1 \rceil$ times differentiable. When $\nu = 1/2 + p$ for $p \in \mathbb{N}_+$ then the Matérn kernel has a simplified form (product of an exponential and a polynomial of order p). When $p = 1/2$, the Matérn kernel is equivalent to the exponential kernel; as $\nu \rightarrow \infty$, the Matérn kernel converges to the squared exponential covariance.

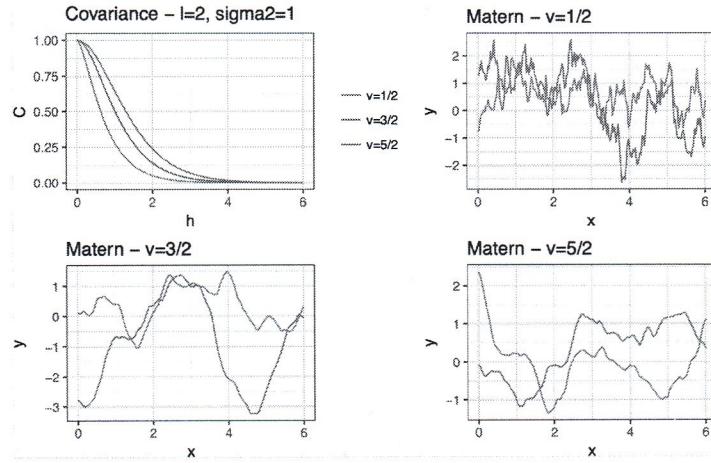


Figure 3.3: Matérn kernel functions, with realizations of resulting Gaussian fields ($d = 1$)