Kernel Design

GP Summer School

Sheffield, June 2013

Nicolas Durrande

We have seen during the introduction lectures that the distribution of a GP Z depends on two functions:

- the mean $m(.)$
- the covariance $K(.,.)$

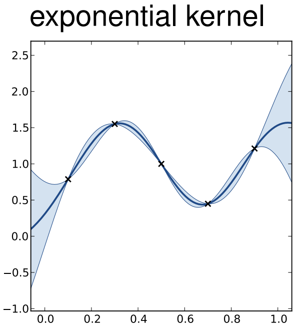
In this talk, we will focus on **the covariance function**:

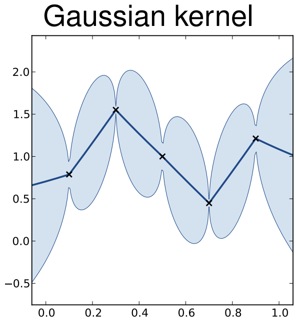
$$K(x, y) = \operatorname{cov}(Z(x), Z(y))$$

The choice of the kernel is of great importance in GP regression

$$m(x) = \mathbf{k}(x)^t \mathrm{K}^{-1} \mathbf{Y}$$
$$c(x,y) = K(x,y) - \mathbf{k}(x)^t \mathrm{K}^{-1} \mathbf{k}(y)$$

Example





K has to reflect the prior belief on the function to approximate

1. **Introduction**
2. **What is a kernel?**
    - Kernels and positive definite functions
    - Stationary kernels
3. **Kernels and positive measures**
    - Bochner's theorem
    - Examples on usual kernels
    - Spectral approximation
4. **Making new from old**
    - Multiplication by a scalar
    - Sum of kernels
    - Product of kernels
    - Multiplication by a function
    - Composition with a function
    - Effect of a linear operator
5. **Conclusion**

We will first recall some definitions

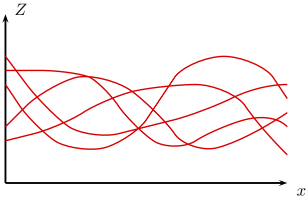
Gaussian vector

A d-dimensional random vector Y is said to be Gaussian iif $a^t Y$ is Gaussian $\forall a \in \mathbb{R}^d$

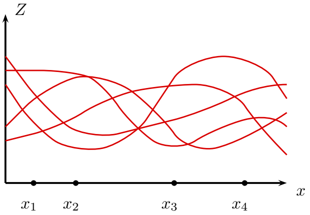
Gaussian process

A random process Z indexed by D is said to be Gaussian iif $(Z(x_1), \dots, Z(x_n))$ is a Gaussian vector $\forall x_i \in D$, $\forall n \in \mathbb{N}$

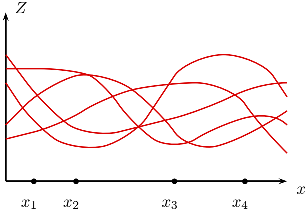
Same definitions with images:

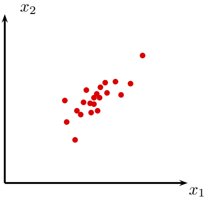


Same definitions with images:

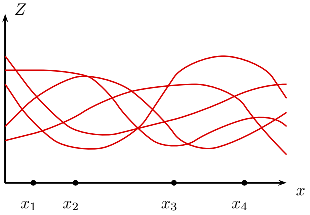


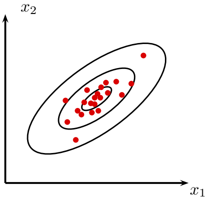
Same definitions with images:



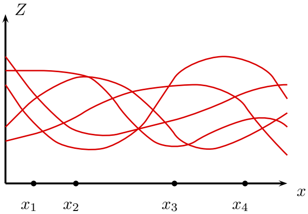


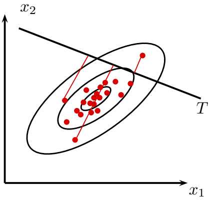
Same definitions with images:



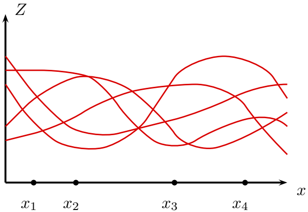


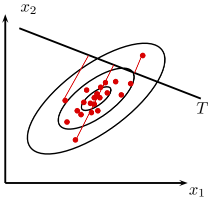
Same definitions with images:



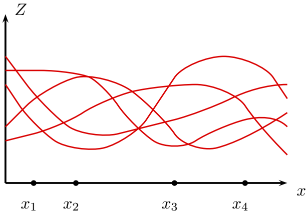


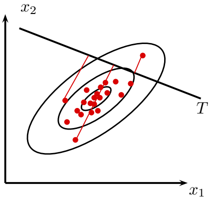
Same definitions with images:

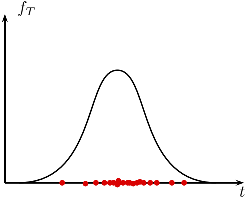




Same definitions with images:







Let Z be a random process. Some properties of kernels can be obtained directly from their definition.

Example

$$K(x,x) = \mathrm{cov}\left(Z(x), Z(x)\right) = \mathrm{var}\left(Z(x)\right) \geq 0$$

$\Rightarrow K(x,x)$ is **positive.**

$$K(x,y) = \mathrm{cov}\left(Z(x), Z(y)\right) = \mathrm{cov}\left(Z(y), Z(x)\right) = K(y,x)$$

$\Rightarrow K(x,y)$ is **symmetric.**

We can obtain a thinner result...

We introduce the random variable $T = \sum_{i=1}^{n} a_i Z(x_i)$ where n, a_i and x_i are arbitrary.

Computing the variance of T gives:

$$\mathrm{var}\left(T\right) = \sum\sum a_i a_j \mathrm{cov}\left(Z(x_i), Z(x_j)\right) = \sum\sum a_i a_j K(x_i, x_j)$$

We thus have:

$$\sum\sum a_i a_j K(x_i, x_j) \geq 0$$

Definition

The functions satisfying the above inequality **for all** $n \in \mathbb{N}$, **for all** $x_i \in D$, **for all** $a_i \in \mathbb{R}$ are called positive semi-definite functions.

We have not assumed here that Z is Gaussian!

We have seen:

$$K \text{ is a covariance} \Rightarrow K \text{ is a positive semi-definite function}$$

The reverse is also true:

Theorem (Loeve)

$$\textit{K corresponds to the covariance of a GP}$$
$$\Updownarrow$$
$$\textit{K is a (symmetric) positive definite function}$$

Major issue

It is often intractable to show that function is positive definite directly from the definition...

A common approach is to use well known kernels such as:

white noise: $K(x, y) = \delta_{x,y}$

bias: $K(x, y) = 1$

exponential: $K(x, y) = \exp\left(-|x - y|\right)$

Brownian: $K(x, y) = \min(x, y)$

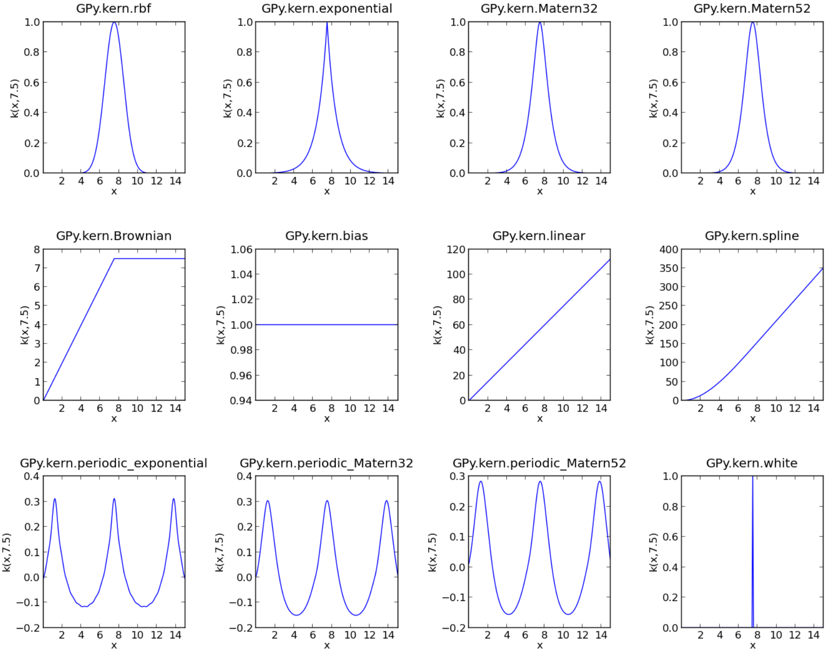
Gaussian: $K(x, y) = \exp\left(-(x - y)^2\right)$

Matérn 3/2: $K(x, y) = (1 + |x - y|) \times \exp\left(-|x - y|\right)$

sinc: $K(x, y) = \dfrac{\sin(|x - y|)}{|x - y|}$

$\vdots$

Kernels that are a function of $|x - y|$ are called **stationary** kernels.


GPy.kern.rbf
GPy.kern.exponential
GPy.kern.Matern32
GPy.kern.Matern52
GPy.kern.Brownian
GPy.kern.bias
GPy.kern.linear
GPy.kern.spline
GPy.kern.periodic_exponential
GPy.kern.periodic_Matern32
GPy.kern.periodic_Matern52
GPy.kern.white
k(x,7.5)
x

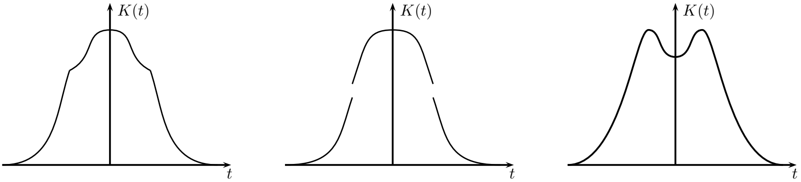
Let $K(x, y) = \tilde{K}(|x - y|)$ be a stationary kernel.

Properties

- If $\tilde{K}$ is n times differentiable in 0, then it is n times differentiable everywhere.
- The maximum value of $\tilde{K}(t)$ is reached in $t = 0$.

Example

The following functions are not valid covariance structures



Theorem (Bochner)

A stationary function $k(x, y) = \tilde{k}(|x - y|)$ is positive definite if and only if $\tilde{k}$ can be represented as

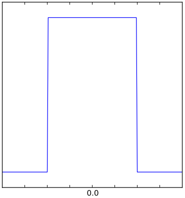
$$\tilde{k}(t) = \int_{\mathbb{R}} e^{i\omega t} \mathrm{d}\mu(\omega)$$

where μ is a finite positive measure.

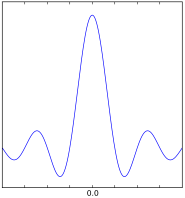
This result is very useful to prove the positive definiteness of stationary functions.

Example

We consider the following measure:



Its Fourier transform gives $\tilde{k}(t) = \dfrac{\sin(t)}{t}$:



As a consequence, $K(x, y) = \dfrac{\sin(x - y)}{x - y}$ is a valid covariance function.

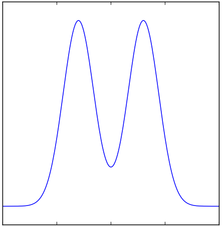
Bochner theorem can be used to prove the positive definiteness of many usual stationary kernels

- The Gaussian is the Fourier transform of itself
  $\Rightarrow$ it is psd.
- $\delta_{x,y}$ is the inverse Fourier transform of the constant function
  $\Rightarrow$ it is psd.
- the constant function is the inverse Fourier transform of $\delta_{x,y}$
  $\Rightarrow$ it is psd.

Spectral approximation with a mixture of Gaussian (A. Wilson, ICML 2013)

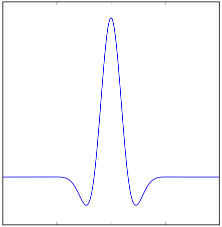
The inverse Fourier transform of a (symmetrised) non centred Gaussian is:

$\mu(\omega)$



$$\xrightarrow[\mathcal{F}^{-1}]{}$$

$\tilde{k}(t)$

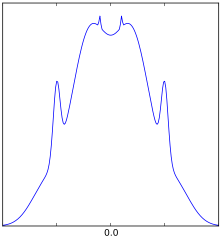


This can be generalised to a measure based on the sum of Gaussians.

Spectral approximation with a mixture of Gaussian (A. Wilson, ICML 2013)

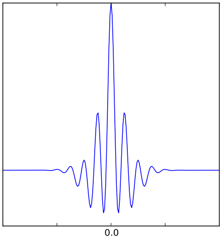
We obtain a kernel that is parametrised by the means and the bandwidths of Gaussians bells in the measure space:

$\mu(\omega)$



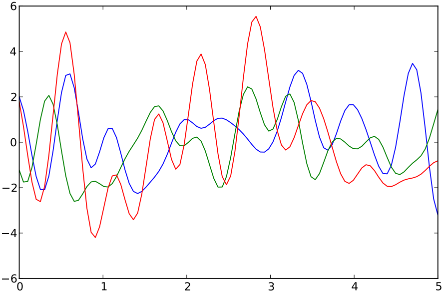
$$\xrightarrow[\mathcal{F}^{-1}]{}$$

$\tilde{k}(t)$



Spectral approximation with a mixture of Gaussian (A. Wilson, ICML 2013)

The sample paths have the following aspect:



We have seen that it is difficult to prove directly the positive semi-definiteness of a function.

For all $n \in \mathbb{N}$, **for all** $x_i \in D$, **for all** $a_i \in \mathbb{R}$

$$\sum \sum a_i a_j K(x_i, x_j) \geq 0$$

However, many operations can be applied to a psd function while retaining this property. This is often called **making new from old**.

We will discuss the following operations:

- Multiplication by a scalar
- Sum of kernels
- Product of kernels
- Multiplication by a function
- Composition with a function
- Effect of a linear operator

Multiplication by a scalar

Hereafter, we assume that K_i is a kernel and that $Z_i \sim \mathcal{N}(0, K_i)$.

Property

Let α be a positive real, then

$$K(x,y) = \alpha K_1(x,y)$$

is a valid kernel.

proof

$\forall n \in \mathbb{N}$, $\forall x_i \in D$, $\forall a_i \in \mathbb{R}$

$$\sum \sum a_i a_j \alpha K(x_i, x_j) = \alpha \sum \sum a_i a_j K(x_i, x_j) \geq 0$$

From a GP point of view, K is the covariance of $\sqrt{\alpha} Z_1$:

$$\mathrm{cov}\left(\sqrt{\alpha} Z_1(x), \sqrt{\alpha} Z_1(y)\right) = \sqrt{\alpha}\sqrt{\alpha}\mathrm{cov}\left(Z_1(x), Z_1(y)\right) = \alpha K(x,y)$$

Sum of kernels

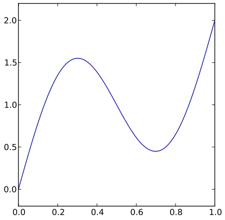
Let $f_1, \ f_2$ be two functions $\mathbb{R} \rightarrow \mathbb{R}$:

$$f_1(x) = \sin(2\pi x)$$
$$f_2(x) = 2x$$

The sum $f = f_1 + f_2$ can be understood in two different ways:

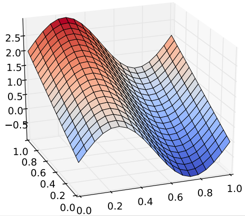
As a function defined over $\mathbb{R}$

$$f(x) = f_1(x) + f_2(x)$$



As a function over $\mathbb{R} \times \mathbb{R}$

$$f(x_1, x_2) = f_1(x_1) + f_2(x_2)$$



Sum of kernels defined over the same space.

Property

$$K(x, y) = K_1(x, y) + K_2(x, y)$$

is a valid covariance structure.

proof

$\forall n \in \mathbb{N}, \forall x_i \in D, \forall a_i \in \mathbb{R}$

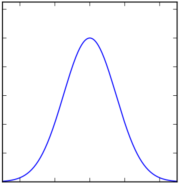
$$\begin{aligned} \sum \sum a_i a_j K(x_i, x_j) &= \sum \sum a_i a_j (K_1(x_i, x_j) + K_2(x_i, x_j)) \\ &= \sum \sum a_i a_j K_1(x_i, x_j) + \sum \sum a_i a_j K_2(x_i, x_j) \geq 0 \end{aligned}$$

Remark:

- From a GP point of view, K is the kernel of $Z(x) = Z_1(x) + Z_2(x)$

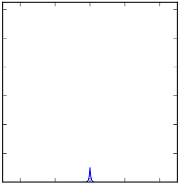
Example

We can sum a Gaussian and an exponential kernel:



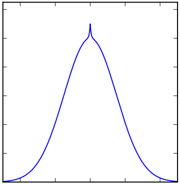
Gaussian

$+$



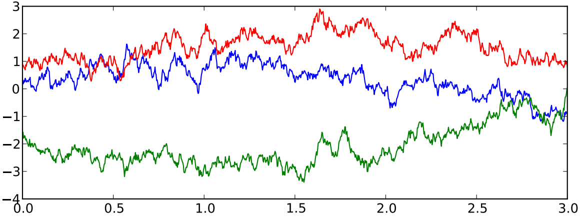
Exponential

$=$



psd kernel

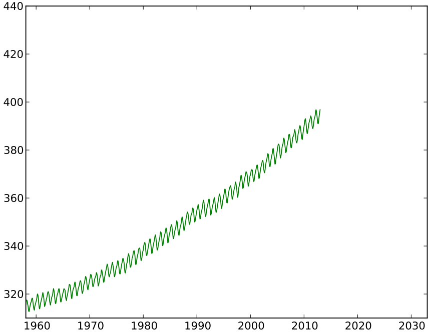
We obtain the following GP sample paths:



In practice, summing kernels is very useful.

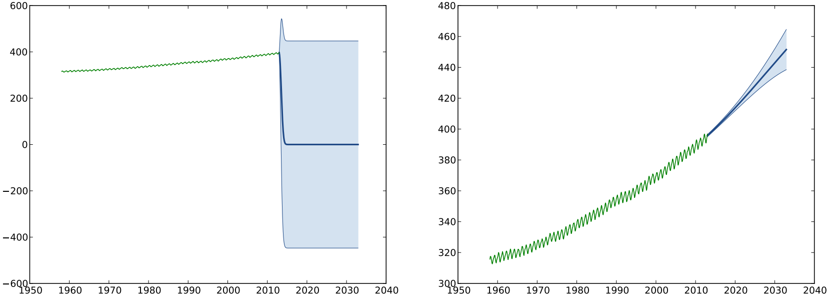
Example (The Mauna Loa observatory dataset)

This famous dataset compiles the monthly CO_2 concentration in Hawaii since 1958.



Let's try to predict the concentration for the next 20 years.

We first consider a squared-exponential kernel:



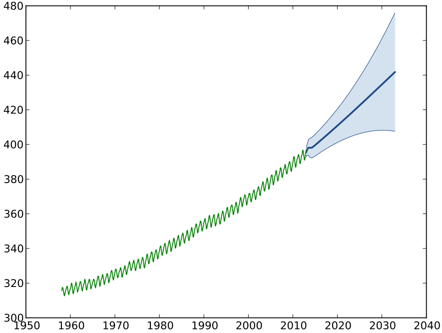
The results are terrible!

What happen if we sum both kernels?

$$k(x,y) = \sigma_1^2 k_{rbf1}(x,y) + \sigma_2^2 k_{rbf2}(x,y)$$

What happen if we sum both kernels?

$$k(x,y) = \sigma_1^2 k_{rbf1}(x,y) + \sigma_2^2 k_{rbf2}(x,y)$$



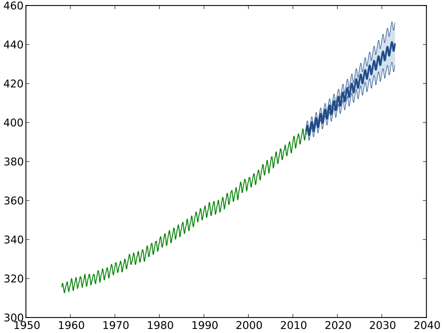
The model is drastically improved!

We can try the following kernel:

$$k(x, y) = \sigma_0^2 x^2 y^2 + \sigma_1^2 k_{rbf1}(x, y) + \sigma_2^2 k_{rbf2}(x, y) + \sigma_3^2 k_{per}(x, y)$$

We can try the following kernel:

$$k(x, y) = \sigma_0^2 x^2 y^2 + \sigma_1^2 k_{rbf1}(x, y) + \sigma_2^2 k_{rbf2}(x, y) + \sigma_3^2 k_{per}(x, y)$$



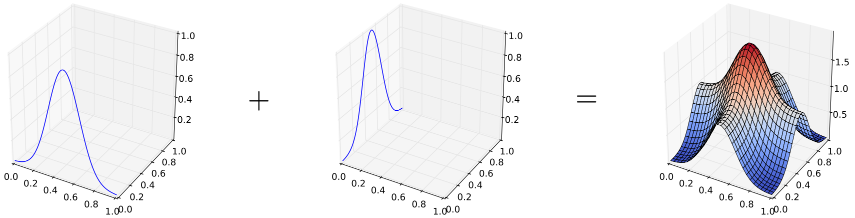
Once again, the model is significantly improved.

Sum of kernels defined over different spaces

Property

$$K(\mathbf{x},\mathbf{y}) = K_1(x_1,y_1) + K_2(x_2,y_2) \tag{1}$$

is valid covariance structure.

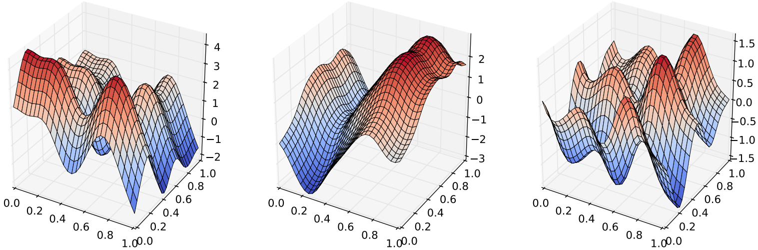


Remark:

- From a GP point of view, K is the kernel of $Z(\mathbf{x}) = Z_1(x_1) + Z_2(x_2)$

Sum of kernels defined over different spaces

We can have a look at a few sample paths from Z:



$\Rightarrow$ They are additive (up to a modification)

Tensor Additive kernels are very useful for

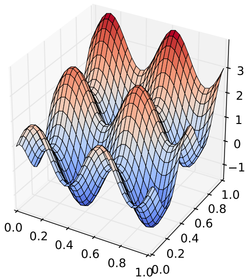
- Approximating additive functions
- Building models over high dimensional inputs spaces

Sum of kernels defined over different spaces

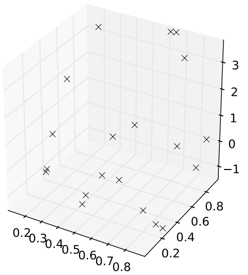
Approximating an additive function

We consider the test function $f(x) = \sin(4\pi x_1) + \cos(4\pi x_2) + 2x_2$ and a set of 20 observation in $[0, 1]^2$

Test function



Observations



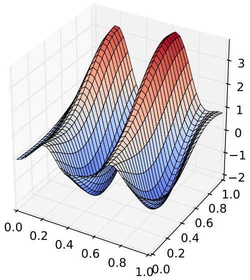
Sum of kernels defined over different spaces

Approximating an additive function

We obtain the following models:

Gaussian kernel

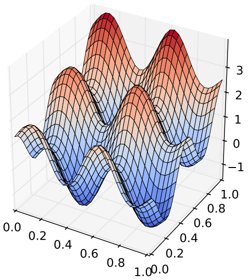
Mean predictor



RMSE is 1.06

Additive Gaussian kernel

Mean predictor



RMSE is 0.12

Sum of kernels defined over different spaces

Approximating an additive function

Remark

- It is straightforward to show that the mean predictor is additive

$$\begin{aligned} m(\mathbf{x}) &= (\mathbf{k_1}(x_1) + \mathbf{k_2}(x_2))^t (\mathrm{K}_1 + \mathrm{K}_2)^{-1} \mathbf{Y} \\ &= \underbrace{\mathbf{k_1}(x_1)^t (\mathrm{K}_1 + \mathrm{K}_2)^{-1} \mathbf{Y}}_{m_1(x_1)} + \underbrace{\mathbf{k_2}(x_2)^t (\mathrm{K}_1 + \mathrm{K}_2)^{-1} \mathbf{Y}}_{m_2(x_2)} \end{aligned}$$

$\Rightarrow$ The mean predictor shares the prior behaviour.

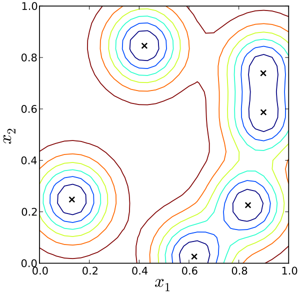
Sum of kernels defined over different spaces

Approximating an additive function

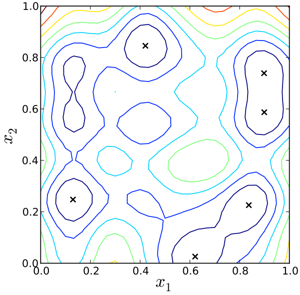
Remark

- The prediction variance has interesting features

pred. var. with kernel product



pred. var. with kernel sum

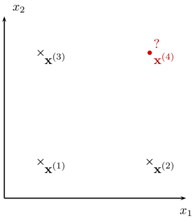


Let's consider a toy example to illustrate this.

Sum of kernels defined over different spaces

Approximating an additive function

Provided that f is additive, we want to predict $f(\mathbf{x}^{(4)})$ knowing $f(\mathbf{x}^{(1)})$, $f(\mathbf{x}^{(2)})$ and $f(\mathbf{x}^{(3)})$ of an additive function,



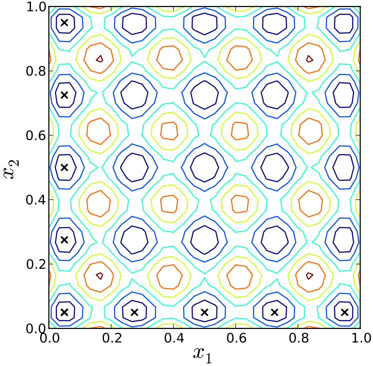
Given an additive GP, we compute the prediction variance in $\mathbf{x}^{(4)}$

$$
\begin{aligned}
c(\mathbf{x}^{(4)}, \mathbf{x}^{(4)}) &= \mathrm{var}\left(Z(\mathbf{x}^{(4)}) | Z(\mathbf{x}^{(1)}), Z(\mathbf{x}^{(2)}), Z(\mathbf{x}^{(3)})\right) \\
&= \mathrm{var}\left(Z(\mathbf{x}^{(2)}) + Z(\mathbf{x}^{(3)}) - Z(\mathbf{x}^{(1)}) | Z(\mathbf{x}^{(1)}), Z(\mathbf{x}^{(2)}), Z(\mathbf{x}^{(3)})\right) \\
&= 0
\end{aligned}
$$

Sum of kernels defined over different spaces

Approximating an additive function

Using this property we can construct a design of experiment that covers the space with only $cst \times d$ points!

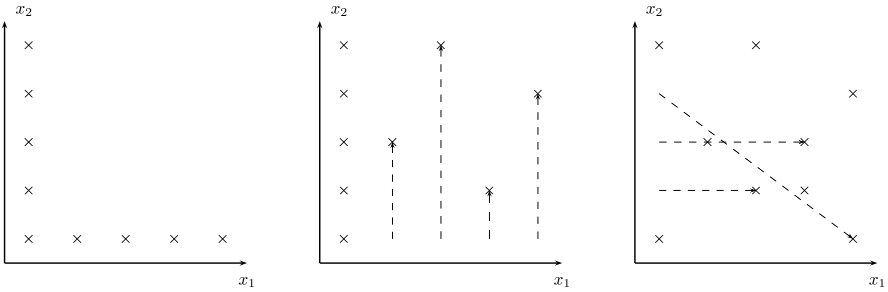


Prediction variance

Sum of kernels defined over different spaces

Approximating an additive function

Note that the distribution of the points can be modified for a better coverage of the space:

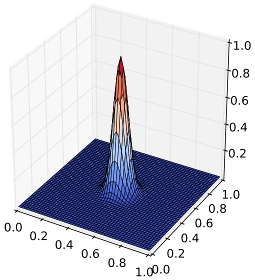


Sum of kernels defined over different spaces

High-dimensional modelling We assume here $f : \mathbb{R}^d \rightarrow \mathbb{R}$

Stationary kernels

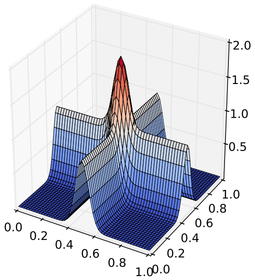
$$K(\mathbf{x}, \mathbf{y}) = f(|\mathbf{x} - \mathbf{y}|)$$



$\Rightarrow$ *cst^d* points are required to cover the space!

Additive kernels

$$K(\mathbf{x}, \mathbf{y}) = K_1(x_1, y_1) + K_2(x_2, y_2)$$



$\Rightarrow$ *$cst \times d$* points are required to cover the space!

Product of kernels

What about the product of kernels?

As previously, two products can be defined

- Over the same space
- Over the tensor product space

Product over the same space

Property

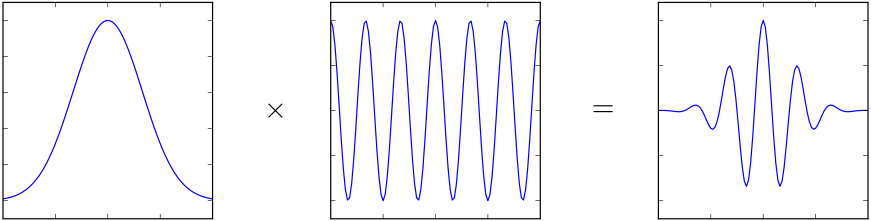
Let $K_1, \ K_2$ be two kernels over $D \times D$, then

$$K(x, y) = K_1(x, y) \times K_2(x, y)$$

is a kernel.

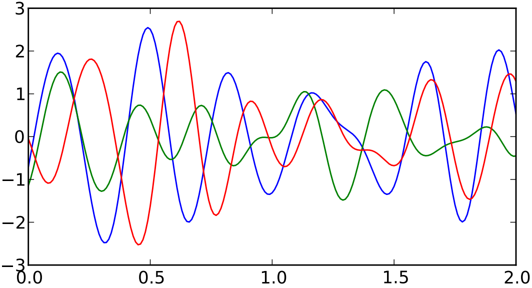
Example

We consider the product of a squared exponential with a cosine:



Product over the same space

Examples of sample paths from the previous kernel:



Product over the tensor space

Property

Let $K_1, \ K_2$ be two kernels resp over $D_1 \times D_1$ and $D_2 \times D_2$, then

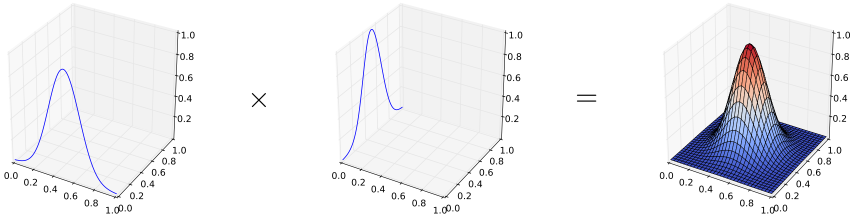
$$K(\mathbf{x}, \mathbf{y}) = K_1(x_1, y_1) \times K_2(x_1, y_1)$$

is a kernel over $(D_1 \times D_2) \times (D_1 \times D_2)$.

Tensor product can be used to obtain covariance structures in higher dimension.

Example

We compute the product of two squared exponential kernels

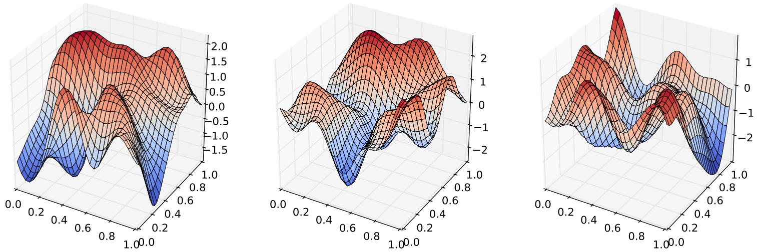


We have:

$$K(\mathbf{x},\mathbf{y}) = e^{-(x_1-y_1)^2} \times e^{-(x_2-y_2)^2} = e^{-\sum(x_i-y_i)^2} = e^{-||\mathbf{x}-\mathbf{y}||^2}$$

$\Rightarrow$ We can recognise here a 2D squared exponential kernel.

Here is a few sample paths from Z:



This GP **cannot be seen** as the product of two independent GPs with kernels K_1 and K_2

$$Z(\mathbf{x}) \neq Z_1(x_1) \times Z_2(x_2)$$

Multiplication by a function

Property

Let f be an arbitrary function over D_1, then

$$K(x,y) = f(x)f(y)K_1(x,y)$$

is a kernel over $D_1 \times D_1$.

proof

$$\sum\sum a_i a_j K(x_i,x_j) = \sum\sum \underbrace{a_i f(x_i)}_{b_i} \underbrace{a_j f(x_j)}_{b_j} K_1(x_i,x_j) \geq 0$$

Remarks:

- This property is a generalization of the multiplication by a scalar
- $f(x)f(y)$ corresponds to the covariance of $Z_2(x) = \alpha f(x)$ with $\alpha \sim \mathcal{N}(0,1)$. The property can thus be seen as the product of two kernels defined over the same space.

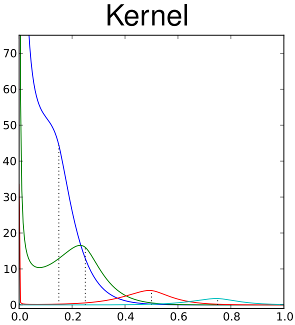
Example

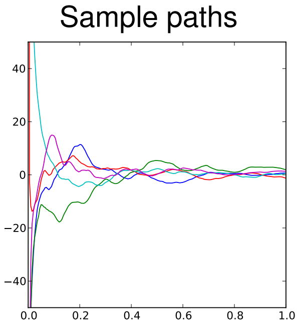
We illustrate the previous property with $f(x) = \frac{1}{x}$ and a Matérn 3/2 kernel $K_1(x, y) = (1 + |x - y|)e^{-|x-y|}$.

Example

We illustrate the previous property with $f(x) = \frac{1}{x}$ and a Matérn 3/2 kernel $K_1(x, y) = (1 + |x - y|)e^{-|x-y|}$.

We obtain:





- This property can be seen as a (nonlinear) rescaling of the output space

Composition with a function

Property

Let K_1 be a kernel over $D_1 \times D_1$ and f be an arbitrary function $D \to D_1$, then

$$K(x, y) = K_1(f(x), f(y))$$

is a kernel over $D \times D$.

proof

$$\sum \sum a_i a_j K(x_i, x_j) = \sum \sum a_i a_j K_1(\underbrace{f(x_i)}_{y_i}, \underbrace{f(x_j)}_{y_j}) \geq 0$$

Remarks:

- K corresponds to the covariance of $Z(x) = Z_1(f(x))$
- This can be seen as a (nonlinear) rescaling of the input space

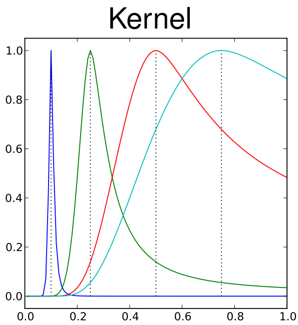
Example

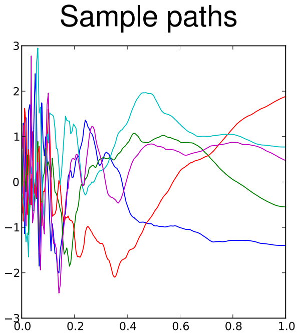
We consider $f(x) = \frac{1}{x}$ and a Matérn 3/2 kernel $K_1(x, y) = (1 + |x - y|)e^{-|x-y|}$.

Example

We consider $f(x) = \frac{1}{x}$ and a Matérn 3/2 kernel $K_1(x, y) = (1 + |x - y|)e^{-|x-y|}$.

We obtain:





Effect of a linear operator

Property

Let L be a linear operator that commutes with the covariance, then

$$K(x, y) = L_x(L_y(K_1(x, y)))$$

is a kernel.
proof
K is the kernel of $L_x(Z_x)$

Remarks:

- The RKHS framework allows to give proper conditions for the above property, but it is out of the scope of this talk.

We want to approximate a function $[0,1] \to \mathbb{R}$ that is symmetric with respect to 0.5. We will consider 2 linear operators:

$$L_1 : f(x) \to \begin{cases} f(x) & x < 0.5 \\ f(1-x) & x \geq 0.5 \end{cases}$$

$$L_2 : f(x) \to \frac{f(x) + f(1-x)}{2}.$$

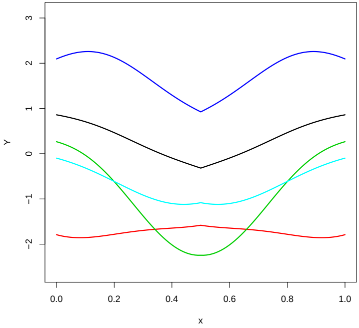
Those operators transform any function into a symmetric function.

Let $K_1 = L_1(L_1(K))$ and $K_2 = L_2(L_2(K))$ be their associated kernels.

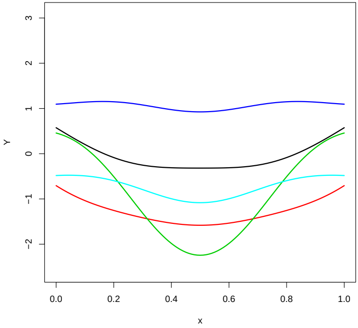
Effect of a linear operator: example (Ginsbourger, AFST 2013)

Examples of associated sample paths are

K_1



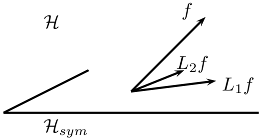
K_2



The differentiability is not always respected!

Effect of a linear operator

Ideally, we want to extract the subspace of symmetric functions in $\mathcal{H}$



and to define L as the orthogonal projection onto $\mathcal{H}_{sym}$

$\Rightarrow$ This can be difficult... but it raises interesting questions!

Effect of a linear operator

We now consider another example:

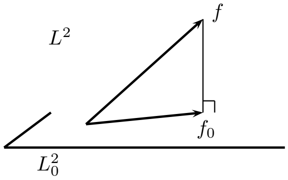
We want to approximate a function f that is exactly zero mean:

$$\int_D f(x)\mathrm{d}x = 0$$

Can we build a kernel that takes into account this property?

It is straightforward to build a linear operator that centres functions:

$$L : f(x) \to f_0(x) = f(x) - \int_D f(s)\mathrm{d}s$$



Effect of a linear operator

Let's apply L to a GP Z with kernel K:

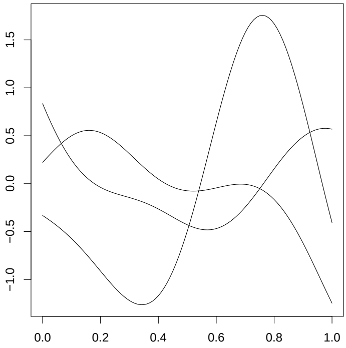
$$Z_0(x) = L(Z)(x) = Z(x) - \int Z(s)\mathrm{d}s$$

and compute the covariance of Z_0:

$$\begin{aligned}
K_0(x,y) &= \mathrm{cov}\left(Z_0(x), Z_0(y)\right) \\
&= \mathrm{cov}\left(Z(x) - \int Z(s)\mathrm{d}s, Z(y) - \int Z(s)\mathrm{d}s\right) \\
&= \mathrm{cov}\left(Z(x), Z(y)\right) - \mathrm{cov}\left(Z(x), \int Z(s)\mathrm{d}s\right) \\
&\qquad - \mathrm{cov}\left(Z(y), \int Z(s)\mathrm{d}s\right) + \mathrm{var}\left(\int Z(s)\mathrm{d}s\right) \\
&= K(x,y) - \int K(x,s)\mathrm{d}s - \int K(y,s)\mathrm{d}s - \iint K(s,t)\mathrm{d}s\mathrm{d}t
\end{aligned}$$

Effect of a linear operator

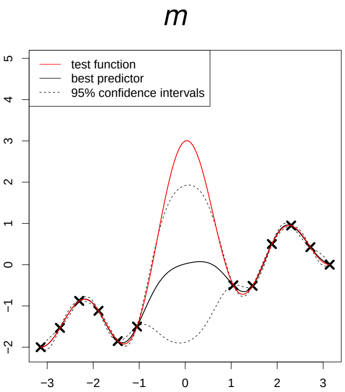
We can use K_0 to simulate sample paths from Z_0:

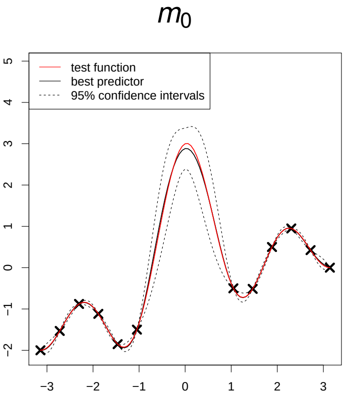


These sample paths are exacty zero-mean!

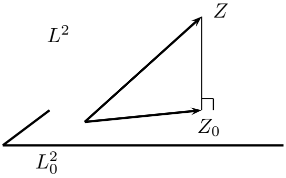
Effect of a linear operator

We can compare the predictions of two models m and m_0 respectively based on K and K_0.





Effect of a linear operator

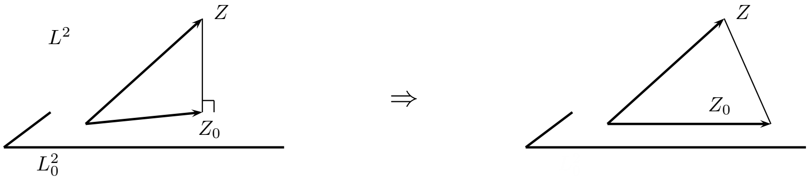


Are Z_0 and $Z - Z_0$ independent?

$$\begin{aligned}
&\operatorname{cov}\left(Z_0(x), Z(y) - Z_0(y)\right) \\
&\quad = \operatorname{cov}\left(Z(x) - \int Z(s)\mathrm{d}s, Z(y) - Z(y) + \int Z(s)\mathrm{d}s\right) \\
&\quad = \int K(x,s)\mathrm{d}s - \int K(s,t)\mathrm{d}s\mathrm{d}t \neq 0
\end{aligned}$$

$\Rightarrow$ They are not!

Effect of a linear operator



Are Z_0 and $Z - Z_0$ independent?

$$
\begin{aligned}
&\operatorname{cov}\left(Z_0(x), Z(y) - Z_0(y)\right)\\
&\quad = \operatorname{cov}\left(Z(x) - \int Z(s)\mathrm{d}s, Z(y) - Z(y) + \int Z(s)\mathrm{d}s\right)\\
&\quad = \int K(x,s)\mathrm{d}s - \int K(s,t)\mathrm{d}s\mathrm{d}t \neq 0
\end{aligned}
$$

$\Rightarrow$ They are not!

Effect of a linear operator

The alternative here is to change the way to center the functions:

$$L_\perp : f(x) \to f(x) - g(x) \int_D f(s) \mathrm{d}s$$

where $\int g(s)\mathrm{d}s = 1$. It can be shown that:

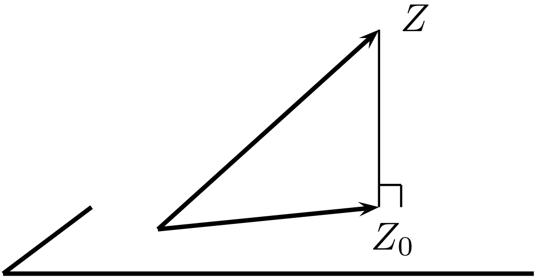
$$g(x) = \frac{\displaystyle\int_D k(x,s)\mathrm{d}s}{\displaystyle\int_D k(s,t)\mathrm{d}s\mathrm{d}t}$$

gives:

$$\mathrm{cov}\left(Z_0(x), Z(y) - Z_0(y)\right) = 0$$

Effect of a linear operator

We finally obtain:



In a space where the orthogonality is meaningful for the GP Z!

Small recap

We have seen that

- The choice of the kernel has to reflect the prior belief about the function to approximate.
- Kernels can (and should) be tailored to the problem at hand.

Making new from old

Although a direct proof of the positive definiteness of a function is often intractable, it is possible to

- multiply kernels
- sum kernels
- multiply a kernel by a function
- compose a kernel with a function

Linear application

If we have a linear application that transforms any function into a function satisfying the desired property, it is possible to build a GP fulfilling the requirements.

Any questions ?