MATH 590: Meshfree Methods

Chapter 3: Examples of Kernels

Greg Fasshauer

Department of Applied Mathematics
Illinois Institute of Technology

Fall 2014

Outline

Isotropic Radial Kernels

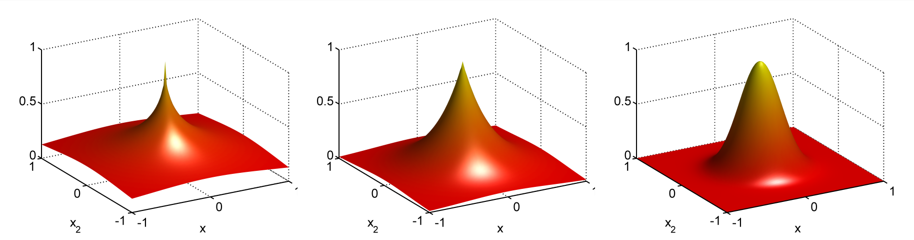
- Of the form

$$K(\boldsymbol{x}, \boldsymbol{z}) = \kappa(\|\boldsymbol{x} - \boldsymbol{z}\|), \quad \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d, \quad \kappa : \mathbb{R}_0^+ \to \mathbb{R},$$

Example

Powered exponential kernel (plotted with $\beta = 0.5, 1, 2$, $\varepsilon = 3$)

$$\kappa(r) = \mathrm{e}^{-(\varepsilon r)^{\beta}}, \quad \beta \in (0, 2]$$



- The family of powered exponential kernels is common in the statistics and machine learning literature since the two parameters ε and β provide flexibility with respect to scale and smoothness.
- However, the powered exponential kernel is smooth only for $\beta = 2$, i.e., the Gaussian.
- They are positive definite on $\mathbb{R}^d$ for all d.
- The case $\beta = 1$ is known as the Ornstein–Uhlenbeck kernel, and also corresponds to the Matérn kernel with $\beta = \frac{d+1}{2}$ (see next).
- The Gaussian is sometimes referred to as squared exponential in the machine learning or statistics literature.

Example

Matérn (or Sobolev) kernel (plotted with $d = 2$, $\varepsilon = 3$)

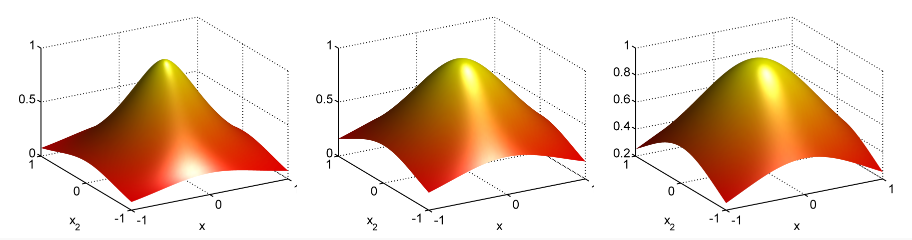
$$\kappa(\varepsilon r) = \frac{K_{d/2-\beta}(\varepsilon r)}{(\varepsilon r)^{d/2-\beta}}, \quad \beta > \frac{d}{2}$$

K_ν: modified Bessel functions of the second kind of order ν

$$\kappa(\varepsilon r) = (1+\varepsilon r)e^{-\varepsilon r}, \qquad \text{(when } \beta = (d+3)/2)$$

$$\kappa(\varepsilon r) = (1+\varepsilon r + \frac{1}{3}(\varepsilon r)^2)e^{-\varepsilon r}, \qquad \text{(when } \beta = (d+5)/2)$$

$$\kappa(\varepsilon r) = (1+\varepsilon r + \frac{2}{5}(\varepsilon r)^2 + \frac{1}{15}(\varepsilon r)^3)e^{-\varepsilon r}, \qquad \text{(when } \beta = (d+7)/2)$$



- Matérn kernels are popular in the statistics and approximation theory communities.

- They are fundamental solutions of the d-dimensional iterated modified Helmholtz operator in Euclidean coordinates, i.e.,

$$\mathcal{D} = \left(-\nabla^2 + \varepsilon^2 \mathcal{I}\right)^{\beta},$$

  with $\mathcal{I}$ the identity operator.

- The parameters ε and β specify scale and smoothness of the kernel, respectively.

- Matérn kernels generate classical Sobolev spaces $H^{\beta}(\mathbb{R}^d)$ as their RKHSs.

- They are positive definite on $\mathbb{R}^d$, but only when $\beta > \frac{d}{2}$.

Example

(Inverse) Multiquadric kernels (plotted with $\varepsilon = 3$)

$$\kappa(\varepsilon r) = (1 + \varepsilon^2 r^2)^\beta, \quad \beta \in \mathbb{R} \setminus \mathbb{N}_0$$

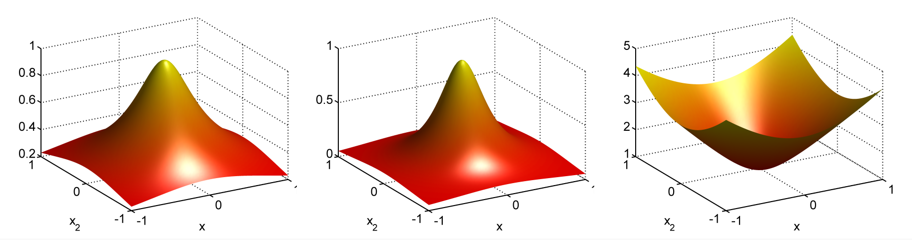
$\beta < 0$: inverse MQs (positive definite)

$\beta > 0$: MQs (conditionally positive definite of different orders)

$$\kappa(\varepsilon r) = \frac{1}{\sqrt{1 + \varepsilon^2 r^2}}, \qquad \text{(IMQ)}$$

$$\kappa(\varepsilon r) = \frac{1}{1 + \varepsilon^2 r^2}, \qquad \text{(IQ or Cauchy)}$$

$$\kappa(\varepsilon r) = \sqrt{1 + \varepsilon^2 r^2}, \qquad \text{(MQ)}$$



- Popular mostly in approximation theory and engineering applications.
- The IQ kernel is equivalent to the rational quadratic kernel (see, e.g., [Gen02]) since

$$\frac{1}{1+\varepsilon^2 r^2} = 1 - \frac{r^2}{\theta + r^2}$$

with $\theta = 1/\varepsilon^2$. This kernel is sometimes recommended as a computationally cheaper alternative to the Gaussian kernel in the machine learning literature.

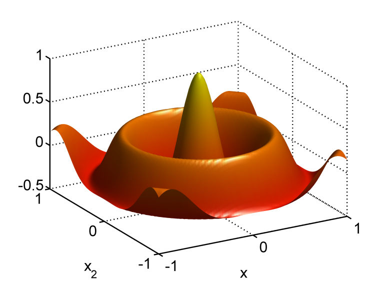
- (Inverse) MQ kernels are (conditionally) positive definite on $\mathbb{R}^d$ for all d.

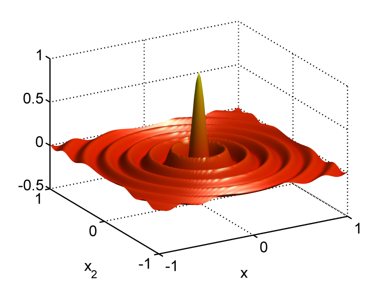
Oscillatory kernels (plotted with $\varepsilon = 10$, $d = 2$)

$$\kappa(\varepsilon r) = \frac{J_{d/2-1}(\varepsilon r)}{(\varepsilon r)^{d/2-1}}, \qquad \text{(Poisson or Bessel)}$$

$$\kappa(\varepsilon r) = \frac{\sin(\varepsilon r)}{\varepsilon r}, \qquad \text{(wave, Poisson with } d = 3\text{)}$$

J_ν: Bessel functions of the first kind of order ν





Bessel kernels were introduced in [FLW06]. The wave kernel sometimes appears in machine learning (see, e.g., [Gen02]).

They are positive definite only in dimension $\leq d$.

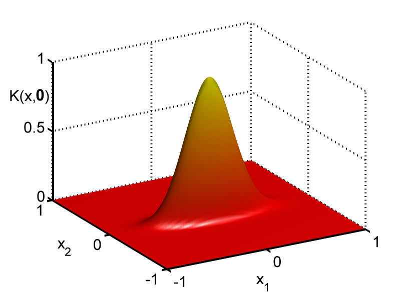
Anisotropic Radial Kernels

Any isotropic radial kernel can be turned into an anisotropic radial kernel by using a weighted 2-norm instead of an unweighted one.

Example (Anisotropic Gaussian)

$$K(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{z})^T \mathsf{E}(\boldsymbol{x}-\boldsymbol{z})},$$

with E a symmetric positive definite matrix.
If $\mathsf{E} = \varepsilon^2 \mathsf{I}_d$, with I_d a $d \times d$ identity matrix, then the kernel is isotropic.



- Anisotropic kernels are not common in the approximation theory literature. They have been
  - analyzed, e.g., in [Bax06, BDL10] and
  - applied, e.g., in [CBM+03, CLMM06].
- But they are very popular in the literature on information-based complexity, e.g., [NW08].
- [FHW12a, FHW12b] used $\mathsf{E} = \text{diag}(\varepsilon_1^2, \ldots, \varepsilon_d^2)$, a diagonal matrix with dimension-dependent shape parameters, to avoid the curse of dimensionality and obtain dimension-independent error bounds.

Remark

Some authors have applied a different scale to each basis function in the RBF interpolation expansion resulting in, e.g.,

$$s(\boldsymbol{x}) = \sum_{j=1}^{N} c_j e^{-\varepsilon_j^2 \|\boldsymbol{x}-\boldsymbol{x}_j\|^2}, \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

Now the interpolant is no longer generated by a single kernel and the theoretical foundation must be reconsidered.

The most promising paper to address this approach — especially on a theoretical level — is [BLRS14], where the problem is tackled by embedding a d-dimensional interpolation problem into $\mathbb{R}^{d+1}$ so that the additional dimension houses the locally varying shape parameter. In $\mathbb{R}^{d+1}$ one then works with a "standard" kernel with fixed global shape.

Translation Invariant Kernels

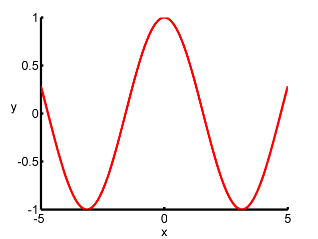
- A kernel is called translation invariant (or stationary in the statistics literature) if $K(\boldsymbol{x}+\boldsymbol{h}, \boldsymbol{z}+\boldsymbol{h}) = K(\boldsymbol{x}, \boldsymbol{z})$ for any $\boldsymbol{h} \in \mathbb{R}^d$. This means that K is a function of the difference of $\boldsymbol{x}$ and $\boldsymbol{z}$, i.e., it's of the form

$$K(\boldsymbol{x}, \boldsymbol{z}) = \widetilde{K}(\boldsymbol{x} - \boldsymbol{z}).$$

Example

Cosine kernel

$$K(x, z) = \cos(x - z)$$



- In the literature on integral equations, translation invariant kernels are often called convolution kernels.
- Discretized convolution kernels (i.e., matrices) find applications in image processing.
- The Fourier transform is the ideal tool to analyze translation invariant kernels. This results in, e.g., Bochner's theorem.
- Any nonnegative (infinite) linear combination of kernels of the form $K_n(x,z) = \cos(n(x-z))$ is positive definite and translation invariant on $\mathbb{R}$.
  - E.g., periodic univariate splines can be represented with the kernel

$$K(x,z) = \sum_{n=1}^{\infty} \frac{2}{(2n\pi)^{2\beta}} \cos\left(2n\pi(x-z)\right)$$

    whose RKHS is $H^{\beta}_{\text{per}}(0,1)$ (see [Wah90, Chapter 2]).
- To get a kernel in higher dimensions we can take a tensor product of one-dimensional translation invariant kernels, e.g.,

$$K(\boldsymbol{x},\boldsymbol{z}) = \prod_{\ell=1}^{d} \sum_{n=0}^{\infty} \alpha_{n,\ell} K_n(x_\ell, z_\ell), \qquad \alpha_{n,\ell} \geq 0.$$

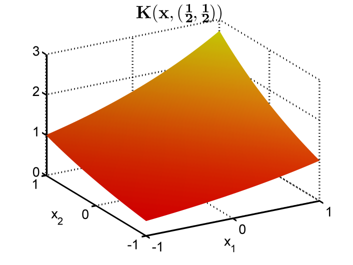
Power Series Kernels

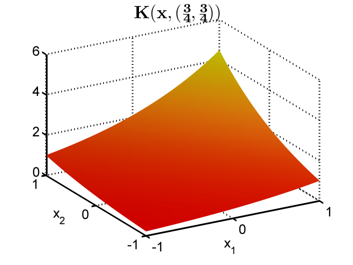
- Of the form [Zwi08]:

$$K(\boldsymbol{x},\boldsymbol{z}) = \sum_{\alpha\in\mathbb{N}_0^d} w_\alpha \frac{\boldsymbol{x}^\alpha}{\alpha!}\frac{\boldsymbol{z}^\alpha}{\alpha!}, \qquad \sum_{\alpha\in\mathbb{N}_0^d} \frac{w_\alpha}{\alpha!^2} < \infty,$$

Example (Exponential kernel)

$$K(\boldsymbol{x},\boldsymbol{z}) = e^{\boldsymbol{x}\cdot\boldsymbol{z}} = \sum_{n=0}^{\infty} \frac{1}{n!}(\boldsymbol{x}\cdot\boldsymbol{z})^n = \sum_{\alpha\in\mathbb{Z}^d} \frac{1}{|\alpha|!}\binom{|\alpha|}{\alpha}\boldsymbol{x}^\alpha\boldsymbol{z}^\alpha$$





Example (Taylor series kernels [ZS13])

$$K(x,z) = \frac{1}{(1-z\overline{x})^2} = \sum_{n=0}^{\infty}(n+1)z^n\overline{x}^n, \qquad \text{(Bergman kernel)}$$

$$K(x,z) = \frac{1}{1-z\overline{x}} = \sum_{n=0}^{\infty} z^n\overline{x}^n, \qquad \text{(Hardy or Szegő kernel)}$$

$$K(x,z) = -\frac{\ln(1-z\overline{x})}{z\overline{x}} = \sum_{n=0}^{\infty}\frac{1}{n+1}z^n\overline{x}^n, \qquad \text{(Dirichlet kernel)}$$

Here $x, z \in \mathbb{D}$, the open complex unit disk, i.e., $\mathbb{D} = \{x \in \mathbb{C} : |x| < 1\}$.

- Native spaces:
  - Bergman space $B^2 = L^2(\mathbb{D})$, the space of analytic functions in $\mathbb{D}$ that are square summable with respect to planar Lebesgue measure.
  - Hardy space H^2, the space of analytic functions in $\mathbb{D}$ with square summable Taylor coefficients. $H^2 \subset B^2$.
  - Dirichlet space $\mathcal{D}$, the space of analytic functions in D whose derivatives are in B^2.

- Other examples of series kernels are
  - Fourier-type series such as the periodic spline kernels,

$$K(x,z) = \sum_{n=1}^{\infty} \frac{2}{(2n\pi)^{2\beta}} \cos\left(2n\pi(x-z)\right).$$

  - Kernels specified via their Mercer/Hilbert–Schmidt series such as

$$\begin{aligned} K(x,z) &= \sum_{n=1}^{\infty} \frac{8}{(2n-1)^2\pi^2} \sin\left((2n-1)\frac{\pi x}{2}\right) \sin\left((2n-1)\frac{\pi z}{2}\right) \\ &= \min(x,z). \end{aligned}$$

Dot Product Kernels

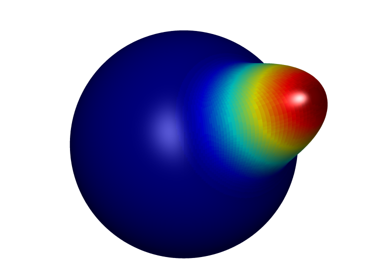
These kernels depend on $\boldsymbol{x}$ and $\boldsymbol{z}$ only through their dot product. They are also known as ridge functions (or zonal kernels if $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{S}^2$).

- Zonal kernels are of the form

$$K(\boldsymbol{x}, \boldsymbol{z}) = \tilde{\kappa}(\boldsymbol{x} \cdot \boldsymbol{z}), \quad \boldsymbol{x}, \boldsymbol{z} \in \mathbb{S}^2, \quad \tilde{\kappa} : [-1, 1] \to \mathbb{R}$$

Example (Spherical Gaussian kernel)

$$K(\boldsymbol{x}, \boldsymbol{z}) = \tilde{\kappa}(\boldsymbol{x} \cdot \boldsymbol{z}) = \mathrm{e}^{-2\varepsilon(1-\boldsymbol{x} \cdot \boldsymbol{z})}, \quad \tilde{\kappa}(t) = \mathrm{e}^{-2\varepsilon(1-t)}$$



Example (Polynomial kernel)

$$K(\boldsymbol{x},\boldsymbol{z}) = \left(\varepsilon + \boldsymbol{x}^T\boldsymbol{z}\right)^\beta, \qquad \boldsymbol{x},\boldsymbol{z} \in \mathbb{R}^d$$

- Plays an important role in machine learning.
- It is positive definite for all $\varepsilon \geq 0$ and $\beta \in \mathbb{N}_0$.
- The special case $\varepsilon = 0$ and $\beta = 1$ is known as the linear kernel.

Example (Sigmoid kernel)

$$K(\boldsymbol{x},\boldsymbol{z}) = \tanh(1 + \varepsilon\boldsymbol{x}^T\boldsymbol{z}), \qquad \boldsymbol{x},\boldsymbol{z} \in \mathbb{R}^d$$

- Also popular in machine learning.
- But this kernel is **not** positive definite for any choice of ε.

Remark

- *Ridge functions are discussed, e.g., in [CL99, Chapter 22] or [Pin13].*
- *They first arose in the context of computerized tomography [LS75].*
- *Zonal functions on spheres $\mathbb{S}^d$ in $\mathbb{R}^{d+1}$ can be analyzed using Mercer series.*
  - *The expansion can be written in terms of Legendre or Gegenbauer polynomials (and ultimately spherical harmonics).*
  - *This was done in, e.g., [Men99, RS96, XC92] (see also [SS02, Section 4.6]).*

Tensor Product Kernels

Weighted tensor products of various univariate kernels also produce general anisotropic kernels.

Example (Product of the Brownian motion kernel)

$$K(\boldsymbol{x},\boldsymbol{z}) = \prod_{\ell=1}^{d} \left(1 + \varepsilon_\ell \min(x_\ell, z_\ell)\right), \qquad \varepsilon_1 \geq \varepsilon_2 \geq \ldots \geq \varepsilon_d \geq 0,$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$.

- Neither radially nor translation invariant.
- Positive definite in $[0,1]^d$.

Remark

Such kernels play an important role in the theory of Monte-Carlo and quasi Monte-Carlo methods, where they are used to avoid the curse of dimensionality.

Remark

A related kernel is the kernel for fractional Brownian motion (see, e.g., [BTA04])

$$K(\boldsymbol{x},\boldsymbol{z}) = \frac{1}{2}\left(\|\boldsymbol{x}\|^{2\beta} + \|\boldsymbol{z}\|^{2\beta} - \|\boldsymbol{x}-\boldsymbol{z}\|^{2\beta}\right), \qquad \boldsymbol{x},\boldsymbol{z} \in \mathbb{R}^d.$$

However, this kernel is not a tensor product kernel.
For $\beta = \frac{1}{2}$ and $d = 1$ this simplifies to the standard Brownian motion kernel.

Remark

The linear covariance kernel is actually a radial kernel even though it is obtained by adding the kernels of two independent Brownian motions:

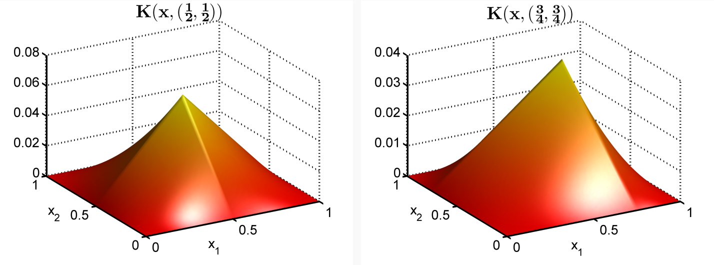
$$\begin{aligned} K(x,z) &= \min(x,z) + \min(1-x, 1-z) \\ &= \min(x,z) + 1 - \max(x,z) \\ &= 1 - |x-z|. \end{aligned}$$

We can also view this as a positive definite modification of the (conditionally negative definite) norm kernel.

Example

Brownian bridge product kernel

$$K(\boldsymbol{x}, \boldsymbol{z}) = \prod_{\ell=1}^{d} \left(\min(x_\ell, z_\ell) - x_\ell z_\ell\right)$$



- Neither radially nor translation invariant.
- Positive definite in $[0, 1]^d$.

Compactly Supported (Radial) Kernels

- One of the benefits of using compactly supported kernels is that — with an appropriate scaling — they lead to sparse kernel matrices.
- We concentrate on the Wendland family.
- Other families have been introduced by Buhmann, Gneiting or Wu (see [Fas07, Chapter 11]), as well as Johnson [Joh12].
- We will not do much with compactly supported kernels in this class.
- These kernels are discussed in detail in [Fas07, Wen05].
- Notation:
  - We use $r = \|\boldsymbol{x} - \boldsymbol{z}\|$ to indicate we are working with radial kernels, i.e., $K(\boldsymbol{x}, \boldsymbol{z}) = \kappa(\|\boldsymbol{x} - \boldsymbol{z}\|)$.
  - Below, $\doteq$ denotes equality up to a constant factor.

Original Wendland kernels [Wen95]

The family of kernels $\kappa_{d,k}$ includes

$$\kappa_{d,0} \doteq (1-r)_+^\ell$$
$$\kappa_{d,1} \doteq (1-r)_+^{\ell+1}\left((\ell+1)r+1\right)$$
$$\kappa_{d,2} \doteq (1-r)_+^{\ell+2}\left(\frac{\ell^2+4\ell+3}{3}r^2+(\ell+2)r+1\right)$$
$$\kappa_{d,3} \doteq (1-r)_+^{\ell+3}\left(\frac{\ell^3+9\ell^2+23\ell+15}{15}r^3+\frac{6\ell^2+36\ell+45}{15}r^2+(\ell+3)r+1\right)$$

- d: $K_{d,k}$ strictly positive definite on $\mathbb{R}^d \times \mathbb{R}^d$
- k: smoothness index, i.e., $\kappa_{d,k} \in C^{2k}(\mathbb{R})$
- ℓ: auxiliary variable with value $\ell = \lfloor \frac{d}{2} + k + 1 \rfloor$

Associated reproducing kernel Hilbert space:

$$\mathcal{H}_{K_{d,k}}(\Omega) = H^{k+(d+1)/2}(\mathbb{R}^d) \quad \text{(classical Sobolev space)}$$

Remark

- *The construction of [Wen95] with RKHS $H^{k+(d+1)/2}(\mathbb{R}^d)$ does not allow for Sobolev spaces of integer order when d is even.*
- *This, it appears that some functions are missing.*
- *This gap was filled when Schaback [Sch11] derived the so-called "missing" Wendland functions (see also [Hub12, CH14]).*

“Missing” Wendland kernels

Typical examples of the family $\kappa_{\ell,k}$ are (see [CSW14, Hub12, Sch11])

$$\kappa_{2,\frac{1}{2}}(r) \doteq (1+2r^2)\sqrt{1-r^2} + 3r^2 \log\left(\frac{r}{1+\sqrt{1-r^2}}\right)$$

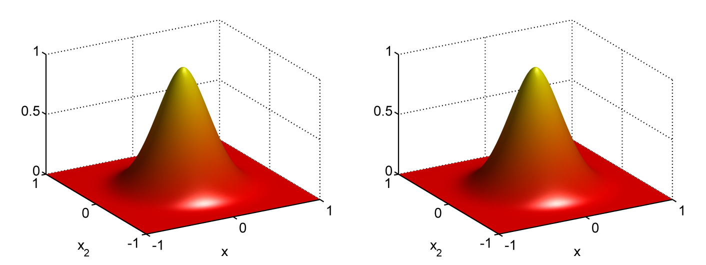
$$\kappa_{3,\frac{3}{2}}(r) \doteq \left(1-7r^2-\frac{81}{4}r^4\right)\sqrt{1-r^2} - \frac{15}{4}r^4(6+r^2)\log\left(\frac{r}{1+\sqrt{1-r^2}}\right)$$

These formulas hold for $r \in [0,1]$ and the functions are zero otherwise.

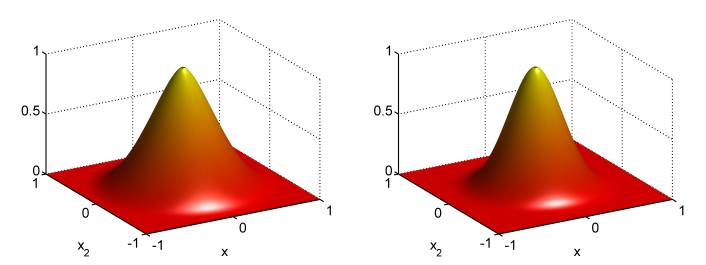
- ℓ: Sobolev smoothness, as before $\ell = \lfloor \frac{d}{2} + k + 1 \rfloor$
- k: half-integer, connected to smoothness of $\kappa_{\ell,k}$
- d: space dimension, but $K_{2,\frac{1}{2}}$ and $K_{3,\frac{3}{2}}$ both strictly positive definite on $\mathbb{R}^2 \times \mathbb{R}^2$

Associated reproducing kernel Hilbert space:

$$\mathcal{H}_{K_{2,\frac{1}{2}}}(\Omega) = H^2(\mathbb{R}^2), \quad \mathcal{H}_{K_{3,\frac{3}{2}}}(\Omega) = H^3(\mathbb{R}^2)$$



“Original” Wendland kernels: $\kappa_{3,1}$ (left, C^2) and $\kappa_{3,2}$ (right, C^4)



“Missing” Wendland kernels: $\kappa_{2,1/2}$ (left, C^1) and $\kappa_{3,3/2}$ (right, C^3)

Remark

- *Schaback [Sch11] derived the "missing" Wendland functions using fractional derivatives.*
- *In contrast to the "original" Wendland functions, these new functions are no longer polynomials on their support.*
- *Hubbert [Hub12] gives closed form representations of both the "original" and the "missing" Wendland functions in terms of associated Legendre functions (of the first and second kinds).*
- *Chernih [CSW14] showed that, as their smoothness increases, all (appropriately normalized) Wendland functions converge to Gaussians.*

General multiscale kernels [Opf06] are of the form

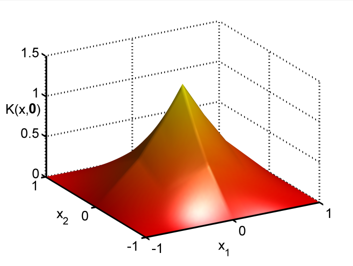
$$K(\boldsymbol{x},\boldsymbol{z}) = \sum_{j\geq 0} w_j K_j(\boldsymbol{x},\boldsymbol{z}) = \sum_{j\geq 0} w_j \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \phi(2^j\boldsymbol{x}-\boldsymbol{k})\phi(2^j\boldsymbol{z}-\boldsymbol{k}),$$

with $w_j > 0$ and ϕ a compactly supported (possibly refinable) function whose shifts (at level j) produces the single-scale kernel K_j.

Example (Multiscale piecewise linear kernel)

$$K(\boldsymbol{x},\boldsymbol{z}) = \sum_{j=0}^{3} 2^{-2j} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \phi(2^j\boldsymbol{x}-\boldsymbol{k})\varphi(2^j\boldsymbol{z}-\boldsymbol{k})$$

with $\phi(\boldsymbol{x}) = \prod_{\ell=1}^{d}(1-x_\ell)_+$



- [Opf06] described the RKHSs of these kernels.
- He used them in wavelet-like applications such as image compression.
- Very little work has been performed otherwise with these kernels.

Space-Time Kernels

- Many problems have both a spatial as well as a temporal component, so the idea to construct and use space-time kernel is natural.
- The most common approach is to use a tensor product kernel that factors into a spatial and a temporal component.
- But sometimes the data does not seem to allow such separability since it contains spatio-temporal interactions which a separable model would not be able to pick up on (see, e.g., [CH99, GGG07]).

In the RBF literature these kernels are rare.

Li & Mao [LM11] solved an ill-posed inverse heat conduction problem using an anisotropic IMQ kernel

$$K((\boldsymbol{x},s),(\boldsymbol{z},t)) = \frac{1}{\sqrt{1+\varepsilon^2\|\boldsymbol{x}-\boldsymbol{z}\|^2+\gamma^2(s-t)^2}}, \;\; \boldsymbol{x},\boldsymbol{z}\in\mathbb{R}^d, \;\; s,t\in\mathbb{R},$$

where $d = 1, 2$.

The spatial coordinates are augmented by an additional time coordinate, but note the use of two different scale parameters.

In the statistics literature space-time kernels are more common.

- Stein uses kernels that are translation invariant in both space and time, i.e., of the form $K((\boldsymbol{x},s),(\boldsymbol{z},t)) = \widetilde{K}(\boldsymbol{x}-\boldsymbol{z}, s-t)$. He derives
  - generalizations of Matérn kernels [Ste05], and
  - power law covariance functions (which generalize polyharmonic splines) [Ste13].
- Porcu [PMB07] allows for spatial anisotropy with temporal translation invariance leading to kernels such as, e.g.,

$$K((\boldsymbol{x},s),(\boldsymbol{z},t)) = \frac{\exp\left(-\frac{|s-t|^2}{K_{\text{space}}(\boldsymbol{x},\boldsymbol{z})}\right)}{\sqrt{K_{\text{space}}(\boldsymbol{x},\boldsymbol{z})}}, \qquad \boldsymbol{x},\boldsymbol{z} \in \mathbb{R}^d,\ s,t \in \mathbb{R},$$

where $K_{\text{space}}(\boldsymbol{x},\boldsymbol{z}) = \log\left(2 + \frac{1}{2}\left(2\varepsilon(\boldsymbol{x}+\boldsymbol{z}) - \frac{1+\varepsilon(\boldsymbol{x}+\boldsymbol{z})}{1+\varepsilon(\boldsymbol{x}-\boldsymbol{z})}\right)\right)$.

Learned Kernels

In the machine learning literature one finds kernels that are learned directly from the data.

- Micchelli and Pontil [MP05] suggest learning the kernel via regularization techniques.
  - They start with a — possibly uncountable — set $\mathbb{K}$ of kernels and then determine the optimal kernel for a given set of N pieces of data $\{(\boldsymbol{x}_i, y_i) : i = 1, \ldots, N\}$ as a finite convex combination of kernels from $\mathbb{K}$.
  - The set $\mathbb{K}$ is assumed to be compact and convex, and then the optimal learned kernel is obtained by solving a convex optimization problem.
  - Once the kernel K has been found, the kernel approximation is obtained by solving a finite-dimensional convex optimization problem.
- Lanckriet [LCB$^+$04] suggests that the kernel matrix (instead of the actual kernel) can be learned from the given data by employing semi-definite programming techniques.

Designer Kernels

Some ideas to obtain specially designed custom kernels (or designer kernels):

- Use the basic properties of kernels discussed in Chapter 2, such as adding, multiplying and taking positive linear combinations of kernels.
- Use ideas such as composition with multiply or completely monotone functions (see [Fas07]) to construct new radial kernels.
- Build a kernel via Mercer's theorem by combining an appropriate sequence of "eigenvalues" λ_n with a given set of orthogonal functions.
  - This may mean that the closed form of the kernel may not be known in this case.
  - Good example: iterated Brownian bridge kernels (see later).

References I

[Bax06] B. J. C. Baxter, *Scaling radial basis functions via Euclidean distance matrices*, Computers & Mathematics with Applications **51** (2006), no. 8, 1163–1170.

[BDL10] R. K. Beatson, O. Davidov, and J. Levesley, *Error bounds for anisotropic RBF interpolation.*, Journal of Approximation Theory **162** (2010), 512–527.

[BLRS14] Mira Bozzini, Licia Lenarduzzi, Milvia Rossini, and Robert Schaback, *Interpolation with variably scaled kernels*, IMA Journal of Numerical Analysis (2014).

[BTA04] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, Dordrecht, 2004.

[CBM+03] J. C Carr, R. K Beatson, B. C McCallum, W. R Fright, T. J McLennan, and T. J Mitchell, *Smooth surface reconstruction from noisy range data*, Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia (New York), GRAPHITE '03, ACM, 2003, pp. 119–126.

References II

[CH99] Noel Cressie and Hsin-cheng Huang, *Classes of nonseparable, spatio-temporal stationary covariance functions*, Journal of the American Statistical Association **94** (1999), 1330–1340.

[CH14] Andrew Chernih and Simon Hubbert, *Closed form representations and properties of the generalised Wendland functions*, Journal of Approximation Theory **177** (2014), 17–33.

[CL99] E. W. Cheney and W. A. Light, *A Course in Approximation Theory*, Brooks/Cole (Pacific Grove, CA), 1999.

[CLMM06] G. Casciola, D. Lazzaro, L. B. Montefusco, and S. Morigi, *Shape preserving surface reconstruction using locally anisotropic radial basis function interpolants*, Computers & Mathematics with Applications **51** (2006), no. 8, 1185–1198.

[CSW14] A. Chernih, I. H. Sloan, and R. S. Womersley, *Wendland functions with increasing smoothness converge to a Gaussian*, Advances in Computational Mathematics **40** (2014), no. 1, 185–200.

References III

[Fas07] G. E. Fasshauer, *Meshfree Approximation Methods with* MATLAB, Interdisciplinary Mathematical Sciences, vol. 6, World Scientific Publishing Co., Singapore, 2007.

[FHW12a] G. E. Fasshauer, F. J. Hickernell, and H. Woźniakowski, *On dimension-independent rates of convergence for function approximation with Gaussian kernels*, SIAM J. Numer. Anal. **50** (2012), no. 1, 247–271.

[FHW12b] G. E. Fasshauer, Fred J. Hickernell, and H. Woźniakowski, *Average case approximation: convergence and tractability of Gaussian kernels*, Monte Carlo and Quasi-Monte Carlo Methods 2010 (H. Woźniakowski and L. Plaskota, eds.), Springer Proceedings in Mathematics and Statistics, no. 23, Springer, Berlin, 2012, pp. 309—324.

[FLW06] B. Fornberg, E. Larsson, and G. Wright, *A new class of oscillatory radial basis functions*, Computers & Mathematics with Applications **51** (2006), no. 8, 1209–1222.

[Gen02] Marc G. Genton, *Classes of kernels for machine learning: A statistics perspective*, J. Mach. Learn. Res. **2** (2002), 299–312.

References IV

[GGG07] Tilmann Gneiting, Marc G. Genton, and Peter Guttorp, *Geostatistical space-time models, stationarity, separability and full symmetry*, Statistics of Spatio-Temporal Systems (B. Finkenstaedt, L. Held, and V. Isham, eds.), Chapman & Hall/CRC Press, 2007, pp. 151–175.

[Hub12] Simon Hubbert, *Closed form representations for a class of compactly supported radial basis functions*, Advances in Computational Mathematics **36** (2012), no. 1, 115–136.

[Joh12] Michael J. Johnson, *Compactly supported, piecewise polyharmonic radial functions with prescribed regularity*, Constructive Approximation **35** (2012), no. 2, 201–223.

[LCB+04] Gert R. G. Lanckriet, Nello Cristianini, Peter Bartlett, Laurent El Ghaoui, and Michael I. Jordan, *Learning the kernel matrix with semidefinite programming*, Journal of Machine Learning Research **5** (2004), 27–72.

[LM11] Z. Li and X.-Z. Mao, *Global space-time multiquadric method for inverse heat conduction problem*, International Journal for Numerical Methods in Engineering **85** (2011), no. 3, 355–379.

References V

[LS75] B. F. Logan and L. A. Shepp, *Optimal reconstruction of a function from its projections*, Duke Mathematical Journal **42** (1975), no. 4, 645–659.

[Men99] V. A. Menegatto, *Strict positive definiteness on spheres*, Analysis **19** (1999), no. 3, 217—233.

[MP05] Charles A. Micchelli and Massimiliano Pontil, *Learning the kernel function via regularization*, J. Mach. Learn. Res. **6** (2005), 1099–1125.

[NW08] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems Volume 1: Linear Information*, EMS Tracts in Mathematics, no. 6, European Mathematical Society, 2008.

[Opf06] R. Opfer, *Multiscale kernels*, Advances in Computational Mathematics **25** (2006), no. 4, 357–380.

[Pin13] Allan Pinkus, *On ridge functions*, Azerbaijan Journal of Mathematics **3** (2013), no. 2.

References VI

[PMB07] E. Porcu, J. Mateu, and M. Bevilacqua, *Covariance functions that are stationary or nonstationary in space and stationary in time*, Statistica Neerlandica **61** (2007), no. 3, 358–382.

[RS96] A. Ron and X. Sun, *Strictly positive definite functions on spheres*, Math. Comp. **65** (1996), no. 216, 1513–1530.

[Sch11] Robert Schaback, *The missing Wendland functions*, Advances in Computational Mathematics **34** (2011), no. 1, 67–81.

[SS02] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press, 2002.

[Ste05] Michael L Stein, *Space-time covariance functions*, Journal of the American Statistical Association **100** (2005), no. 469, 310–321.

[Ste13] Michael L. Stein, *On a class of space-time intrinsic random functions*, Bernoulli **19** (2013), no. 2, 387–408.

References VII

[Wah90] Grace Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.

[Wen95] H. Wendland, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Advances in Computational Mathematics **4** (1995), no. 1, 389–396.

[Wen05] ______, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, vol. 17, Cambridge University Press, 2005.

[XC92] Y. Xu and E. W. Cheney, *Strictly positive definite functions on spheres*, Proc. Amer. Math. Soc. **116** (1992), 977–981.

[ZS13] Barbara Zwicknagl and Robert Schaback, *Interpolation and approximation in Taylor spaces*, Journal of Approximation Theory **171** (2013), 65–83.

[Zwi08] B. Zwicknagl, *Power series kernels*, Constructive Approximation **29** (2008), no. 1, 61–84.