Translation-invariant kernels

Toni Karvonen

Department of Mathematics and Statistics
University of Helsinki, Finland

Workshop on Kernel Approximations and and Space-Filling
Cardiff, UK

1 July 2022

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE

Table of contents

Translation-invariant kernels

Suppose that $\Phi\colon \mathbb{R} \to \mathbb{R}$ is continuous and integrable.

Translation-invariant kernel

A positive-definite kernel $K\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is *translation-invariant* if

$$K(x, y) = \Phi(x - y) \quad \text{for all} \quad x, y \in \mathbb{R}.$$

$$K(x, y) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\,|x - y|}{\lambda} \right)^{\nu} \mathcal{K}_\nu \left(\frac{\sqrt{2\nu}\,|x - y|}{\lambda} \right) \qquad \text{(Matérn)}$$

$$K(x, y) = \frac{1}{1 + (x - y)^2/\lambda^2} \qquad \text{(Cauchy)}$$

$$K(x, y) = \exp\left(-\frac{(x - y)^2}{2\lambda^2} \right) \qquad \text{(Gaussian)}$$

Reproducing kernel Hilbert spaces

Let Ω be a set.

Reproducing kernel Hilbert space

Every PD kernel $K\colon \Omega \times \Omega \to \mathbb{R}$ induces a unique *reproducing kernel Hilbert space* $H_K(\Omega)$ of functions $f\colon \Omega \to \mathbb{R}$. The RKHS is equipped with the inner product $\langle \cdot, \cdot \rangle_K$ and the associated norm $\|\cdot\|_K$.

Restriction of an RKHS

Let $\Omega \subset \mathbb{R}$. The RKHS $H_K(\Omega)$ is a *restriction* of $H_K(\mathbb{R})$:

$$f \in H_K(\Omega) \iff \text{there is en } \textit{extension } f_e \in H_K(\mathbb{R}) \text{ s.t. } f_e|_\Omega = f.$$

Constants in the RKHS

Recall that we assume $\Phi\colon \mathbb{R} \to \mathbb{R}$ is integrable and $K(x, y) = \Phi(x - y)$.

Lemma [standard]

$H_K(\mathbb{R}) \subset L^2(\mathbb{R})$

Corollary

Non-trivial constant functions are not elements of $H_K(\mathbb{R})$.

When are constant functions in $H_K(\Omega)$ if $\Omega \subset \mathbb{R}$ is a bounded interval?

Length-scale estimation and flat limits

In Gaussian process interpolation, the *maximum likelihood estimate* of the length-scale $\lambda > 0$ is

$$\lambda_{\mathrm{ML}} = \arg\min_{\lambda > 0} \left[Y^\top K_\lambda(X, X) Y + \log\det K_\lambda(X, X) \right],$$

where $Y \in \mathbb{R}^n$ are the observed data at points $X = \{x_1, \ldots, x_n\}$.

Theorem [Karvonen & Oates 2022]

Suppose that

- The data are *constant*, in that $Y = (c, \ldots, c) \in \mathbb{R}^n$ for some $c \in \mathbb{R}$;
- $1 \in H_K(\Omega)$ for some interval Ω that contains the origin.

Then

$$\lambda_{\mathrm{ML}} = \infty.$$

Karvonen & Oates (2022). Maximum likelihood estimation in Gaussian process regression is ill-posed. arXiv:2203.09179.

Locality

Tentative Definition I — locality via constant functions

A translation-invariant kernel K is *local* if $1 \in H_K(\Omega)$ for some bounded interval $\Omega \subset \mathbb{R}$.

Tentative Definition II — locality via bounded support

A translation-invariant kernel K is *local* if $H_K(\mathbb{R})$ contains a function supported on some bounded interval $\Omega \subset \mathbb{R}$.

Tentative Definition III — locality via length-scale

A translation-invariant kernel K is *local* if $H_K(\mathbb{R})$ does not depend on the length-scale λ.

Example: Matérns are local by any of these definitions, Cauchy and Gaussian by none.

Table of contents

Analyticity

Analytic function

A function $f : \mathbb{R} \to \mathbb{R}$ is *analytic with convergence radius* $r > 0$ if for every $x_0 \in \mathbb{R}$ the function is given by its Taylor series at x_0:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \quad \text{for all} \quad |x - x_0| < r.$$

Theorem [Sun & Zhou 2008, Theorem 1]

If K is a translation-invariant kernel of the form $K(x, y) = \varphi((x - y)^2)$ for an analytic function φ with convergence radius $r > 0$, then every function in $H_K(\mathbb{R})$ is analytic with convergence radius $\frac{1}{2}\sqrt{r}$.

Sun & Zhou (2008). Reproducing kernel Hilbert spaces associated with analytic translation-invariant Mercer kernels. *Journal of Fourier Analysis and Applications*.

Globality

Analyticity implies *globality* or *non-locality*.

Theorem [Sun & Zhou 2008, Theorem 3]

If K is an analytic translation-invariant kernel and $\Omega \subset \mathbb{R}$ is any interval, then $H_K(\Omega) = H_K(\mathbb{R})$:

$$f \in H_K(\mathbb{R}) \iff f|_\Omega \in H_K(\Omega).$$

Corollary

If K is an analytic translation-invariant kernel and $\Omega \subset \mathbb{R}$ is any interval, then $1 \notin H_K(\Omega)$.

Proof. $f \equiv 1$ is analytic but $f \notin L_2(\mathbb{R}) \supset H_K(\mathbb{R})$. Thus $f|_\Omega \notin H_K(\Omega)$. □

Example: Gaussian and Cauchy are analytic and hence global.

Fourier characterisation

Let

$$\widehat{f}(t) = \int_{\mathbb{R}} f(x) e^{-ixt} \, \mathrm{d}t$$

be the Fourier transform.

Theorem [standard]

Let $K(x, y) = \Phi(x - y)$ be a translation-invariant kernel. Then $H_K(\mathbb{R})$ contains every function $f \in L^2(\mathbb{R})$ such that

$$\|f\|_K^2 = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{|\widehat{f}(t)|^2}{\widehat{\Phi}(t)} \, \mathrm{d}t < \infty. \tag{1}$$

To prove that $f : \Omega \to \mathbb{R}$ is an element of $H_K(\Omega)$ for $\Omega \subsetneq \mathbb{R}$, one thus has to construct an extension $f_e \in L^2(\mathbb{R})$ which satisfies the Fourier decay condition (1).

Examples of Fourier transforms

K = Matérn

$$\widehat{\Phi}(t) \propto \left(\frac{2\nu}{\lambda^2} + t^2\right)^{-(\nu+1/2)}$$

K = Cauchy

$$\widehat{\Phi}(t) \propto \exp(-\lambda|t|)$$

K = Gaussian

$$\widehat{\Phi}(t) \propto \exp\left(-\frac{\lambda^2 t^2}{2}\right)$$

Note: Exponentially decaying $\widehat{f} \implies f$ is analytic

Moment determinacy

The Fourier transform $\widehat{\Phi}$ of K defines the *spectral measure* $\alpha(t) = \widehat{\Phi}(t)$ s.t.

$$K(x, y) = \Phi(x - y) = \int_{\mathbb{R}} e^{i(x-y)t} \alpha(\mathrm{d}t).$$

The spectral measure has the moments

$$c_n = \Phi^{(n)}(0) = i^n \int_{\mathbb{R}} t^n \alpha(\mathrm{d}t).$$

The *Hamburger moment problem* for α is said to be *determinate* if there is no other measure with the moments $(c_n)_{n=0}^{\infty}$.

Theorem [Dette & Zhigljavsky 2021, Theorem 1.1]

Let $\Omega \subset \mathbb{R}$ be any interval. If the Hamburger moment problem for α is determinate, then $1 \notin H_K(\Omega)$.

Dette & Zhigljavsky (2021). Reproducing kernel Hilbert spaces, polynomials, and the classical moment problem. *SIAM/ASA Journal on Uncertainty Quantification*.

Beurling–Malliavin for translation-invariant kernels

A fact of Fourier analysis: If f has bounded support, $\widehat{f}$ cannot be too small.

Beurling–Malliavin for kernels (tentative)

Let $\Omega \subset \mathbb{R}$ be a bounded interval. Then $1 \in H_K(\Omega)$ if there is a Lipschitz function $\omega\colon \mathbb{R} \to [0, \infty)$ and a constant $\alpha > 0$ such that

$$e^{-\omega(\xi)} \leq \alpha\, \widehat{\Phi}(\xi) \quad \text{for all} \quad \xi \in \mathbb{R} \tag{BM-1}$$

and

$$\int_{\mathbb{R}} \frac{\omega(\xi)}{1+\xi^2}\, \mathrm{d}\xi < \infty. \tag{BM-2}$$

Example: $\widehat{\Phi}(\xi) = (1+\xi^2)^{-p}$ for $p > 1/2$. $[\omega(\xi) = \log(1+\xi^2)^p]$

Example: $\widehat{\Phi}(\xi) = \exp(-|\xi|^\beta)$ for $\beta \in (0,1)$. $[\omega(\xi) = |\xi|^\beta]$

Non-example: $\widehat{\Phi}(\xi) = \exp(-|\xi|^\beta)$ for $\beta \geq 1$. $[\omega(\xi) = |\xi|^\beta \gtrsim |\xi|]$

Beurling & Malliavin (1962). On Fourier transforms of measures with compact support. *Acta Mathematica*.

Table of contents

Orthonormal basis of an RKHS

Basic Result

Let I be an infinitely countable index set. If the RKHS $H_K(\mathbb{R})$ is separable and $(\psi_m)_{m\in I}$ is any orthonormal basis of $H_K(\mathbb{R})$, then

$$K(x,y) = \sum_{m\in I} \psi_m(x)\psi_m(y) \quad \text{for all} \quad x,y \in \mathbb{R}.$$

Theorem [Tronarp & Karvonen 2022, Theorem 1.1]

Let $K(x,y) = \Phi(x-y)$ be a translation-invariant kernel. If $(\varphi_m)_{m\in I}$ is an orthonormal basis of $L^2(\mathbb{R})$ and h s.t. $|\widehat{h}(t)| = \widehat{\Phi}(t)^{1/2}$, then

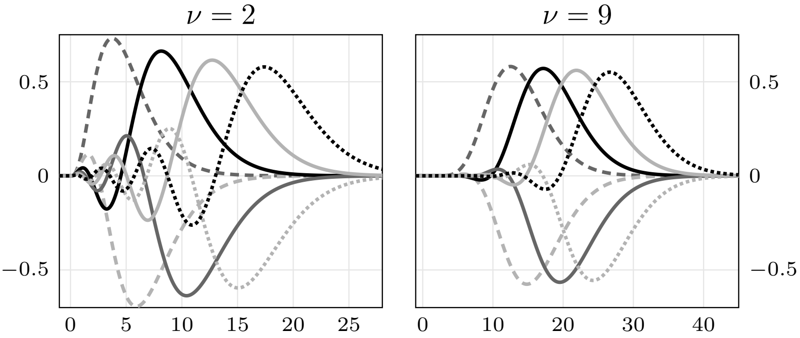
$$\psi_m(x) = \int_{\mathbb{R}} h(x-\tau)\varphi_m(\tau)\,\mathrm{d}\tau$$

form an orthonormal basis of $H_K(\mathbb{R})$.

Tronarp & Karvonen (2022). Orthonormal expansions for translation-invariant kernels. arXiv:2206.08648.

Orthonormal basis for Matérn-ν

$$\psi_{\nu,m}(x) = \frac{1}{\sqrt{2\lambda}} \frac{\nu!}{\sqrt{(2\nu)!}} \sum_{k=0}^{\nu+1} \binom{\nu+1}{k} (-1)^k \underbrace{\varphi_{m+k,\lambda}(x)}_{\text{Laguerre func.}} \qquad [m \in \mathbb{Z}]$$



Tronarp & Karvonen (2022). Orthonormal expansions for translation-invariant kernels. arXiv:2206.08648.

Laguerre functions

The Laguerre functions are

$$\varphi_{m,\lambda}(x) = \mathrm{L}_m(2\lambda x)e^{-\lambda x}\mathbf{1}_{[0,\infty)}(x),$$
$$\varphi_{-m-1,\lambda}(x) = -\mathrm{L}_m(-2\lambda x)e^{\lambda x}\mathbf{1}_{(-\infty,0)}(x)$$

for $m \in \mathbb{N}_0$.

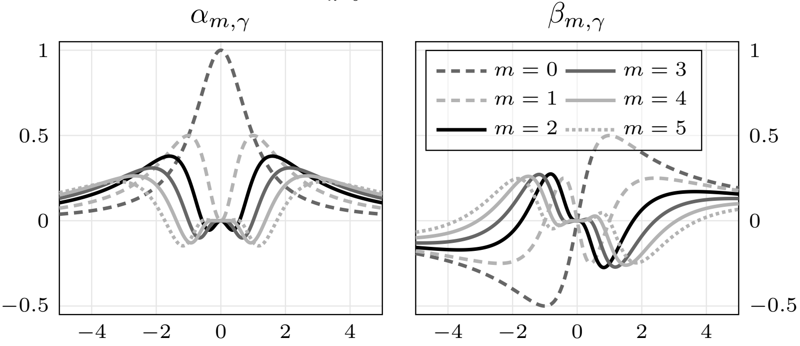
For non-negative indices the RKHS basis functions simplify to

$$\psi^{(\nu)}_{m,\lambda}(x) = \frac{\nu!}{\sqrt{(2\nu)!}}\frac{m!}{(m+\nu+1)!}(2\lambda x)^{\nu+1}\mathrm{L}^{(\nu+1)}_m(2\lambda x)e^{-\lambda x}\mathbf{1}_{[0,\infty)}(x).$$

Orthonormal basis for Cauchy

$$\alpha_{m,\gamma=1/\lambda}(x) = \frac{\gamma}{2} \frac{(-1)^m (ix)^m}{(x^2+\gamma^2)^{m+1}} \sum_{k=0}^{m+1} \binom{m+1}{k} (ix)^k \gamma^{m+1-k} \left(1 - (-1)^{m+1-k}\right) \quad [m \in \mathbb{N}_0]$$

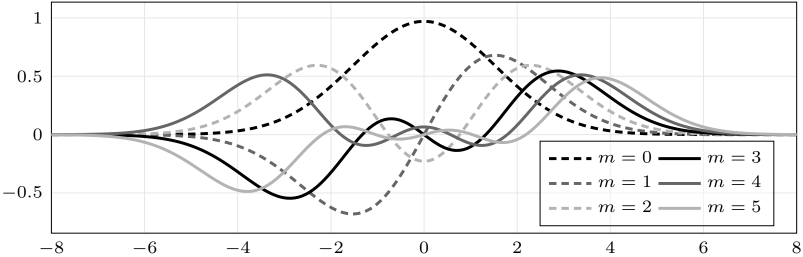
$$\beta_{m,\gamma=1/\lambda}(x) = \frac{\gamma}{i2} \frac{(-1)^m (ix)^m}{(x^2+\gamma^2)^{m+1}} \sum_{k=0}^{m+1} \binom{m+1}{k} (ix)^k \gamma^{m+1-k} \left(1 + (-1)^{m+1-k}\right) \quad [m \in \mathbb{N}_0]$$



Tronarp & Karvonen (2022). Orthonormal expansions for translation-invariant kernels. arXiv:2206.08648.

Orthonormal basis for Gaussian

$$\psi_m(x) = \left(\frac{2\sqrt{2}}{3}\right)^{1/2} \sqrt{\frac{1}{6^m m!}} \exp\left(-\frac{\lambda^2 x^2}{3}\right) \mathrm{H}_m\left(\frac{2\lambda x}{\sqrt{3}}\right) \qquad [m \in \mathbb{N}_0]$$



Tronarp & Karvonen (2022). Orthonormal expansions for translation-invariant kernels. arXiv:2206.08648.

Globally analytic kernels may be problematic

Let

$$\mathbb{V}(x \mid X_n) = K(x,x) - K(x,X_n)^\mathsf{T} K(X_n,X_n)^{-1} K(x,X_n)$$

be the Gaussian process posterior variance given data at n points X_n.

Theorem [Karvonen 2022]

Let $\Omega = [-1,1]$ and $X_n = \{x_1, \ldots, x_n\} \subset \Omega$ be **any** pairwise distinct points. If $K(x,y) = \exp(-(x-y)^2/(2\lambda^2))$ is the Gaussian kernel, then

$$C_{\lambda,1} n^{-1/2} \left(\frac{e}{4\lambda^2} \right)^n n^{-n} \leq \sup_{x \in [-1,1]} \mathbb{V}(x \mid X_n) \leq C_{\lambda,2} n^{-1} \left(\frac{8e}{\lambda^2} \right)^n n^{-n}$$

for certain positive constants $C_{\lambda,1}$ and $C_{\lambda,2}$.

Karvonen (2022). Worst-case analysis of approximation in the Hilbert space of the Gaussian kernel. In preparation.

Locally convergenc analytic kernels are probably fine

Theorem

Let $\Omega = [-a, a]$ and $X_n = \{x_1, \ldots, x_n\} \subset \Omega$ be **any** pairwise distinct points. If $K(x, y) = 1/(1 + (x-y)^2/\lambda^2)$ is the Cauchy kernel, then

$$\sup_{x \in [-a,a]} \mathbb{V}(x \mid X_n) \leq C_\lambda \, n^{-1/2} \left(\frac{16a^2}{\lambda^2} \right)^n \tag{2}$$

for a certain positive constant C_λ.

Proof. Error formula for polynomial interpolation + derivative bound for RKHS functions + Stirling's approximation of the factorial. □

Observe: The rhs in (2) tends to zero if and only if $a < \frac{1}{4}\lambda$.

Local kernels are fine

Theorem [standard]

Let K be a Matérn kernel. Then

$$\sup_{x \in [-1,1]} \mathbb{V}(x \mid X_n) \to 0 \quad \text{as} \quad n \to \infty$$

if and only if X_n is *space-filling*, in that

$$h(X_n) = \sup_{x \in [-1,1]} \min_{i=1,\ldots,n} |x - x_i| \to 0 \quad \text{as} \quad n \to \infty.$$

More Precise Theorem [standard]

If $(X_n)_{n=1}^{\infty}$ is *quasi-uniform* on sufficiently regular and bounded $\Omega \subset \mathbb{R}^d$ and K is a d-dimensional translation-invariant Matérn-ν kernel, then

$$\sup_{x \in \Omega} \mathbb{V}(x \mid X_n) \asymp n^{-2\nu/d}.$$

Conclusion

- Finitely differentiable kernels are local, analytic kernels are non-local.
- It may be dangerous to use a non-local translation-invariant kernel.
- If you really have to or want to, use Cauchy not Gaussian.

---

1. **Beurling & Malliavin (1962)**. On Fourier transforms of measures with compact support. *Acta Mathematica*.
2. **Saitoh (1997)**. *Integral Transforms, Reproducing Kernels and Their Applications*.
3. **Sun & Zhou (2008)**. Reproducing kernel Hilbert spaces associated with analytic translation-invariant Mercer kernels. *Journal of Fourier Analysis and Applications*.
4. **Minh (2010)**. Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. *Constr. Approx*.
5. **Dette & Zhigljavsky (2021)**. Reproducing kernel Hilbert spaces, polynomials, and the classical moment problem. *SIAM/ASA Journal on Uncertainty Quantification*.
6. **Karvonen & Oates (2022)**. Maximum likelihood estimation in Gaussian process regression is ill-posed. arXiv:2203.09179.
7. **Tronarp & Karvonen (2022)**. Orthonormal expansions for translation-invariant kernels. arXiv:2206.08648.