Mean Dimension of Radial Basis Functions

Art B. Owen*

Stanford University

Based on joint work with:

Christopher Hoyt, Stanford University

Opinions are my own, and not those of Stanford, the NSF, or Hitachi, Ltd.

arXiv:2302.09689

Radial basis functions

$$\phi(\|\boldsymbol{x} - \boldsymbol{c}\|) \quad \text{for } \boldsymbol{x}, \boldsymbol{c} \in \mathbb{R}^d$$

E.g., scattered data interpolation

$$f(\boldsymbol{x}) \approx \tilde{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \beta_i \phi\bigl(\|\boldsymbol{x} - \boldsymbol{x}_i\|\bigr)$$

Versus ridge functions

$$\phi(\boldsymbol{x}^\mathsf{T}\theta) \quad \theta^\mathsf{T}\theta = 1$$

What we find

Some RBFs become essentially additive in high dimensions

Why we care

That limits their usefulness as

- approximators
- covariance functions

Other RBFs

Not necessarily additive

Same for ridge functions

Concentration of measure

High dimensional and Lipschitz

$\Longrightarrow$ Nearly constant*

Donoho (2000)

Variation around constant

Depends

*$\boldsymbol{x} \sim \mathcal{N}(0, I_d)$ and f Lipschitz

$\Rightarrow f(\boldsymbol{x})$ has Gaussian tails

RBF interpolation

$$\tilde{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \beta_i \phi(\|\boldsymbol{x} - \boldsymbol{x}_i\|)$$

Scattered data $\boldsymbol{x}_i$

e.g., IID

Choose $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$ so that

$$\tilde{f}(\boldsymbol{x}_i) = f(\boldsymbol{x}_i)$$

Doable for

Strictly positive definite RBFs

some ‘conditionally positive definite’ RBFs

See Fasshauer (2007)

Polynomial interpolation not always effective

Some strictly PSD RBFs

$r = \|\boldsymbol{x} - \boldsymbol{c}\|$	e.g., $\boldsymbol{c} = \boldsymbol{x}_i$
$\phi(r) = \exp(-\theta^2 r^2)$	Gaussian / squared exponential
$\phi(r) = (1 + \theta^2 r^2)^p \quad p < 0$	Generalized inverse multiquadric
$\phi(r) = (1 + \theta^2 r^2)^{-1/2}$	Inverse multiquadric
$\phi(r) = (1 + \theta^2 r^2)^{1/2}$	Multiquadric Hardy (1971)

Schoenberg condition

$$\phi(r) = \int_0^\infty e^{-r^2 t^2} \mu(\mathrm{d}t)$$

For strictly PSD RBF, any $d \geqslant 1$

Fasshauer (2007) Theorem 3.8

$\phi(\cdot) \geqslant 0$ nondecreasing

ANOVA

For $x_j \sim F_j$ independent & $0 < \mathbb{E}(f(\boldsymbol{x})^2) < \infty$

Fisher & Mackenzie, Hoeffding, Sobol', Efron & Stein

Notation

$[d] \equiv \{1, 2, \dots, d\}$ For $u \subseteq [d]$, $\boldsymbol{x}_u = (x_j)_{j \in u}$

ANOVA decomp

$$f(\boldsymbol{x}) = \sum_{u \subseteq [d]} f_u(\boldsymbol{x})$$

$f_u(\boldsymbol{x})$ only depends on $\boldsymbol{x}_u$

Recursive definition

$$f_\varnothing(\boldsymbol{x}) = \mathbb{E}(f(\boldsymbol{x}))$$

$$f_u(\boldsymbol{x}) = \mathbb{E}\Biggl(f(\boldsymbol{x}) - \sum_{v \subsetneq u} f_v(\boldsymbol{x}) \mid \boldsymbol{x}_u\Biggr) = \mathbb{E}(f(\boldsymbol{x}) \mid \boldsymbol{x}_u) - \sum_{v \subsetneq u} f_v(\boldsymbol{x})$$

ANOVA continued

Recall $f(\boldsymbol{x}) = \sum_{u\subseteq[d]} f_u(\boldsymbol{x})$

From orthogonality

$$\sigma^2 = \mathrm{Var}(f(\boldsymbol{x}))$$

$$\sigma_u^2 = \mathrm{Var}(f_u(\boldsymbol{x}))$$

$$\sigma^2 = \sum_{u\subseteq[d]} \sigma_u^2$$

Mean dimension

$$\nu(f) = \sum_{u \subseteq [d]} |u| \frac{\sigma_u^2}{\sigma^2}$$

E.g., for $d = 3$

$$\nu(f) = \frac{\sigma_{\{1\}}^2 + \sigma_{\{2\}}^2 + \sigma_{\{3\}}^2 + 2\sigma_{\{1,2\}}^2 + 2\sigma_{\{1,3\}}^2 + 2\sigma_{\{2,3\}}^2 + 3\sigma_{\{1,2,3\}}^2}{\sigma^2}$$

Additivity

$$R_{\mathrm{add}}^2 = \frac{1}{\sigma^2} \sum_{j=1}^{d} \sigma_{\{j\}}^2$$

$$\nu(f) \leqslant 1 + \epsilon \implies R_{\mathrm{add}}^2 \geqslant 1 - \epsilon$$

Sobol' indices

$$\bar{\tau}_j^2 = \sum_{v:\, j \in v} \sigma_v^2$$

Jansen

$$\bar{\tau}_j^2 = \frac{1}{2}\mathbb{E}\Bigl((f(\boldsymbol{x}) - f(\boldsymbol{z}_j{:}\boldsymbol{x}_{-j}))^2\Bigr)$$

Liu & O

$$\underbrace{\sum_{u \subseteq [d]} |u|\sigma_u^2}_{2^d \text{ terms}} = \underbrace{\sum_{j=1}^{d} \bar{\tau}_j^2}_{d \text{ terms}}$$

Pick freeze

For $\boldsymbol{z}_j{:}\boldsymbol{x}_{-j}$ replace x_j by independent z_j

Main result

For generalized multiquadrics, independent x_j

$$f(\boldsymbol{x}) = (a + \|\boldsymbol{x} - \boldsymbol{c}\|^2)^p \qquad a \geqslant 0 \qquad p < 1$$

Then under conditions

$$\nu(f) = 1 + O(d^{-1})$$ With explicit constant

More notation

$$z_j = \begin{cases} a + (x_1 - c_1)^2, & j = 1 \\ (x_j - c_j)^2, & j > 1. \end{cases}$$

Study

Mean dim of

$$\Bigl(\sum_{j=1}^{d} z_j\Bigr)^p$$

equals mean dim in $\boldsymbol{x}$

Transforming to z

$$z_j = \begin{cases} a + (x_1 - c_1)^2, & j = 1 \\ (x_j - c_j)^2, & j > 1. \end{cases}$$

Define

$$\mu_j = \mathbb{E}(z_j) \qquad \sigma_j^2 = \mathrm{Var}(z_j) \qquad \mu_j^{(k)} = \mathbb{E}((z_j - \mu_j)^k)$$

$$\mu_{1:d} = \sum_{j=1}^d \mu_j \qquad \sigma_{1:d}^2 = \sum_{j=1}^d \sigma_j^2 \qquad \mu_{1:d}^{(k)} = \sum_{j=1}^d \mu_j^{(k)}$$

Under conditions (up to 6 moments)

$$\nu(f) \leqslant 1 + \frac{(p-1)^2}{2} \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + O\Bigl(\frac{1}{d^2}\Bigr)$$

Weaker conditions

$$\nu(f) \to 1$$

Proof strategy

$$\nu(f) = \frac{\sum_{j=1}^{d} \bar{\tau}_j^2}{\sigma^2}$$

$$\limsup_{d\to\infty} \frac{\sum_{j=1}^{d} \bar{\tau}_j^2}{p^2\sigma_{1:d}^2/(\mu_{1:d})^2} \leqslant 1$$

$$\liminf_{d\to\infty} \frac{\mathrm{Var}(f(\boldsymbol{z}))}{p^2\sigma_{1:d}^2/(\mu_{1:d})^2} \geqslant 1$$

Ingredients

Taylor expansions

moment bounds

Jansen identity

negative moment bounds

central limit theorem

Gaussian RBF

$$f(\boldsymbol{x}) = \prod_{j=1}^{d} \exp(-(x_j - c_j)^2/\vartheta^2)$$

Mean dim of a product

For $f(\boldsymbol{x}) = \prod_{j=1}^d g_j(x_j)$

$$\nu(f) = \frac{\sum_{j=1}^d \rho_j}{1 - \prod_{j=1}^d (1 - \rho_j)}$$

$$\rho_j = \mathrm{Var}(z_j)/\mathbb{E}(z_j^2) \in [0, 1]$$

O (2003)

Then

$$\frac{\partial}{\partial \rho_k} \nu(f) \geqslant 0$$

Hoyt & O (2023)

So $\cdots$ larger $\rho_j \Rightarrow$ larger ν

Gaussian RBF theorem

Pick θ to get any $0 < \rho_j < 1$

$\Rightarrow$ get any $1 < \nu(f) < d$

Independent x_j with

$$z_j = \frac{(x_j - c_j)^2}{\theta^2} \qquad \mathrm{Var}(x_j) > 0 \qquad \text{PDFs } h_j(x_j) \leqslant M$$

Then

$$f(\boldsymbol{x}) = f(\boldsymbol{z}) = \prod_{j=1}^{d} \exp(-z_j)$$

can attain any $1 < \nu(f) < d$ by choosing θ

Speculation

This may be why Gaussian RBFs dominate other RBFs in machine learning

Ridge in RQMC/QMC

$$\boldsymbol{x} \sim \mathcal{N}(0, I_d) \qquad f_d(\boldsymbol{x}) = g(\theta^\mathsf{T}\boldsymbol{x})$$

Family of test functions with same mean and variance

In any dimension

For Lipschitz $g(\cdot)$

$$\sup_{d \geqslant 1} \nu(f_d) < \infty$$

Get modest limiting mean dimension, e.g., $\leqslant 2$ or 3.

Speculation

This may be why ridge functions dominate Gaussian RBFs in machine learning

Covariance functions

Gaussian process (GP) model for $f(\boldsymbol{x})$ on $\boldsymbol{x} \in \mathbb{R}^d$

$$\mathbb{E}(f(\boldsymbol{x})) = \mu(x) \quad \mathrm{Cov}(f(\boldsymbol{x}), f(\tilde{\boldsymbol{x}})) = \Sigma(\boldsymbol{x}, \tilde{\boldsymbol{x}})$$

For a GP

$$\Sigma(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \phi\bigl(\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|\bigr)$$

ϕ is additive $\Rightarrow$ f is additive

Additive f is a sum of independent GPs

Speculation

This may be why Gaussian covariances dominate multiquadrics in GP models

Keister function

$$\int_{\mathbb{R}^d} e^{-\|\boldsymbol{x}\|^2}\cos(\|\boldsymbol{x}\|)\,\mathrm{d}\boldsymbol{x}$$

Keister (1996), Capstick & Keister (1996)

Resembles problems from physics   Rewrite

$$\mathbb{E}\Bigl(\cos\Bigl(\frac{\|\boldsymbol{x}\|}{2}\Bigr)\Bigr) \qquad \boldsymbol{x}\sim\mathcal{N}(0,I_d)$$

QMC context

This function treats all inputs symmetrically

Yet we still see good QMC integration

Papageorgiou & Traub (1997)

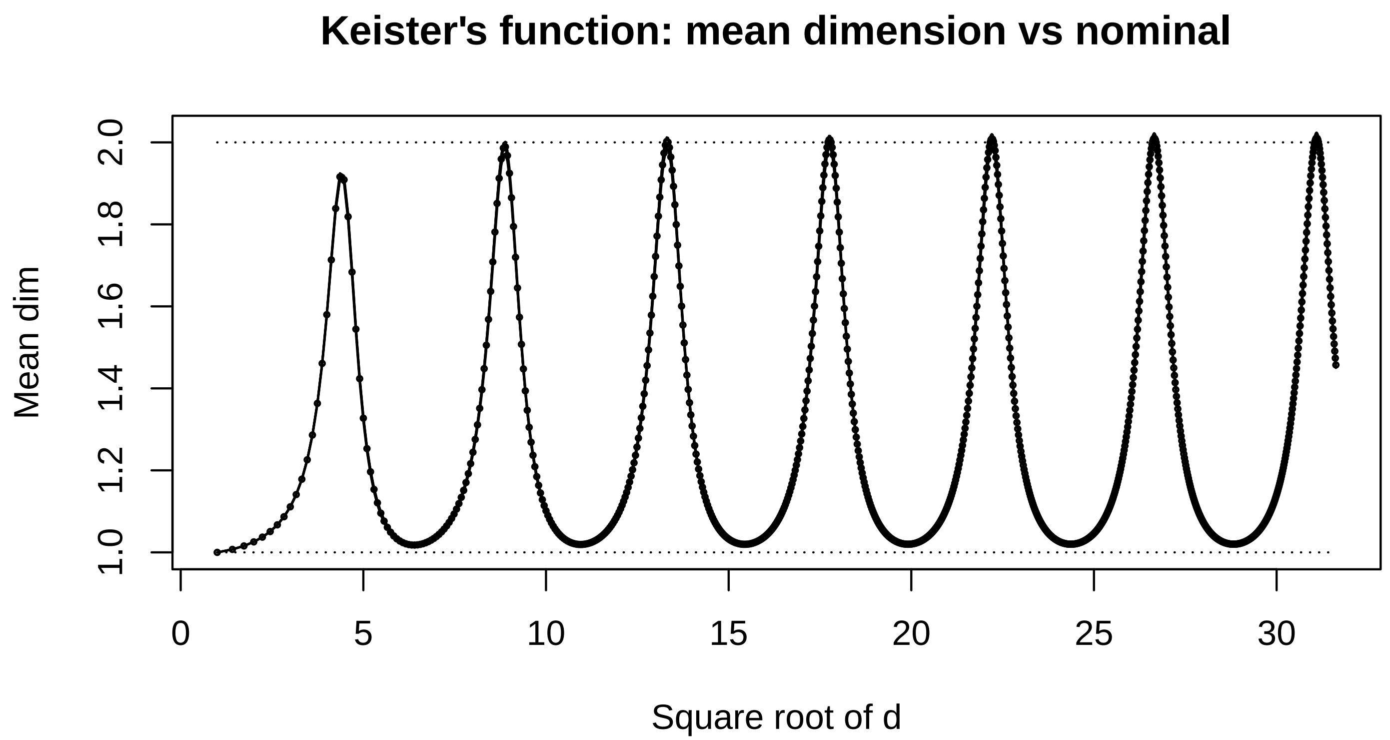
Mean dimension

It has modest mean dimension in any d

and smoothness

$\Rightarrow$ QMC accuracy not surprising

Keister mean dimension



$\nu(f)$ vs $\sqrt{d}$

Five independent QMC estimates overlaid

Based on Jansen identity and 3 χ^2 random variables

Why the oscillation?

$$\|\boldsymbol{x}\|^2 \sim \chi^2_{(d)} \approx \mathcal{N}(d, 2d)$$

By delta method (Taylor's theorem)

$$\frac{\|\boldsymbol{x}\|}{2} \approx \mathcal{N}\Bigl(\frac{\sqrt{d}}{2}, \frac{1}{4}\Bigr)$$

$\approx 99.9\%$ of $\mathcal{N}(\alpha, 1/4)$ is in $\alpha \pm 1.65$

Distn of $\|\boldsymbol{x}\|/2$ uses $\approx$ half period of cosine

Centering

$\frac{\sqrt{d}}{2\pi} \approx$ integer $\Rightarrow$ $\cos(\cdot)$ nearly pure quadratic $\Rightarrow$ $\nu(f) \approx 2$

$\frac{\sqrt{d}}{2\pi} \approx$ integer $+\frac{1}{2}$ $\Rightarrow$ $\cos(\cdot)$ nearly linear $\Rightarrow$ $\nu(f) \approx 1$

Thanks

- Christopher Hoyt, co-author
- Masayoshi Mase and Naofumi Hama, comments
- NSF IIS-1837931 and DMS-2152780
- Hitachi, Ltd.
- Session organizers: Mathias Sonnleitner & Mario Ullrich
- MCM organizers

arXiv:2302.09689