Strong Tractability of Function Approximation Using Radial Basis Function Methods

Fred J. Hickernell

Department of Applied Mathematics
Illinois Institute of Technology
Email: hickernell@iit.edu, Web: www.iit.edu/~hickernell

Joint work with Greg Fasshauer and Henryk Woźniakowski
Supported by grants NSF-DMS-0713848, DE-SC0002100, and NSF-DMS-0914345

Happy Birthday, Stefan!
Thanks to Henryk, Leszek, Peter, and the conference hosts

August 15–19, 2010

Summary

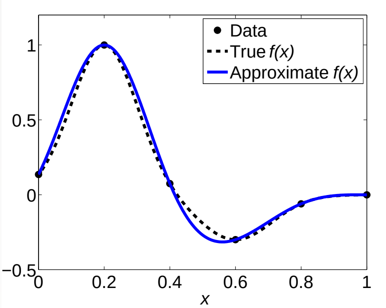
Excellent dimension-independent convergence for function approximation requires the kernel to have

- a high degree of smoothness plus
- a shape parameter that decays quickly with dimension.

Why Approximate Functions

Find $\tilde{f}$ to approximate f given the data:

$$y_i = f(\boldsymbol{x}_i), \quad i = 1, \ldots, n.$$



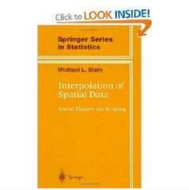
Observational Data — No way to observe $f(\boldsymbol{x})$ at $\boldsymbol{x}$ other than $\{\boldsymbol{x}_i\}_{i=1}^n$

Computer Experiments — Evaluating $f(\boldsymbol{x})$ is expensive

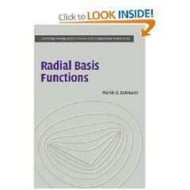
Solving Differential Equations — f is input to a (partial) differential equation solver

Kernel Methods — aka (Smoothing, Regression) Splines, Kriging, Radial Basis Function Methods, etc.

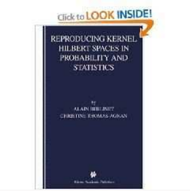
Wahba (1990)



Stein (1999)



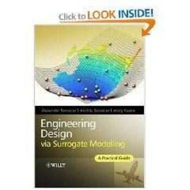
Buhmann (2003)



Berlinet and Thomas-Agnan (2004)

Wendland (2005)

Fasshauer (2007)



Forrester et al. (2008)

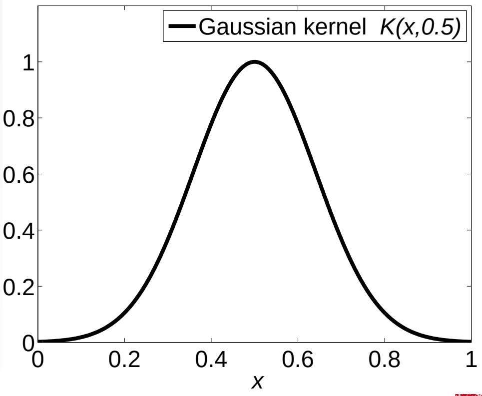
Splines

Given a symmetric & (strictly) positive definite kernel function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, plus data $y_i = f(\boldsymbol{x}_i), i = 1, \ldots, n$, define

$$\tilde{f}(\boldsymbol{x}) = \sum_{j=1}^{n} K(\boldsymbol{x}, \boldsymbol{x}_j) c_j,$$

where the c_j are given by the interpolation conditions

$$y_i = \tilde{f}(\boldsymbol{x}_i) = \sum_{j=1}^{n} K(\boldsymbol{x}_i, \boldsymbol{x}_j) c_j.$$



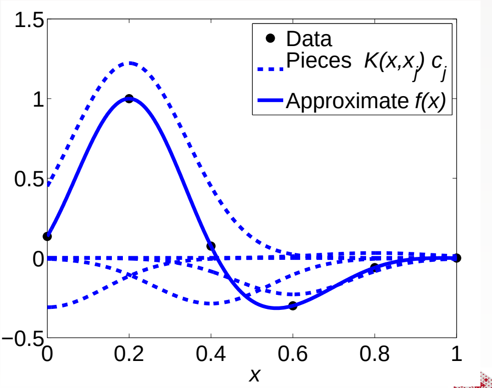
Splines

Given a symmetric & (strictly) positive definite kernel function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, plus data $y_i = f(\boldsymbol{x}_i), i = 1, \ldots, n$, define

$$\tilde{f}(\boldsymbol{x}) = \sum_{j=1}^{n} K(\boldsymbol{x}, \boldsymbol{x}_j) c_j,$$

where the c_j are given by the interpolation conditions

$$y_i = \tilde{f}(\boldsymbol{x}_i) = \sum_{j=1}^{n} K(\boldsymbol{x}_i, \boldsymbol{x}_j) c_j.$$



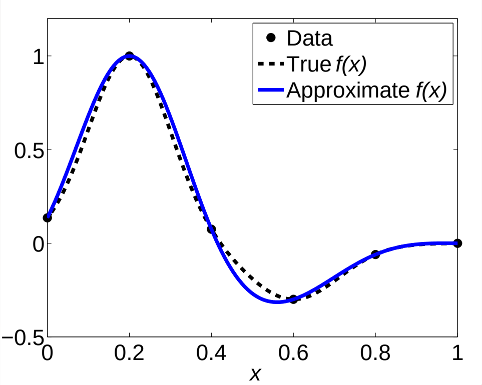
Splines

Given a symmetric & (strictly) positive definite kernel function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, plus data $y_i = f(\boldsymbol{x}_i), i = 1, \ldots, n$, define

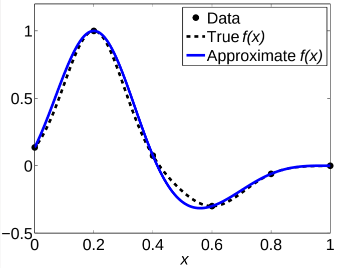
$$\tilde{f}(\boldsymbol{x}) = \sum_{j=1}^{n} K(\boldsymbol{x}, \boldsymbol{x}_j) c_j,$$

where the c_j are given by the interpolation conditions

$$y_i = \tilde{f}(\boldsymbol{x}_i) = \sum_{j=1}^{n} K(\boldsymbol{x}_i, \boldsymbol{x}_j) c_j.$$



Deterministic Perspective



Suppose that f lies in $\mathcal{H}(K)$, the Hilbert space with reproducing kernel K, i.e.,

$$f(\boldsymbol{x}) = \langle K(\cdot, \boldsymbol{x}), f \rangle_{\mathcal{H}(K)} \quad \forall f \in \mathcal{H}(K),\ \boldsymbol{x} \in \mathcal{X}.$$

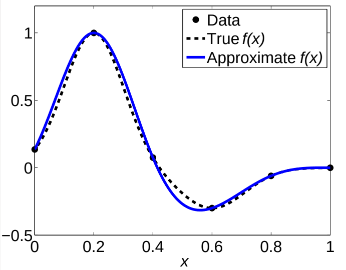
The spline $\tilde{f}$ is optimal in the face of a clever adversary:

$$\mathrm{e}(n; \{\boldsymbol{x}_i\}_{i=1}^n) = \mathrm{error}(\tilde{f}(\boldsymbol{x})) = \min_{g \in \mathbb{R}} \mathrm{error}(\underbrace{g}_{\text{guess}}),$$

where the true function fits the data observed:

$$\mathrm{error}(\underbrace{g}_{\text{guess}}) := \sup \left\{ |f(\boldsymbol{x}) - \underbrace{g}_{\text{guess}}| : \|f\|_{\mathcal{H}(K)} \le \alpha^2,\ y_i = f(\boldsymbol{x}_i)\ \forall i \right\}.$$

Deterministic Perspective



Suppose that f lies in $\mathcal{H}(K)$, the Hilbert space with reproducing kernel K, i.e.,

$$f(\boldsymbol{x}) = \langle K(\cdot, \boldsymbol{x}), f \rangle_{\mathcal{H}(K)} \qquad \forall f \in \mathcal{H}(K),\ \boldsymbol{x} \in \mathcal{X}.$$

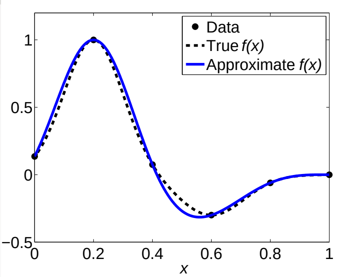
The spline $\tilde{f}$ is optimal in the face of a clever adversary:

$$\mathrm{e}(n; \{\boldsymbol{x}_i\}_{i=1}^n) = \mathrm{error}(\tilde{f}) = \min_{g \in \mathcal{L}_p} \mathrm{error}(\underbrace{g}_{\text{guess}}),$$

where the true function fits the data observed:

$$\mathrm{error}(\underbrace{g}_{\text{guess}}) := \sup \left\{ \|f - \underbrace{g}_{\text{guess}}\|_p : \|f\|_{\mathcal{H}(K)} \le \alpha^2,\ y_i = f(\boldsymbol{x}_i)\ \forall i \right\}.$$

Deterministic Perspective



Suppose that f lies in $\mathcal{H}(K)$, the Hilbert space with reproducing kernel K, i.e.,

$$f(\boldsymbol{x}) = \langle K(\cdot, \boldsymbol{x}), f \rangle_{\mathcal{H}(K)} \quad \forall f \in \mathcal{H}(K),\ \boldsymbol{x} \in \mathcal{X}.$$

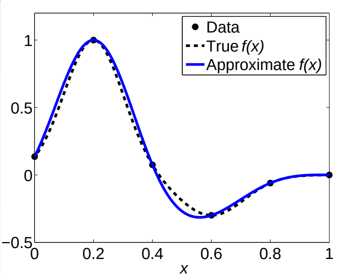
The spline $\tilde{f}$ is optimal in the face of a clever adversary:

$$\mathrm{e}(n; \{L_i\}_{i=1}^n) = \mathrm{error}(\tilde{f}) = \min_{g \in \mathcal{L}_p} \mathrm{error}(\underbrace{g}_{\text{guess}}),$$

where the true function fits the data observed:

$$\mathrm{error}(\underbrace{g}_{\text{guess}}) := \sup \left\{ \|f - \underbrace{g}_{\text{guess}}\|_p : \|f\|_{\mathcal{H}(K)} \le \alpha^2,\ y_i = L_i(f)\ \forall i \right\}.$$

Stochastic Perspective



Suppose that f is a zero mean Gaussian stochastic process with covariance kernel K, i.e.,

$$E[f(\boldsymbol{x})f(\boldsymbol{t})] = K(\boldsymbol{x},\boldsymbol{t}), \quad \boldsymbol{x},\boldsymbol{t} \in \mathcal{X}.$$

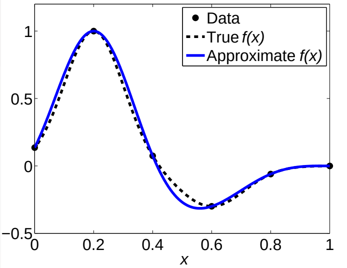
It follows that the spline $\tilde{f}$ is optimal in the face of a crazy adversary:

$$\mathrm{e}(n;\{\boldsymbol{x}_i\}_{i=1}^n) = \mathrm{error}(\tilde{f}(\boldsymbol{x})) = \min_{g\in\mathbb{R}} \mathrm{error}(\underbrace{g}_{\text{guess}}),$$

where the true function fits the data observed:

$$\mathrm{error}(\underbrace{g}_{\text{guess}}) := \left\{E\left[|f(\boldsymbol{x}) - \underbrace{g}_{\text{guess}}|^q \;\middle|\; y_i = f(\boldsymbol{x}_i)\ \forall i\right]\right\}^{1/q}$$

Stochastic Perspective



Suppose that f is a zero mean Gaussian stochastic process with covariance kernel K, i.e.,

$$E[f(\boldsymbol{x})f(\boldsymbol{t})] = K(\boldsymbol{x},\boldsymbol{t}), \quad \boldsymbol{x},\boldsymbol{t} \in \mathcal{X}.$$

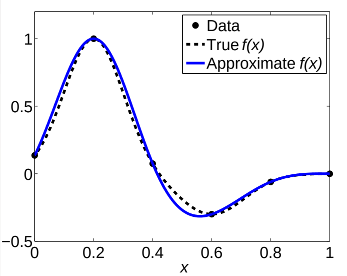
It follows that the spline $\tilde{f}$ is optimal in the face of a crazy adversary:

$$\mathrm{e}(n;\{\boldsymbol{x}_i\}_{i=1}^n) = \mathrm{error}(\tilde{f}) = \min_{g\in\mathcal{L}_2} \mathrm{error}(\underbrace{g}_{\text{guess}}),$$

where the true function fits the data observed:

$$\mathrm{error}(\underbrace{g}_{\text{guess}}) := \left\{E\left[\|f - \underbrace{g}_{\text{guess}}\|_p^q \;\middle|\; y_i = f(\boldsymbol{x}_i)\ \forall i\right]\right\}^{1/q}$$

Stochastic Perspective



Suppose that f is a zero mean Gaussian stochastic process with covariance kernel K, i.e.,

$$E[f(\boldsymbol{x})f(\boldsymbol{t})] = K(\boldsymbol{x},\boldsymbol{t}), \quad \boldsymbol{x},\boldsymbol{t} \in \mathcal{X}.$$

It follows that the spline $\tilde{f}$ is optimal in the face of a crazy adversary:

$$\mathrm{e}(n;\{L_i\}_{i=1}^n) = \mathrm{error}(\tilde{f}) = \min_{g\in\mathcal{L}_2} \mathrm{error}(\underbrace{g}_{\text{guess}}),$$

where the true function fits the data observed:

$$\mathrm{error}(\underbrace{g}_{\text{guess}}) := \left\{E\left[\|f - \underbrace{g}_{\text{guess}}\|_p^q \;\middle|\; y_i = L_i(f)\ \forall i\right]\right\}^{1/q}$$

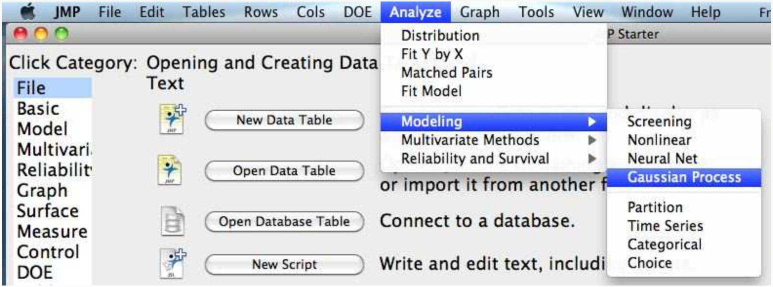
Known Convergence & Tractability Results

Conditions on f, $\{\boldsymbol{x}_i\}_{i=1}^n$	$\mathrm{e}(n, \{\boldsymbol{x}_i\}_{i=1}^n)$, $\mathcal{L}_\infty$ worst case
f has deriv. up to total order α, covering radius of $\{\boldsymbol{x}_i\}_{i=1}^n$ is $\preceq C_1(d) n^{-1/d}$ (Wendland, 2005)	$\preceq C_2(d) n^{-\alpha/d}$, curse of dimensionality?
f periodic, mixed partial deriv. up to order α in each variable, $K(\boldsymbol{x},\boldsymbol{t}) = \prod_{\ell=1}^d [1 + \gamma_\ell^2 K_1(x_\ell, t_\ell)]$, γ_ℓ indicates the importance of variable ℓ, lattice designs (Zeng et al., 2006; Kuo et al., 2006, 2008; Zeng et al., 2009), similar results for digital nets	$\preceq n^{-\frac{r(\boldsymbol{\gamma})-1/2}{2-1/(2r(\boldsymbol{\gamma}))}}$ for $r(\boldsymbol{\gamma}) > 1/2$ $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots)$ $r(\boldsymbol{\gamma}) = \sup\{\beta \le \alpha \mid \sum_{\ell=1}^\infty \gamma_\ell^{1/\beta} < \infty\}$ dimension independent, but suboptimal convergent rate
As above, but for unknown designs (Kuo et al., 2009)	$\preceq n^{-\frac{r(\boldsymbol{\gamma})-1/2}{1+1/(2r(\boldsymbol{\gamma}))}}$ for $r(\boldsymbol{\gamma}) > 1/2$

$a(n,d) \preceq n^{-p}$ means $a(n,d) \le C n^{-p+\epsilon}$ for all $\epsilon > 0$

Gaussian Kernel with a Nonhomogeneous Shape Parameter

The JMP statistical software Gaussian process modeling module:



uses a Gaussian covariance kernel:

$$K(\boldsymbol{x}, \boldsymbol{t}) = e^{-\gamma_1^2(x_1-t_1)^2 - \cdots - \gamma_d^2(x_d-t_d)^2},$$

where the shape parameter γ_j reflects the importance of the variable x_j.
How accurate is the fit?

Summary of Tractability for Gaussian Kernel (Fasshauer et al., 2010)

$$K(\boldsymbol{x},\boldsymbol{t}) = e^{-\gamma_1^2(x_1-t_1)^2-\cdots-\gamma_d^2(x_d-t_d)^2}, \quad \boldsymbol{x},\boldsymbol{t} \in \mathbb{R}^d$$

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots), \qquad r(\boldsymbol{\gamma}) = \sup\left\{\beta > 0 \,\middle|\, \sum_{\ell=1}^{\infty} \gamma_\ell^{1/\beta} < \infty\right\}$$

Design	Absolute, $\mathrm{e}(n)$	Normalized, $\frac{\mathrm{e}(n)}{\mathrm{e}(0)}$
	Worst Case $\mathcal{L}_2$ Error	
Best $\{\boldsymbol{x}_i\}_{i=1}^n$	$\preceq n^{-\max\left(\frac{r(\boldsymbol{\gamma})}{1+1/(2r(\boldsymbol{\gamma}))},\frac{1}{4}\right)}$	$\preceq n^{-\frac{r(\boldsymbol{\gamma})}{1+1/(2r(\boldsymbol{\gamma}))}}$ if $r(\boldsymbol{\gamma}) > 1/2$
Best $\{L_i\}_{i=1}^n$	$\asymp n^{-\max(r(\boldsymbol{\gamma}),1/2)}$	$\asymp n^{-r(\boldsymbol{\gamma})}$ if $r(\boldsymbol{\gamma}) > 0$
	Average Case $\mathcal{L}_2$ Error	
Best $\{L_i\}_{i=1}^n$	$\asymp n^{-r(\boldsymbol{\gamma})+1/2}$ if $r(\boldsymbol{\gamma}) > 1/2$	

Summary of Tractability for Gaussian Kernel (Fasshauer et al., 2010)

$$K(\boldsymbol{x},\boldsymbol{t}) = e^{-\gamma_1^2(x_1-t_1)^2-\cdots-\gamma_d^2(x_d-t_d)^2}, \quad \boldsymbol{x},\boldsymbol{t} \in \mathbb{R}^d$$

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots), \qquad \text{e.g., } \gamma_\ell = \ell^{-r(\boldsymbol{\gamma})}$$

Design	Absolute, $\mathrm{e}(n)$	Normalized, $\frac{\mathrm{e}(n)}{\mathrm{e}(0)}$
	Worst Case $\mathcal{L}_2$ Error	
Best $\{\boldsymbol{x}_i\}_{i=1}^n$	$\preceq n^{-\max\left(\frac{r(\boldsymbol{\gamma})}{1+1/(2r(\boldsymbol{\gamma}))},\frac{1}{4}\right)}$	$\preceq n^{-\frac{r(\boldsymbol{\gamma})}{1+1/(2r(\boldsymbol{\gamma}))}}$ if $r(\boldsymbol{\gamma}) > 1/2$
Best $\{L_i\}_{i=1}^n$	$\asymp n^{-\max(r(\boldsymbol{\gamma}),1/2)}$	$\asymp n^{-r(\boldsymbol{\gamma})}$ if $r(\boldsymbol{\gamma}) > 0$
	Average Case $\mathcal{L}_2$ Error	
Best $\{L_i\}_{i=1}^n$	$\asymp n^{-r(\boldsymbol{\gamma})+1/2}$ if $r(\boldsymbol{\gamma}) > 1/2$	

The Best Design Using Linear Functional Data

Decompose the reproducing/covariance kernel into its eigenfunction expansion:

$$K(\boldsymbol{x},\boldsymbol{t}) = \sum_{\ell=1}^{\infty} \lambda_{(\ell)} \varphi_{(\ell)}(\boldsymbol{x}) \varphi_{(\ell)}(\boldsymbol{t}), \quad \int_{\mathcal{X}} K(\boldsymbol{x},\boldsymbol{t}) \varphi_{(\ell)}(\boldsymbol{t})\, \rho(\boldsymbol{t}) \mathrm{d}\boldsymbol{t} = \lambda_{(\ell)} \varphi_{(\ell)}(\boldsymbol{x}).$$

where $\lambda_{(\ell)}$ denotes the ℓ^{th} largest eigenvalue. The design yielding the smallest error is $\left\{ \langle \varphi_{(i)}, \cdot \rangle_{\mathcal{H}(K)} \right\}_{i=1}^{n}$.

$\mathcal{L}_{2,\rho}$ error:

$$\mathrm{e}(n) = \begin{cases} \sqrt{\lambda_{(n+1)}}, & \text{worst case} \\ \sqrt{\sum_{\ell=n+1}^{\infty} \lambda_{(\ell)}}, & \text{average case} \end{cases}$$

Gaussian Kernel Eigenvalues Known

For the Gaussian kernel

$$K(\boldsymbol{x},\boldsymbol{t}) = e^{-\gamma_1^2(x_1-t_1)^2-\cdots-\gamma_d^2(x_d-t_d)^2} = \sum_{\boldsymbol{j}\in\mathbb{N}^d} \tilde{\lambda}_{d,\boldsymbol{\gamma},\boldsymbol{j}}\varphi_{d,\boldsymbol{\gamma},\boldsymbol{j}}(\boldsymbol{x})\varphi_{d,\boldsymbol{\gamma},\boldsymbol{j}}(\boldsymbol{t}),$$

$$\rho(\boldsymbol{t}) = \frac{1}{\pi^{d/2}}e^{-\|\boldsymbol{t}\|_2^2},$$

$$\tilde{\lambda}_{d,\boldsymbol{\gamma},\boldsymbol{j}} = \prod_{\ell=1}^{d}(1-\omega_{\gamma_\ell})\omega_{\gamma_\ell}^{j_\ell-1}, \qquad \omega_\gamma = \frac{\gamma^2}{\frac{1}{2}(1+\sqrt{1+4\gamma^2})+\gamma^2}.$$

Moreover, the $n+1^{\text{st}}$ largest eigenvalue may be bounded in terms of the sums of powers of eigenvalues:

$$\lambda_{(n+1)} \le \left(\frac{1}{n+1}\sum_{\ell=1}^{n+1}\lambda_{(\ell)}^{\tau}\right)^{1/\tau} \le \frac{1}{(n+1)^{1/\tau}} \underbrace{\prod_{\ell=1}^{d}\frac{1-\omega_{\gamma_\ell}}{(1-\omega_{\gamma_\ell}^{\tau})^{1/\tau}}}_{\text{Uniformly bounded in } d?} .$$

Summary of Tractability for Gaussian Kernel (Fasshauer et al., 2010)

$$K(\boldsymbol{x},\boldsymbol{t}) = e^{-\gamma_1^2(x_1-t_1)^2-\cdots-\gamma_d^2(x_d-t_d)^2}, \quad \boldsymbol{x},\boldsymbol{t} \in \mathbb{R}^d$$

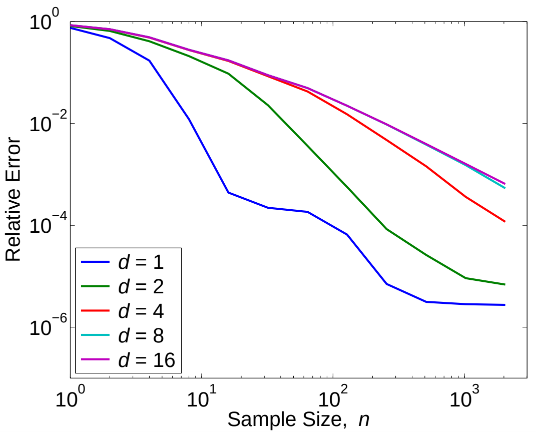
$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots), \qquad \text{e.g., } \gamma_\ell = \ell^{-r(\boldsymbol{\gamma})}$$

Design	Absolute, $\mathrm{e}(n)$	Normalized, $\frac{\mathrm{e}(n)}{\mathrm{e}(0)}$
	Worst Case $\mathcal{L}_2$ Error	
Best $\{\boldsymbol{x}_i\}_{i=1}^n$	$\preceq n^{-\max\left(\frac{r(\boldsymbol{\gamma})}{1+1/(2r(\boldsymbol{\gamma}))},\frac{1}{4}\right)}$	$\preceq n^{-\frac{r(\boldsymbol{\gamma})}{1+1/(2r(\boldsymbol{\gamma}))}}$ if $r(\boldsymbol{\gamma}) > 1/2$
Best $\{L_i\}_{i=1}^n$	$\asymp n^{-\max(r(\boldsymbol{\gamma}),1/2)}$	$\asymp n^{-r(\boldsymbol{\gamma})}$ if $r(\boldsymbol{\gamma}) > 0$
	Average Case $\mathcal{L}_2$ Error	
Best $\{L_i\}_{i=1}^n$	$\asymp n^{-r(\boldsymbol{\gamma})+1/2}$ if $r(\boldsymbol{\gamma}) > 1/2$	

Numerical Experiments

$$f(\boldsymbol{x}) = \sum_{k=1}^{2} \frac{A_k}{1 + B_{k1}^2\gamma_1^2(x_1 - Z_{k1})^2 + \cdots + B_{kd}^2\gamma_d^2(x_d - Z_{kd})^2}$$

with i.i.d.
$A_1, \ldots, A_d \sim U(-1,1)$,
$B_{11}, \ldots, B_{2d} \sim U(0,1)$,
$Z_{11}, \ldots, Z_{2d} \sim U(0,1)$,
and $\gamma_\ell = 2^{3-\ell}$.



Outstanding Issues

- Average case $\mathcal{L}_2$ tractability for function values
- Tractability of $\mathcal{L}_\infty$ approximation
- Ill-conditioning of Gram matrix K used to obtain the spline coefficients
- General radial kernels, $K(\boldsymbol{x}, \boldsymbol{t}) = \kappa(\gamma_1^2(x_1 - t_1)^2 + \cdots + \gamma_d^2(x_d - t_d)^2)$
- Geometric design criterion guaranteeing optimal exponent of strong tractability
- A good design assuming that $r(\boldsymbol{\gamma})$ is large, but not knowing the particular γ_ℓ
- Noisy data

References I

Berlinet, A., Thomas-Agnan, C., 2004. Reproducing Kernel Hilbert Spaces in Probability and Statistics. Kluwer Academic Publishers, Boston.

Buhmann, M. D., 2003. Radial Basis Functions. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge.

Fasshauer, G. E., 2007. Meshfree Approximation Methods with MATLAB. Vol. 6 of Interdisciplinary Mathematical Sciences. World Scientific Publishing Co., Singapore.

Fasshauer, G. E., Hickernell, F. J., Woźniakowski, H., 2010. Rate of convergence and tractability of the radial function approximation problem. Submitted for publication.

Forrester, A. I. J., Sóbester, A., Keane, A. J., 2008. Engineering Design via Surrogate Modelling. Wiley, Chichester.

References II

Kuo, F. Y., Sloan, I. H., Woźniakowski, H., 2006. Lattice rules for multivariate approximation in the worst case setting. In: Niederreiter and Talay (2006), pp. 289–330.

Kuo, F. Y., Sloan, I. H., Woźniakowski, H., 2008. Lattice rule algorithms for multivariate approximation in the average case setting. J. Complexity 24, 283–323.

Kuo, F. Y., Wasilkowski, G. W., Woźniakowski, H., 2009. On the power of standard information for multivariate approximation in the worst case setting. J. Approx. Theory 158, 97–125.

Niederreiter, H., Talay, D. (Eds.), 2006. Monte Carlo and Quasi-Monte Carlo Methods 2004. Springer-Verlag, Berlin.

Stein, M. L., 1999. Interpolation of Spatial Data: Some theory for Kriging. Springer-Verlag, New York.

References III

Wahba, G., 1990. Spline Models for Observational Data. Vol. 59 of CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.

Wendland, H., 2005. Scattered Data Approximation. No. 17 in Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge.

Zeng, X. Y., Kritzer, P., Hickernell, F. J., 2009. Spline methods using integration lattices and digital nets. Constr. Approx. 30, 529–555.

Zeng, X. Y., Leung, K. T., Hickernell, F. J., 2006. Error analysis of splines for periodic problems using lattice designs. In: Niederreiter and Talay (2006), pp. 501–514.