

ON VARIABLE AND RANDOM SHAPE GAUSSIAN INTERPOLATIONS

SUNG NOK CHIU*, LEEVAN LING*, AND MICHAEL MCCOURT†

Abstract. This work focuses on the invertibility of non-constant shape Gaussian asymmetric interpolation matrix, which includes the cases of both variable and random shape parameters. We prove a sufficient condition for that these interpolation matrices are invertible almost surely for the choice of shape parameters. The proof is then extended to the case of anisotropic Gaussian kernels, which is subjected to independent componentwise scalings and rotations. As a corollary of our proof, we propose a parameter free random shape parameters strategy to completely eliminate the need of users' inputs. By studying numerical accuracy in variable precision computations, we demonstrate that the asymmetric interpolation method is not a method with faster theoretical convergence. We show empirically in double precision, however, that these spatially varying strategies have the ability to outperform constant shape parameters in double precision computations. Various random distributions were numerically examined.

1. Interpolation with spatially varying kernels. We consider the standard interpolation problem in $\Omega \subset \mathbb{R}^d$: given data $\{z_k, f(z_k)\}_{k=1}^n =: \{Z, \mathbf{f}_Z\} \subset \Omega \times \mathbb{R}$ for some smooth function f , we seek an interpolant in the form of

$$s_{f,Z} = \sum_{k=1}^n \alpha_k J_k, \quad (1.1)$$

satisfying $s_{f,Z}(z_k) = f(z_k)$ for $k = 1, \dots, n$, where the basis functions $J_k : \Omega \rightarrow \mathbb{R}$, for $k = 1, \dots, n$, are defined by some Gaussian kernel function $J : \Omega \times \Omega \rightarrow \mathbb{R}$ centered at $z_k \in Z$.

We consider the case that each basis function J_k is defined *differently* by its own strictly positive shape parameter, $\epsilon_k \in \mathbb{R}^+$, so that

$$J_k(\cdot) := J(\cdot, z_k; \epsilon_k) = \exp(-\epsilon_k^2 \|\cdot - z_k\|^2). \quad (1.2)$$

All n such shape parameters can be stored in a vector $\boldsymbol{\epsilon} := \{\epsilon_i\}_{i=1}^n \in (\mathbb{R}^+)^n$ which can ease notation below. This non-constant shape parameters framework is commonly referred to as interpolation with *variable shape kernels* and with *random shape kernels*, respectively, when $\epsilon_k = \epsilon(z_k)$ for some smooth function ϵ and when ϵ_k is a positive random variable following some probability distribution. Despite the lack of theory, there are numerous instances of numerical evidence in the literature suggesting such an approach can improve accuracy and numerical stability; see for instance [3, 6, 12,

*Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

†SigOpt, Inc.

16, 19, 22] for some related work in the past two decades.

EXAMPLE 1.1. *Let $Z = \{0, 1e-10, 1\} \subset \mathbb{R}$ and J be the standard Gaussian kernel. The condition numbers of the traditional interpolation matrices are $4.7e+21$, $3.2e+20$ and $5.0e+19$ for constant shape kernels with shape parameter $\epsilon = 0.5, 1$, and 2 . Using non-constant shape parameters $\boldsymbol{\varepsilon} = (0.5, 2, 1)$ to avoid nearly identical basis, the condition number drops significantly to $8.2e+10$. All condition numbers in this example were computed with 64 digits accuracy. \square*

The only theory available [1] thus far is for a “similar” formulation that handles variable shape parameters by an extra dimension. In their approach, numerical expansions are *not* using basis functions in the form of (1.2). The Gaussian-related basis function there takes the form $\exp(\|z_i - z_k\|^2) \exp(-(\epsilon_i - \epsilon_j)^2)$, for example.

Implementation of asymmetric interpolation with (1.1)–(1.2) is straightforward. One can identify the unknown coefficients in (1.1) by solving the exactly determined linear system

$$\begin{pmatrix} J(z_1, z_1; \epsilon_1) & \cdots & J(z_1, z_n; \epsilon_n) \\ \vdots & \ddots & \vdots \\ J(z_n, z_1; \epsilon_1) & \cdots & J(z_n, z_n; \epsilon_n) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} f(z_1) \\ \vdots \\ f(z_n) \end{pmatrix}, \quad (1.3)$$

or denoted by $\mathbf{J}(Z, Z; \boldsymbol{\varepsilon})\boldsymbol{\alpha} = \mathbf{f}_Z$ in compact matrix notations with $\boldsymbol{\varepsilon} = [\epsilon_1, \dots, \epsilon_n]^T \in (\mathbb{R}^+)^n$. Due to the different scales used in defining these basis functions, this $n \times n$ non-constant shape kernel interpolation matrix $\mathbf{J}(Z, Z; \boldsymbol{\varepsilon})$ is asymmetric in general and the conditions for its invertibility will be addressed in the following section. Whenever (1.3) has a solution, we can express the interpolant in (1.1) in vector form as

$$s_{f,Z}(x) = \mathbf{J}(x, Z; \boldsymbol{\varepsilon})\boldsymbol{\alpha}$$

by using a $1 \times n$ row-vector function

$$\mathbf{J}(x, Z; \boldsymbol{\varepsilon}) := [J(x, z_1; \epsilon_1), \dots, J(x, z_n; \epsilon_n)], \quad x \in \Omega.$$

2. Solvability almost surely. We begin with a small comment on the potential lack of invertibility of asymmetric interpolation matrix $\mathbf{J}(Z, Z; \boldsymbol{\varepsilon})$ in (1.3).

EXAMPLE 2.1. *Let $Z = \{1, 2, 3\} \subset \mathbb{R}$ and $\boldsymbol{\varepsilon}(t) = [t, 0.3, 0.4]^T$ for $t \in \mathbb{R}^+$ as the first shape parameter ϵ_1 . We show the determinant of $\mathbf{J}(Z, Z; \boldsymbol{\varepsilon}(t))$ as a function of $0 < t \leq 2.5$ in the left graph of Figure 2.1. For $t > 2.5$, the determinant value*

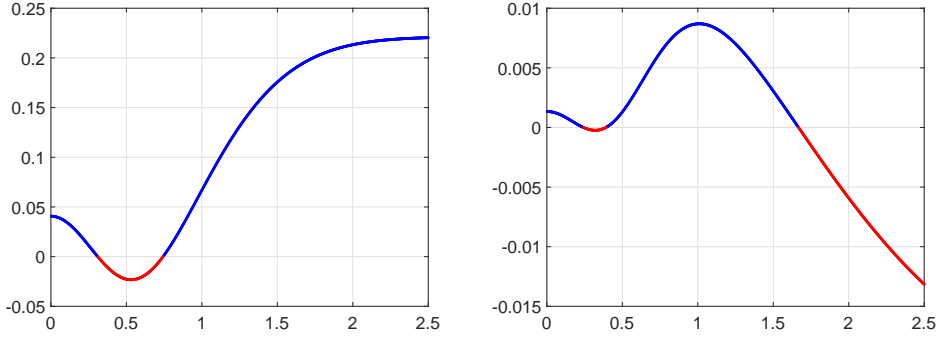


FIG. 2.1. Determinants of $n \times n$ spatially varying kernel interpolation matrices $J(Z, Z; \epsilon)$ as a function of shape parameter ϵ_1 . left: $n = 3$. right: $n = 4$.

monotonically increases and approaches a horizontal asymptote from below. There are two “bad” values for the shape parameter t that make the determinant zero and the asymmetric interpolation matrix singular. Similarly, we construct a 4×4 examples by $Z = \{0.5, 1, 2, 3\} \subset \mathbb{R}$ and $\epsilon(t) = [t, 0.5, 0.3, 0.4]^T$ for $t \in \mathbb{R}^+$ and show the resulting determinant in right graph of Figure 2.1. This time, we see three roots in the determinant function. \square

The data points and shape parameters used in Example 1.1 are arbitrary. It is easy to construct examples so that the determinant of $\det J(Z, Z; \epsilon(t))$ is always nonzero for all $t \in \mathbb{R}^+$. The following theorem inserts a condition for invertibility without any restriction on how $\epsilon \in (\mathbb{R}^+)^n$ behaves.

THEOREM 2.2. *For any set of data points $Z \subset \Omega$ with distinct pairwise distances such that, for any $1 \leq i, j, k \leq n$, $\|z_i - z_k\| \neq \|z_j - z_k\|$ if $i \neq j$, the asymmetric Gaussian interpolation matrix $J(Z, Z; \epsilon)$ in (1.3) is invertible for ϵ in almost everywhere of $(\mathbb{R}^+)^n$ with respect to the Lebesgue measure.*

PROOF: The idea of the proof is to sequentially rank-1 perturb the columns of an invertible matrix to generate a sequence of invertible matrices which end with the desired asymmetric interpolation matrix. We begin with a symmetric interpolation matrix of $A_0 := J(Z, Z; \mathbf{1})$, which must be invertible by the fact that kernel J is symmetric positive definite. The first step towards arriving at $J(Z, Z; \epsilon)$ in (1.3) is to perturb the first column to associate it with the shape parameter $[\epsilon]_1$. We do so by defining $A_1 := A_0 + \mathbf{p}_1 \mathbf{e}_1^T$ with $\mathbf{p}_1 \in \mathbb{R}^n$, whose entries are given by

$$[\mathbf{p}_1]_i := J(z_i, z_1; \epsilon_1) - J(z_i, z_1; 1), \quad 1 \leq i \leq n, \quad (2.1)$$

so that the first column of \mathbf{A}_0 matches with that of $\mathbf{J}(Z, Z; \boldsymbol{\varepsilon})$. Now we examine the invertibility of \mathbf{A}_1 by its determinant. By the equality

$$\det(\mathbf{A}_1) = \det(\mathbf{A}_0 + \mathbf{p}_1 \mathbf{e}_1^T) = (1 + \mathbf{e}_1^T \mathbf{A}_0^{-1} \mathbf{p}_1) \det(\mathbf{A}_0),$$

we know $\det(\mathbf{A}_1) = 0$ if and only if

$$\mathbf{p}_1 \in \{\mathbf{p} \in \mathbb{R}^n : (\mathbf{e}_1^T \mathbf{A}_0^{-1}) \mathbf{p} = -1\} =: \mathcal{S}. \quad (2.2)$$

But \mathbf{p}_1 is not arbitrary; from (2.1), we can see that the set of all \mathbf{p}_1 for $\epsilon_1 > 0$ is a space curve in \mathbb{R}^n , which we denote as $\gamma := \{\mathbf{p}_1(\epsilon_1) : \epsilon_1 > 0\}$. Together, we know \mathbf{A}_1 is invertible as long as $\mathbf{p}_1(\epsilon_1) \notin \gamma \cap \mathcal{S}$.

We already know that the whole space curve is confined in the hyperplane $x_1 = 0$ since $[\mathbf{p}_1]_1 = 0$ for all $\mathbf{p}_1 \in \gamma$. The inhomogeneous equation in (2.2) ensures that \mathcal{S} cannot be the entire $x_1 = 0$ hyperplane, or otherwise the equation has a trivial solution. To complete the proof, it remains to show that γ is not locally confined in any \mathbb{R}^{n-2} dimensional hyperplane by showing that all but one of its curvatures are nonzero. Together, we know the set of intersection $\gamma \cap \mathcal{S}$ is measure zero and, hence, \mathbf{A}_1 is invertible for $[\boldsymbol{\varepsilon}]_1$ almost everywhere in \mathbb{R}^+ .

It is more convenient to work with $t = ([\boldsymbol{\varepsilon}]_1)^2 > 0$; we re-parameterize the space curve γ as

$$\gamma(t) := \left\{ \mathbf{p} : [\mathbf{p}]_i = J(z_i, z_1; \sqrt{t}) - J(z_i, z_1; 1), \quad 1 \leq i \leq n \right\} \subset \mathbb{R}^n. \quad (2.3)$$

Let $s = s_\gamma(t)$, $(\mathbf{T}_1(t), \dots, \mathbf{T}_n(t))$, and $(\kappa_1(t), \dots, \kappa_{n-1}(t))$ denote the arc-length, the generalized TNB frame, and generalized curvatures of $\gamma(t) \subset \mathbb{R}^n$ respectively. By applying chain rule and Frenet-Serret formulas repeatedly, the derivative vectors of $\gamma(t)$ can be expressed as

$$\mathbf{D}_\gamma(t) := [\gamma'(t), \gamma''(t), \dots, \gamma^{(n)}(t)] = [\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_n] \mathbf{U}(t), \quad (2.4)$$

where \mathbf{U} is an $n \times n$ upper triangular matrix with

$$\text{diag}(\mathbf{U}(t)) = \left\| \frac{d\gamma}{ds} \right\| \left[\frac{ds}{dt}, \left(\frac{ds}{dt} \right)^2 \kappa_1, \left(\frac{ds}{dt} \right)^3 \kappa_1 \kappa_2, \dots, \left(\frac{ds}{dt} \right)^n \prod_{j=1}^{n-1} \kappa_j \right].$$

Nonzero off-diagonal entries involve higher derivatives of $s(t)$ and derivatives of generalized curvatures.

Next, we compute entries of the derivative matrix explicitly for the Gaussian

kernel J and get

$$[\mathbf{D}_\gamma(t)]_{ij} = (-1)^j \|z_i - z_1\|^{2j} e^{-t\|z_i - z_1\|^2},$$

which is the product of a full-rank diagonal matrix function of t with exponential components and the transpose of a Vandermonde matrix generated by the set of scalars

$$\{0, \|z_2 - z_1\|^2, \dots, \|z_n - z_1\|^2\}.$$

This t -independent Vandermonde matrix is of rank $n - 1$ if $\|z_i - z_1\|$ are all distinct and non-zero. Combined with (2.4), they imply that the rank of $\mathbf{U}(t)$ is also $n - 1$ and consequently that $\kappa_j \neq 0$, $1 \leq j \leq n - 2$, and $\kappa_{n-1} = 0$ for all $t > 0$. Thus, the number of $\epsilon_1 > 0$ that makes \mathbf{A}_1 singular is finite.

After applying the same argument iteratively to perturb the second through n th columns, we show that the asymmetric Gaussian interpolation matrix $\mathbf{J}(Z, Z; \boldsymbol{\epsilon})$ is invertible for $\boldsymbol{\epsilon}$ almost everywhere in $(\mathbb{R}^+)^n$ if all pairwise distances of Z are distinct. \square

3. Anisotropic Gaussian kernels. We consider the case that each basis function J_k is *componentwise* scaled differently by a matrix $\boldsymbol{\mathcal{E}} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n] \in (\mathbb{R}^+)^{d \times n}$ formed by a set of n strictly positive shape parameter vectors $\{\boldsymbol{\epsilon}_i\}_{i=1}^n \subset (\mathbb{R}^+)^d$ so that the basis function centered at z_k

$$J(\cdot, z_k; \boldsymbol{\epsilon}_k) = \exp\left(-(\cdot - z_k)^T \text{diag}(\boldsymbol{\epsilon}_k)^2 (\cdot - z_k)\right)$$

is no longer radially symmetric. The corresponding anisotropic Gaussian interpolation matrix on $Z \subset \Omega$ now takes the form of

$$\mathbf{J}(Z, Z; \boldsymbol{\mathcal{E}}) := \begin{pmatrix} J(z_1, z_1; \boldsymbol{\epsilon}_1) & \cdots & J(z_1, z_n; \boldsymbol{\epsilon}_n) \\ \vdots & \ddots & \vdots \\ J(z_n, z_1; \boldsymbol{\epsilon}_1) & \cdots & J(z_n, z_n; \boldsymbol{\epsilon}_n) \end{pmatrix}. \quad (3.1)$$

The proof of Theorem 2.2 can be modified to this anisotropic setting.

THEOREM 3.1. *For any set of data points $Z \subset \Omega$ with distinct non-zero pairwise component distances such that, for any $1 \leq i, j, k \leq n$, $1 \leq \ell \leq d$, $|(z_i)_\ell - (z_k)_\ell| \neq |(z_j)_\ell - (z_k)_\ell|$ whenever $i \neq j$, the anisotropic Gaussian interpolation matrix $\mathbf{J}(Z, Z; \boldsymbol{\mathcal{E}})$ in (3.1) is invertible for $\boldsymbol{\mathcal{E}}$ in almost everywhere of $(\mathbb{R}^+)^{d \times n}$ with respect to the Lebesgue measure.*

PROOF: We apply the rank-1 perturbation technique in the proof of Theorem 2.2. The differential geometries of curves and hyperplane remain unchanged in the anisotropic setup, but we need to consider different space curves. To begin, we consider the space curve given as

$$\gamma(t) := \left\{ \mathbf{p} : [\mathbf{p}]_i = J(z_i, z_1; [\sqrt{t}, \mathbf{1}_{d-1}]) - J(z_i, z_1; \mathbf{1}_d), \quad 1 \leq i \leq n \right\} \subset \mathbb{R}^n, \quad (3.2)$$

where $\mathbf{1}_\ell$ denotes the all-ones vector in \mathbb{R}^ℓ . The corresponding derivative matrix explicitly is now

$$[D_\gamma(t)]_{ij} = (-1)^j |(z_i)_1 - (z_1)_1|^{2j} e^{-t|(z_i)_1 - (z_1)_1|^2} e^{-\|z_i - z_1\|^2 + |(z_i)_1 - (z_1)_1|^2},$$

which is the product of a full rank diagonal matrix and the transpose of a Vandermonde matrix generated by the set of scalars

$$\{0, |(z_2)_1 - (z_1)_1|^2, \dots, |(z_n)_1 - (z_1)_1|^2\}.$$

Under the assumption of distinct non-zero pairwise component distances, the Vandermonde matrix is of rank $n - 1$. We can ensure this perturbed matrix, denoted by $\mathbf{A}_{0,1}$, is invertible for $[\boldsymbol{\varepsilon}_1]_1$ in almost everywhere of \mathbb{R}^+ . To obtain the first column of (3.1), we need d perturbations in total:

$$\mathbf{A}_0 \rightarrow \mathbf{A}_{0,1} \rightarrow \dots \rightarrow \mathbf{A}_{0,d} =: \mathbf{A}_1.$$

The proof can be completed after nd iterations to arrive at \mathbf{A}_n , at each of which only a measure zero set of parameters $[\boldsymbol{\varepsilon}_k]_\ell \in \mathbb{R}^+$, $1 \leq k \leq n$ and $1 \leq \ell \leq d$, leads to singular matrices. \square

An even more general form of *rotated* anisotropic Gaussian is

$$J(\cdot, x_k; \boldsymbol{\varepsilon}_k, \boldsymbol{\theta}_k) = \exp\left(-(\cdot - z_k)^T \mathbf{R}_k^T \text{diag}(\boldsymbol{\varepsilon}_k)^2 \mathbf{R}_k (\cdot - z_k)\right), \quad (3.3)$$

where $\mathbf{R}_k = \mathbf{R}(\boldsymbol{\theta}_k) = \mathbf{R}_{x_1}([\boldsymbol{\theta}_k]_1) \cdots \mathbf{R}_{x_d}([\boldsymbol{\theta}_k]_d)$ is some elemental rotation matrix defined by the vector $\boldsymbol{\theta}_k \in \mathbb{R}^d$. Collecting all $1 \leq k \leq n$ angle-vectors $\boldsymbol{\theta}_k$ in a matrix $\boldsymbol{\Theta} \in \mathbb{R}^{d \times n}$ allows us to define an interpolation matrix $J(Z, Z; \boldsymbol{\varepsilon}, \boldsymbol{\Theta})$.

COROLLARY 3.2. *Under the assumptions in Theorem 3.1, the rotated anisotropic Gaussian interpolation matrix $J(Z, Z; \boldsymbol{\varepsilon}, \boldsymbol{\Theta})$ generated by basis (3.3) is invertible with respect to $\boldsymbol{\varepsilon} \in (\mathbb{R}^+)^{d \times n}$ and $\boldsymbol{\Theta} \in \mathbb{R}^{d \times n}$ in almost everywhere of $(\mathbb{R}^+)^{d \times n} \times \mathbb{R}^{d \times n}$ with respect to the Lebesgue measure.*

PROOF: Note that the proof of Theorem 3.1 only relies on the matrix in the previous iteration being invertible. At the k -th column update, we can operate on a rotated set of points instead of the given Z . The argument in the proof remains valid as far as the rotated pairwise component distances are all distinct and non-zero. Let $r_{ij} := z_i - z_j$ subject to $\ell - 1$ previous successful rotation so that the matrix $\mathbf{A}_{k,\ell-1}$ is invertible. The assumption of Theorem 3.1 is equivalent to $|(r_{ij})_\ell| > 0$ and $|(r_{il})_\ell| \neq |(r_{jl})_\ell|$ for any $1 \leq i, j, l \leq n$ and component $1 \leq \ell \leq d$. For each fixed (i, j, l) , only finitely many angles $\theta \in [0, 2\pi)$, or countably many in \mathbb{R} , for rotations about the x_ℓ -axis can be obtained from solving the following equalities for the rotated distances:

$$|(\mathbf{R}_{x_\ell}(\theta)r_{il})_{\ell'}| = |(\mathbf{R}_{x_\ell}(\theta)r_{jl})_{\ell'}| \text{ or } |(\mathbf{R}_{x_\ell}(\theta)r_{il})_{\ell'}| = 0, \quad (3.4)$$

for each $\ell' \neq \ell$. If (3.4) holds, it means rotating about the x_ℓ -axis by θ will result in non-distinct or non-zero pairwise component distances. However, there are only countably many such angles θ for each component of $\boldsymbol{\theta}_k$, and hence are of zero measure. \square

It worths noting the invertibility theorems presented here are Z location dependent. As adding more rows to an invertible matrix is safe with respect to the condition number, see [14, Thm. 2.2], solvability of kernel-based least-squares function approximations [5, 11, 13, 21] with non-constant shape parameters can also be safeguarded. In this setup, kernels were centered at Z and collocation were done at $X \supset Z$ such that the interpolation matrix studied in this work is a submatrix of the collocation matrix.

4. Numerical accuracy in variable/double precisions. Numerical evidence in literature suggests that asymmetric interpolation methods can numerically outperform the traditional symmetric approach [3, 14, 23]. In [3], the authors showed that non-constant shape parameters approach yields interpolation matrices with certain eigenvalue patterns as $\|\epsilon\|_\infty \rightarrow 0$, which differ from the constants shape parameters. In this section, we take a deeper look into these observations without letting basis go flat. Firstly, the following example shows that there are no “theoretical” advantages in terms of convergence for asymmetric interpolations.

EXAMPLE 4.1. *We consider interpolation problems on equally distributed data points using Gaussian basis. Computations were carried out with 512 digits precision to eliminate the effect of ill-conditioning. For symmetric interpolation, we used constant integer shape parameters $1 \leq \epsilon \leq 6$. For the asymmetric case, we use random shape parameters that follows $\chi^2(3)$ -distribution, which is the distribution used in [23].*

The bandlimited function $f = \text{sinc}(2x + 0.5) : [-1, 1] \rightarrow \mathbb{R}$ is considered. It lies in

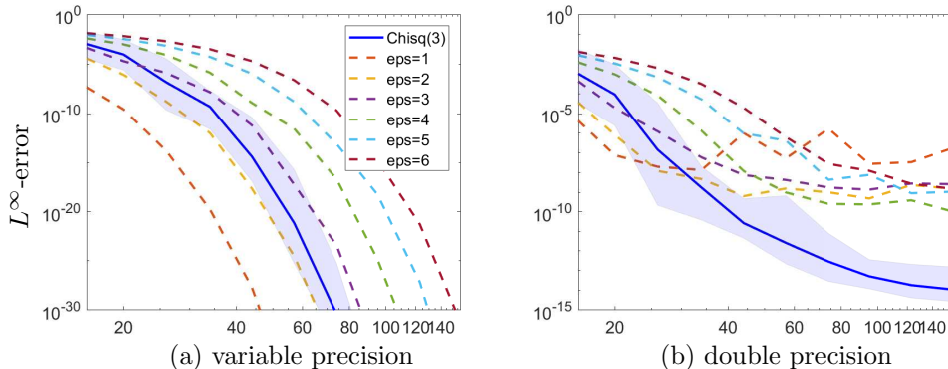


FIG. 4.1. L^∞ -convergence profiles against the number of data points n for interpolating $f_1 = \text{sinc}(2x + 0.5)$ in (a) variable precision and (b) double precision.

the reproducing kernel Hilbert space of all scaled Gaussian kernels. The well-known exponential convergence of symmetric Gaussian interpolation (towards functions in its reproducing kernel Hilbert space) is clearly observed here. The same convergence behavior can be seen in asymmetric interpolations. In terms of accuracy, asymmetric interpolations (with these particular choices of random shape parameters) perform similarly to $\epsilon = 2$ or 3 . The resulting maximum interpolation error profiles were shown in the left half of Figure 4.1. For the asymmetric approach, we show the median of all test runs in a solid line and the full range of error in shade.

For comparison, the right half of Figure 4.1 shows the double precision results of the same test. In presence of rounding error, asymmetric interpolation clearly suffers less from the ill-conditioning of the interpolation basis and yields better accuracy. \square

Example 4.1 suggests that the improved accuracy we observed in double precision is solely a numerical issue. To gain more insights, we will analyze the effect of non-constant shape parameters on the singular values of the asymmetric interpolation matrix. First, the following theorem inserts an upper bound on the maximum singular value.

THEOREM 4.2. *The largest singular values $\sigma_1(\mathbf{J}(Z, Z; \boldsymbol{\epsilon}))$ of the variable shape Gaussian interpolation matrix in (1.3), for $k = 1, \dots, n$, are bounded above by*

$$\sigma_1(\mathbf{J}(Z, Z; \boldsymbol{\epsilon})) \leq \sqrt{\sum_{\epsilon \in \boldsymbol{\epsilon}} \lambda_{\max}^2(\mathbf{J}(Z, Z; \epsilon))},$$

where $\lambda_{\max}(\mathbf{J}(Z, Z; \epsilon))$ denotes the maximum eigenvalue of the standard symmetric interpolation matrix $\mathbf{J}(Z, Z; \epsilon)$ corresponding to the scaled Gaussian kernel $J(\cdot, \cdot; \epsilon)$ with a constant shape parameter ϵ .

PROOF: Let $\mathbf{c} = [c_1, \dots, c_n]^T$ with $\|\mathbf{c}\| = 1$ be an eigenvector of $\mathbf{J}(Z, Z; \boldsymbol{\epsilon})\mathbf{J}^T(Z, Z; \boldsymbol{\epsilon})$ corresponding to the maximum eigenvalue $\lambda_{\max}(\mathbf{J}(Z, Z; \boldsymbol{\epsilon})\mathbf{J}^T(Z, Z; \boldsymbol{\epsilon}))$. Then, we have

$$\begin{aligned}
\lambda_{\max}(\mathbf{J}(Z, Z; \boldsymbol{\epsilon})\mathbf{J}^T(Z, Z; \boldsymbol{\epsilon})) &= \sum_{i=1}^n \sum_{j=1}^n c_i c_j \sum_{k=1}^n \mathbf{J}(z_i, z_k; \epsilon_k) \mathbf{J}(z_j, z_k; \epsilon_k) \\
&= \sum_{k=1}^n \left(\sum_{i=1}^n c_i \mathbf{J}(z_i, z_k; \epsilon_k) \right)^2 \\
&= \sum_{k=1}^n \left(\sum_{i=1}^n c_i \mathbf{J}(z_i, Z; \epsilon_k) \mathbf{e}_k \right)^2 \\
&= \sum_{k=1}^n \left(\mathbf{c}^T \mathbf{J}(Z, Z; \epsilon_k) \mathbf{e}_k \right)^2, \tag{4.1}
\end{aligned}$$

with \mathbf{e}_k denoting the k -th standard basis of \mathbb{R}^n . Because

$$|\mathbf{c}^T \mathbf{J}(Z, Z; \epsilon_k) \mathbf{e}_k| \leq \|\mathbf{J}(Z, Z; \epsilon_k) \mathbf{c}\|_2 \leq \lambda_{\max}(\mathbf{J}(Z, Z; \epsilon)), \tag{4.2}$$

for each $k = 1, \dots, n$, substituting the rightmost term of (4.2) into (4.1) and then taking the square root of both sides of (4.1) yield the asserted upper bound. \square

In \mathbb{R}^d , the symmetric Gaussian interpolation matrix $\mathbf{J}(Z, Z; \epsilon)$ has $\binom{m+d-1}{d-1}$ eigenvalues of order $\mathcal{O}(\epsilon^{2m})$. From the fact that, when $m = 0$, the maximum eigenvalue is of $\mathcal{O}(1)$, the same property holds for the asymmetric counterpart. In a constant shape parameter approach, the problem of ill-conditioning becomes severe because of the cluster of small eigenvalues on the right of zero.

Gaussian interpolation with non-constant shape parameters improves accuracy by breaking symmetry; this results in sending some of the nearly zero eigenvalues out to the complex plane and away from the origin. For asymmetric linear systems, we focus on singular values instead of eigenvalues (i.e., spectrum). Suppose that an asymmetric Gaussian matrix $\mathbf{J} = \mathbf{J}(Z, Z; \boldsymbol{\epsilon})$ has an eigenvalue decomposition $\mathbf{J} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$.

PROPOSITION 4.3. [9, Sec. 7.3, p.18] If $\mathbf{A}, \mathbf{B}^T \in \mathbb{R}^{m \times n}$ and $k = \min\{m, n\}$, then the singular values satisfy $\sigma_{i+j-1}(\mathbf{A}\mathbf{B}) \leq \sigma_i(\mathbf{A})\sigma_j(\mathbf{B})$ for $i, j = 1, \dots, k$ and $i + j \leq k + 1$.

Applying Proposition 4.3 twice (with $j = 1$) yields

$$\sigma_i(\mathbf{J}) = \sigma_i(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}) \leq \sigma_1(\mathbf{V})\sigma_i(\boldsymbol{\Lambda})\sigma_1(\mathbf{V}^{-1}).$$

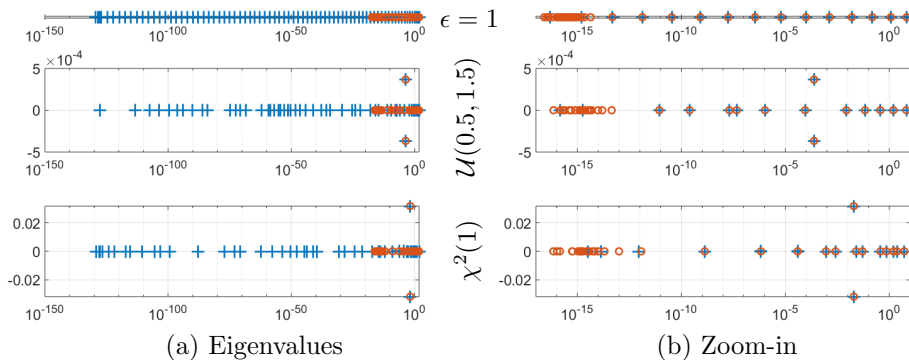


FIG. 4.2. Spectrum of various Gaussian interpolation matrices with (top to bottom) constant shape $\epsilon = 1$, random $\epsilon \sim (\mathcal{U}(0.5, 1.5))^n$, and random $\epsilon \sim (\chi^2(1))^n$. Markers $+$ and \circ indicate computational results from variable and double precision respectively. The left hand panel (a) shows all eigenvalues, whereas the right (b) shows the ones with magnitude of real parts in $[10^{-17}, 10]$.

Moreover, we also have

$$\sigma_i(\Lambda) = \sigma_i(\mathbf{V}^{-1}\mathbf{J}\mathbf{V}) \leq \sigma_1(\mathbf{V}^{-1})\sigma_i(\mathbf{J})\sigma_1(\mathbf{V}).$$

Together with $\kappa(\mathbf{V}) = \sigma_1(\mathbf{V}^{-1})\sigma_1(\mathbf{V})$, we see the connection between singular values and eigenvalues:

THEOREM 4.4. *For any $n \times n$ matrix \mathbf{J} that admits an eigenvalue decomposition $\mathbf{J} = \mathbf{V}\Lambda\mathbf{V}^{-1}$, suppose the eigenvalues are in nonincreasing order so that $|\lambda_i(\Lambda)| = \sigma_i(\Lambda)$ for $1 \leq i \leq n$, then we have*

$$\frac{1}{\kappa(\mathbf{V})}\sigma_i(\mathbf{J}) \leq |\lambda_i(\mathbf{J})| \leq \kappa(\mathbf{V})\sigma_i(\mathbf{J}),$$

where $\kappa(\mathbf{V})$ is the 2-norm condition number of \mathbf{V} .

Theorem 4.4 suggests that some singular values in machine-epsilon magnitude of an asymmetric interpolation matrix \mathbf{J} could allow the corresponding eigenvalues to go into the complex plane and become numerically significant again up to a factor of $\kappa(\mathbf{V})$. The following example demonstrates how random shape parameters affect the spectrum of Gaussian interpolation matrices, and hence, their singular values.

EXAMPLE 4.5. *We consider Gaussian interpolation matrices on $n = 100$ equally spaced data points in $[-1, 1]$. Figure 4.2(a) shows all eigenvalues with positive real part of*

- constant $\epsilon = \mathbf{1}_n$,
- random, independent and identically distributed $\epsilon \sim (\mathcal{U}(0.5, 1.5))^n$, and
- random, independent and identically distributed $\epsilon \sim (\chi^2(1))^n$,

computed under 128 digits (+) and double precision (o). The cluster of + on the left most is numerical artifact and we expect a nice equally space spectrum in log-scale in the absence of rounding error. With shape parameters bounded away from zero, the eigenvalue patterns observed in [3] disappear from both random cases. Figure 4.2(b) zooms into the real axis between 10^{-17} to 10. Both asymmetric interpolation matrices have 45 negative real eigenvalues; 7 and 9 of those have magnitudes greater than $-2.2e - 16$ for $\mathcal{U}(0.5, 1.5)$ and $\chi^2(1)$, respectively. The condition numbers $\kappa(\mathbf{V})$ seen in Theorem 4.4 are around 500 and 10000 in these two random strategies.

From Figure 4.2, asymmetric interpolation matrices still contain many numerically zero eigenvalues. We therefore cannot expect any improvement in terms of condition number from the non-constant shape parameters approach. The effect of allowing complex eigenvalues in the asymmetric approach is a higher (double precision) numerical rank: 13, 19, and 27 respectively for the three test cases reported by MATLAB build-in `rank` function. We see that a larger $\kappa(\mathbf{V})$ results in a higher numerical rank, i.e., the number of numerically nonzero singular values. \square

The message here is that there need not be some intelligent structure for choosing the spatially varying non-constant shape parameters; even a random choice can produce improvements. Because of the new theories, we can safely leverage even unintelligent strategies for improving the performance of the interpolant *in double precision*.

5. Parameter-free random shape strategy. Using non-constant parameter in asymmetric interpolation does not eliminate the need for some choice of shape parameter throughout the domain. For these parameters to be chosen deterministically, some function would be needed to define the parameter ϵ_i given the location x_i and, potentially, all the other locations x_1, \dots, x_n in the problem. For random shape parameters, a user-chosen distribution is needed which will likely have its own free parameters; more discussion on this is presented in Section 6.1. As a corollary of Theorem 2.2, we propose a parameter-free strategy in picking random shape parameters.

For each data point $z_k \in Z \subset \Omega \subset \mathbb{R}^n$, the motivation is to randomly sample its shape parameter ϵ_k that minimizes the risk of having a singular asymmetric interpolation matrix. A zero determinant occurs exactly when the space curve $\gamma(t) = \gamma(t; z_k)$ for z_k in (2.3) intersects with the hyperplane \mathcal{S} in (2.2), which is determined by all the other shape parameters ϵ_j with $j \neq k$. Such dependence also makes the hyperplane \mathcal{S} random. Theorem 2.2 shows that if the distribution of each ϵ_k is absolutely continuous with respect to the Lebesgue measure, we will have almost sure invertibility.

To guarantee that we will not incur the zero probability singular situation, we

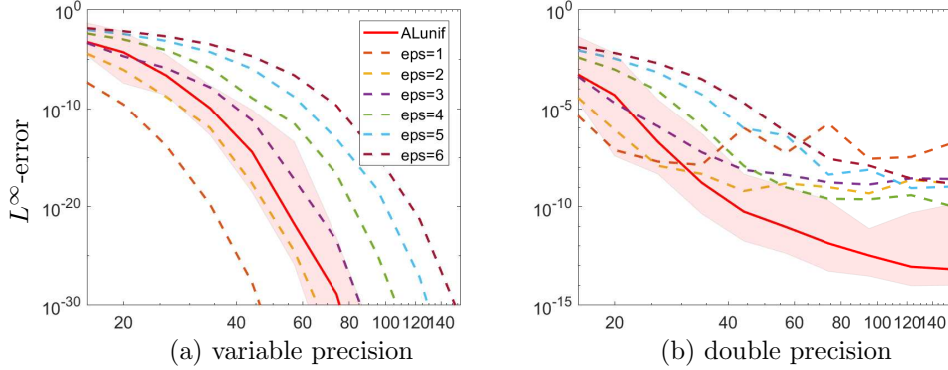


FIG. 5.1. L^∞ -convergence profiles against the number of data points n for interpolating $f_1 = \text{sinc}(2x + 0.5)$ in (a) variable precision and (b) double precision.

suggest that the shape parameter distribution should be chosen in order to get a uniformly distributed space curve value. In other words, we want to pick any point on the space curve, between the start of no change and the end which results in singularity, with equal chance. This can be done by using the arc length variable

$$s(t; z_k) := \int_{\tau=0}^t \|\gamma'(\tau; z_k)\|_{\ell^2(\mathbb{R}^n)} d\tau, \quad 0 < t < \infty.$$

Then, apply inverse transform sampling using the cumulative distribution function $s(t; z_k)/s(\infty; z_k)$ yields an *arc-length uniform distribution*, denoted by

$$\mathcal{A}(t; z_k, Z) := \frac{d}{dt} \frac{s(t; z_k)}{s(\infty; z_k)} = \frac{\|\gamma'(t; z_k)\|_{\ell^2(\mathbb{R}^n)}}{s(\infty; z_k)}, \quad (5.1)$$

for the random shape parameter ϵ_k to follow. One clear advantage is that $\mathcal{A}(t; z_k, Z)$ depends only on the data structure and is free from other parameters. One can easily differentiate (2.3) analytically to obtain a close form formula for $\|\gamma'(t; z_k)\|_{\ell^2(\mathbb{R}^n)}$. Computing (5.1) then requires applications of some numerical quadratures.

EXAMPLE 5.1. We repeat Example 4.1 with random shape parameters that follow the arc-length uniform distribution (5.1) and show the results in Figure 5.1. The mean of all errors of this parameter-free approach is comparable to that of the carefully chosen $\chi^2(3)$ distribution but with a bit wider range of errors.

We emphasize the idea that the data present in this experimental demonstration comes from an analytic function with no noise. The results here do not suggest a unreasonably simplistic picture of RBF interpolation that smaller shape parameters will always yield better outcomes. \square

Following the same line of logic, one can derive a parameter-free distribution for each component of \mathcal{E} in the anisotropic setting. In the following section, we will not go further into this direction, but instead focus on using some standard probability distributions to generate shape parameters.

6. Numerical exploration of different random parameter distributions.

Our empirical analysis of spatially varying shape parameter strategies earlier (in Figure 4.2 and Figure 5.1) has demonstrated that, while in infinite precision there is likely no theoretical benefit to using randomly varying shape parameters, in finite precision there is an opportunity to perform better. This occurs as larger amounts of data leading to an ill-conditioned basis, which leads to a logical conjecture: spatially varying shape parameters can be chosen randomly to outperform a single shape parameter. In this section, we explore that conjecture, with various possible shape parameter distributions on examples from different dimensions.

Unfortunately, the parameter-free random shape strategy in Section 5 is computationally intensive. Any choice of distribution from which we randomly draw shape parameters would, itself, require some parametrization. In this section, we refer to the free parameters associated with such a distribution as *hyperparameters*; this term is common in Bayesian modeling [4, Chapter 5], whereby it defines the distribution of another free parameter¹.

Our initial analysis considers the impact of such hyperparameters when randomly generating random shape parameters. We also try to provide some guidance regarding how the distributions might be chosen in various circumstances.

6.1. Random distributions under consideration. We begin by assuming that some base ϵ_0 value is present: this would be the single shape parameter which would be otherwise used in the symmetric setting. Several possible distributions are considered with this ϵ_0 value as their mean and variance left as a free parameter (when it varies freely from the mean). Other distributions could also be considered and may be preferable in some settings; these are meant to provide some initial insights.

One distribution of interest is a *log-uniform* distribution ϵ_u with components

$$\log([\epsilon_u]_k) \sim \text{Unif}(\log(\nu/\tau), \log(\nu\tau)), \quad 1 \leq k \leq n, \quad (6.1a)$$

where $\nu > 0$ and $\tau \in (0, 1)$ represent a sort of center and spread of the distribution, respectively. This distribution produces points which are logarithmically spaced

¹Interestingly, if kernel interpolation is interpreted through Gaussian processes [2], the shape parameter ϵ would be called a hyperparameter, because the interpolation coefficients are the actual *parameters* of interest in the model. In that setting, τ would be a hyper-hyperparameter, or simply just another hyperparameter. See [18] for an in depth treatment of the topic.

between ν/τ and $\nu\tau$. Because

$$\begin{aligned}\mathbb{E}([\boldsymbol{\varepsilon}_u]_k) &= \nu(\tau - 1/\tau)/(2 \log \tau), \\ \text{Var}([\boldsymbol{\varepsilon}_u]_k) &= (\nu(\tau - 1/\tau)/(2 \log \tau))^2 \left(\log \tau \frac{\tau + 1/\tau}{\tau - 1/\tau} - 1 \right),\end{aligned}$$

and we require $\mathbb{E}([\boldsymbol{\varepsilon}_u]_k) = \epsilon_0$, this fixes $\nu = 2\epsilon_0 \log(\tau)/(\tau - 1/\tau)$ for all $1 \leq k \leq n$. We choose τ for a desired variance $\text{Var}([\boldsymbol{\varepsilon}_u]_k) > 0$ by solving the nonlinear equation

$$\log \tau \frac{\tau + 1/\tau}{\tau - 1/\tau} = 1 + \frac{\text{Var}([\boldsymbol{\varepsilon}_u]_k)}{\epsilon_0^2}.$$

We can also consider the *log-normal* distribution $\boldsymbol{\varepsilon}_n$ with components

$$\log([\boldsymbol{\varepsilon}_n]_k) \sim \mathcal{N}(\mu, \sigma^2), \quad 1 \leq k \leq n, \quad (6.1b)$$

for $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. For this random variable,

$$\mathbb{E}([\boldsymbol{\varepsilon}_n]_k) = e^{\mu + \sigma^2/2} \quad \text{and} \quad \text{Var}([\boldsymbol{\varepsilon}_n]_k) = (e^{\sigma^2/2} - 1)e^{2\mu + \sigma^2}.$$

Requiring $\mathbb{E}([\boldsymbol{\varepsilon}_n]_k) = \epsilon_0$ and given a desired variance $\text{Var}([\boldsymbol{\varepsilon}_n]_k) > 0$ produces parameter values

$$\sigma^2 = \log(1 + \text{Var}([\boldsymbol{\varepsilon}_n]_k)/\epsilon_0^2) \quad \text{and} \quad \mu = \log\left(\epsilon_0^2/\sqrt{\epsilon_0^2 + \text{Var}([\boldsymbol{\varepsilon}_n]_k)}\right).$$

Another distribution, which was used in [23], is the *Chi-squared* distribution $\boldsymbol{\varepsilon}_\chi$ with components

$$[\boldsymbol{\varepsilon}_\chi]_k \sim \chi^2(\nu), \quad 1 \leq k \leq n. \quad (6.1c)$$

This distribution has only one free parameter, which is fixed to $\nu = \epsilon_0$ so that the mean is ϵ_0 . This produces a fixed variance of $\text{Var}([\boldsymbol{\varepsilon}_\chi]_k) = \nu^2$. The *Gamma* distribution $\boldsymbol{\varepsilon}_g$ generalizes the Chi-Squared distribution, so we may also consider

$$[\boldsymbol{\varepsilon}_g]_k \sim \text{Gamma}(\alpha, \beta), \quad 1 \leq k \leq n, \quad (6.1d)$$

where $\alpha, \beta > 0$ are shape and rate parameters respectively. Because $\mathbb{E}([\boldsymbol{\varepsilon}_g]_k) = \alpha\beta$ and $\text{Var}([\boldsymbol{\varepsilon}_g]_k) = \alpha\beta^2$ we require

$$\alpha = \epsilon_0^2/\text{Var}([\boldsymbol{\varepsilon}_g]_k) \quad \text{and} \quad \beta = \text{Var}([\boldsymbol{\varepsilon}_g]_k)/\epsilon_0.$$

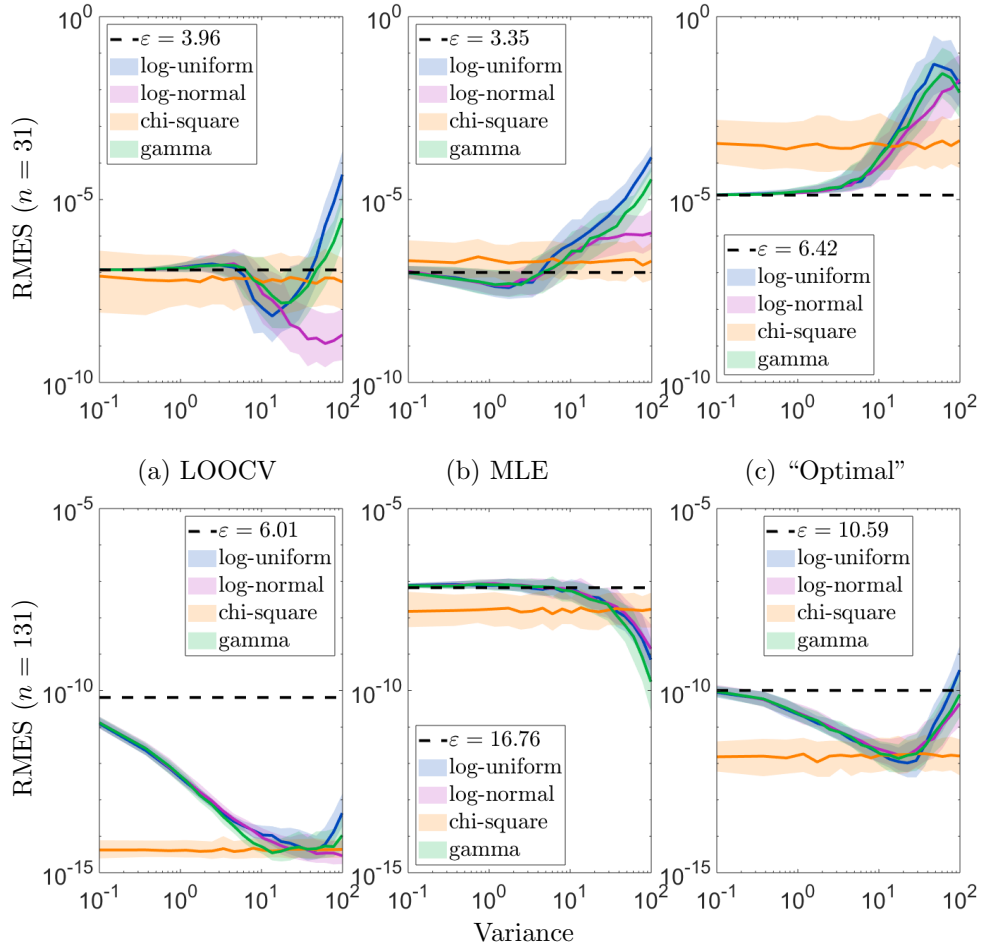


FIG. 6.1. The distributions defined in (6.1) are used to create interpolants of three functions over 250 trials; median and interquartile RMSE values are presented. In each graph, a specific ϵ_0 value is chosen as the distribution mean, and a range of variance values is studied. (top) $n = 31$. (bottom) $n = 131$. (a) The LOOCV ϵ_0 is used on data sampled from the function $f(x) = \sin(6x) - x$. (b) The MLE ϵ_0 is used on data sampled from the function $f(x) = 1 - \tanh(3x) + e^x$. (c) The best ϵ_0 , as judged by the accuracy on the evaluation points, is used on data sampled from the function $f(x) = 1/(1 + 25(x - .3)^2)$.

6.2. Interpolation in 1D. For these first experiments, we use the truncated SVD (tolerance of 10^{-14}) to solve linear systems with the pseudoinverse; we feel that this helps manage any potential disparity of solving symmetric and asymmetric circumstances while also allowing for a consistent treatment of ill-conditioning. In Figure 6.1, we consider interpolant accuracy results from 3 different functions when either $n = 31$ or $n = 131$ evenly spaced points were sampled in the domain $[-1, 1]$. In all tests, 100 evenly spaced evaluation points were used.

For a small number of points, there seems to be little benefit and some potential penalty in using the random shape parameter strategy for many different variance

values. However, with a larger problem size, there is significant potential benefit and reduced potential for penalty. Even in the bottom right graph of Figure 6.1, where we have “cheated” and chosen the best ϵ_0 value possible so as to minimize the error, the use of random shape parameters to produced an asymmetric matrix provided better accuracy (for certain variances).

6.3. Singular value analysis. The top and bottom rows of Figure 6.1 show distinctly different behavior: in the top row, with only $n = 31$ points, the random distribution gives little benefit, whereas in the bottom row, with $n = 131$ points, there is much more consistent benefit. In Figure 6.2 we explore the spread of singular values of these symmetric and asymmetric interpolation matrices to study their role in this discrepancy. To simplify the analysis, the variance was fixed at 10.0 for all the distributions where it was a free parameter.

For smaller matrices (in the top row), essentially all of the singular values are greater than machine precision. In contrast, when $n = 131$ produces larger kernel matrices, the bottom row shows that more of those singular values fall beyond machine precision. The asymmetric kernel matrices have noticeably larger singular values in two of the situations; both of these situations were able to produce more accurate interpolants, as seen in Figure 6.1.

6.4. Interpolation in 2D and 3D. We consider now a 2-dimensional example, at which point additional flexibility arises in the random kernel approximation. We use the standard `MATLAB \` operator for solving linear systems, in lieu of the truncated SVD from Section 6.2. For each kernel, an *anisotropic* kernel is chosen with different shape parameters randomly selected for each dimension. To simplify our analysis, we consider only the Gamma distribution (6.1d) with $\epsilon_0 = \log(n)$ and several possible variances. A random isotropic kernel could also be chosen, but that seems unnecessarily limiting given the freedom we hope to unlock by using randomly selected shape parameters.

The function $\sin(6(x_1^2 + x_2^2))$ on $[-1, 1]^2$ is considered in Figure 6.3; n Halton [20] points were sampled, and the error of the interpolant is computed at 400 Halton points on the convex hull of the points sampled. Figure 6.3(a) shows that the standard isometric symmetric interpolant has its optimal accuracy for larger n with $\epsilon \in [\log(n)/\sqrt{10}, \log(n)\sqrt{10}]$. Figure 6.3(b) shows that there is an opportunity to achieve better interpolant accuracy at large n values, if the variance is suitably large. However, those larger variance values yield worse accuracy for smaller n values where the kernel matrix is still well-conditioned (as suggested in Figure 6.3).

In 3D, we are interested in some relatively fine tuning free approach to get better interpolation accuracy. We consider interpolating $\exp(x + 2y)z$ in $[0, 1]^3$ using the

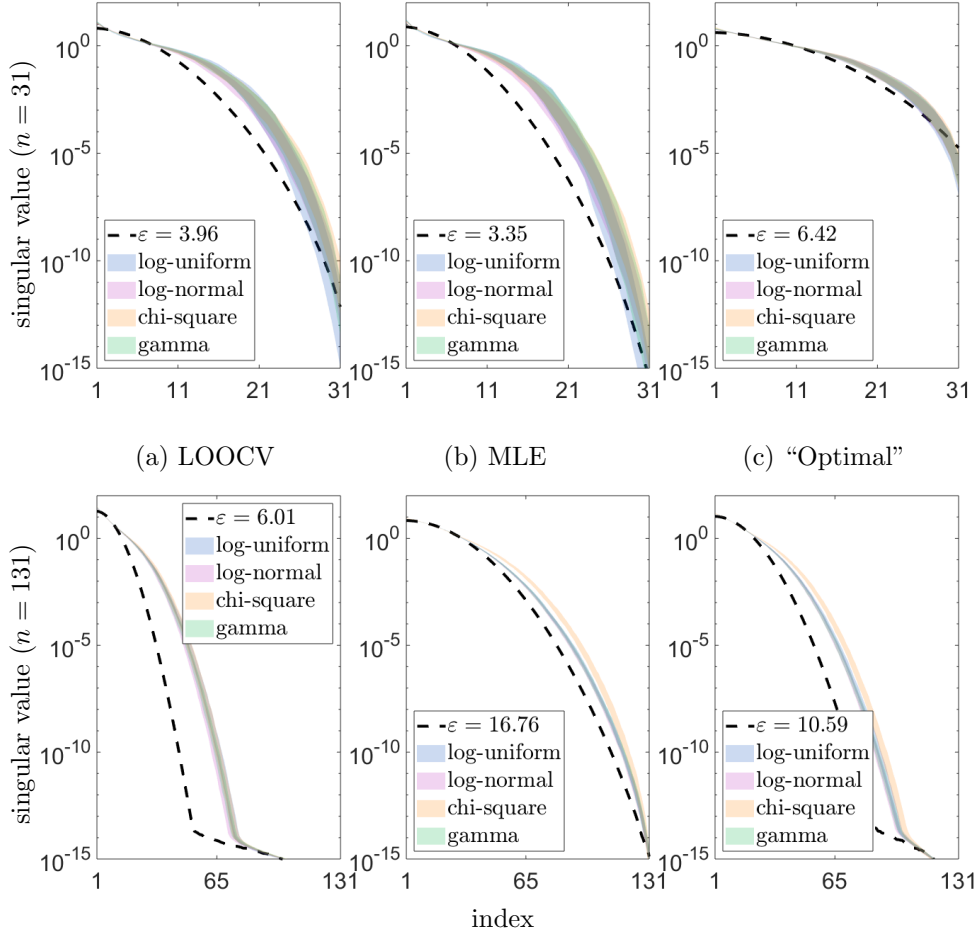


FIG. 6.2. The spread of singular values of symmetric and asymmetric kernel matrices. For the asymmetric kernel matrices, drawn from the distributions in (6.1) with variance fixed at 10, 250 draws were considered and the interquartile range is plotted. (top) $n = 31$. (bottom) $n = 131$. (a) The LOOCV ϵ_0 is used on data sampled from the function $f(x) = \sin(6x) - x$. (b) The MLE ϵ_0 is used on data sampled from the function $f(x) = 1 - \tanh(3x) + e^x$. (c) The best ϵ_0 , as judged by the accuracy on the evaluation points, is used on data sampled from the function $f(x) = 1/(1 + 25(x - .3)^2)$.

anisotropic Gaussian kernel interpolation in Section 3. We use $n = 3000$ Halton points for interpolation and the error is evaluated at 20^3 regularly placed points. Two random shape parameter strategies were considered: Chi-squared distribution $[\epsilon_\chi]_k \sim \chi^2(\max\{\epsilon_0, 1\})$ and log-normal distribution $[\epsilon_n]_k \sim \exp(\mathcal{N}(\log \epsilon_0, 1))$ in Figure 6.4(a) and (b) respectively for $1 \leq k \leq n$. For each parameter ϵ_0 , we run each strategy 100 times to collect the median and range of error. The L_∞ -error resulting from constant shape parameter ϵ_0 is also shown for comparison. Chi-squared distribution seems to be a bad strategy for ϵ_0 . In other cases, it is probabilistically safe to say that random shape parameter interpolation yields better accuracy than the traditional constant shape parameter approach.

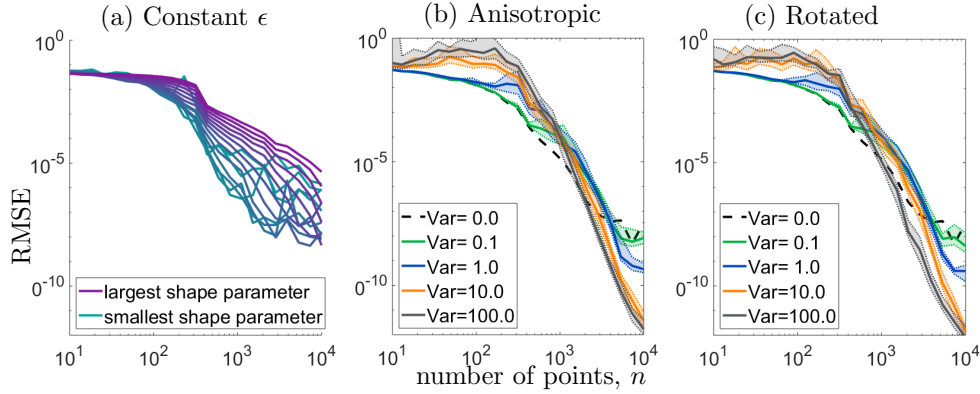


FIG. 6.3. (a) We study symmetric interpolant quality as a function of n for 15 ϵ values logarithmically spaced between $[\log(n)/\sqrt{10}, \log(n)\sqrt{10}]$. (b) Randomly chosen shape parameters are considered with different variance values (including the symmetric case $\text{Var} = 0$) for $\epsilon_0 = \log(n)$; the median outcome of 25 trials is plotted as well as the interquartile range. (c) Rotated kernels, as described in Section 6.5, are used to replace the standard anisotropic kernels with little or no impact on accuracy.

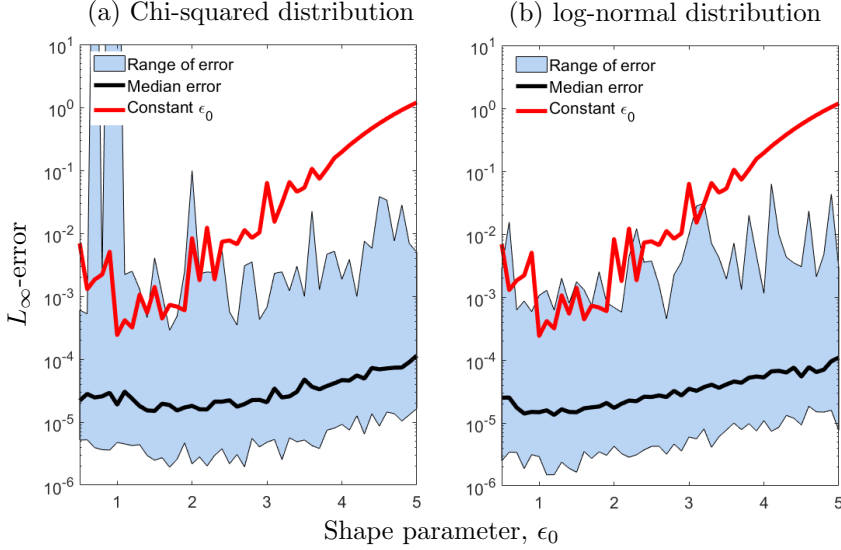


FIG. 6.4. We study the L_∞ -error of a 3D interpolation problem as a function shape/free parameters ϵ_0 that generate random shape parameters by (a) Chi-squared $\chi^2(\max\{\epsilon_0, 1\})$ and (b) log-normal $\exp(\mathcal{N}(\log \epsilon_0, 1))$ distributions. Median and range of error were computed based on 100 trial runs for each tested ϵ_0 .

6.5. Rotated Gaussian kernels. We also experiment with the “rotated” Gaussian kernels defined in Corollary 3.2 to study potential benefits from randomly chosen distance definitions. In this situation, the concept of a random distribution of distance parameters is slightly more complicated because of the rotation present.

Distance is defined as $\|\mathbf{x} - \mathbf{z}\|_E = (\mathbf{x} - \mathbf{z})^T E (\mathbf{x} - \mathbf{z})$ for some symmetric positive definite (SPD) matrix $E \in \mathbb{R}^{d \times d}$. Because all SPD matrices E have a unique Cholesky factorization, they are defined uniquely with $d(d + 1)/2$ values. In our case, however,

we choose to utilize the eigenvalue decomposition in order to define an \mathbf{E} matrix which has directly manipulable length scales. We write

$$\mathbf{E} = \mathbf{V}\mathbf{S}\mathbf{V}^T = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_d \end{pmatrix} \begin{pmatrix} s_1^2 & & \\ & \ddots & \\ & & s_d^2 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_d^T \end{pmatrix}.$$

The s_1, \dots, s_d values are chosen from one of the distributions defined in Section 6.1; again, we use only (6.1d) for ease of analysis. Because \mathbf{V} is orthogonal, there are only $d(d-1)/2$ free parameters; such a matrix can be randomly generated by taking the QR factorization of a $d \times d$ matrix with standard normal random entries [24].

As we can see in Figure 6.3(c), the use of the rotated kernels in that particular experiment provided no immediate benefit or penalty over the standard anisotropic Gaussians. While it seems likely that adding the additional degrees of freedom can yield benefits, more experimentation is required to identify those circumstances.

7. Conclusion and future work. We have proved that, subject to some restrictions on the data under analysis, Gaussian kernels will almost surely produce invertible matrices even when using non-constant shape parameters. The initial result was extended to include Gaussian kernels with an anisotropic sense of distance, both when aligned with the domain axes and rotated away from the standard orientation.

Some theoretical and empirical analysis was conducted to consider the benefit of non-constant shape parameters on the numerical rank of the asymmetric kernel matrices. Experiments were conducted to show that, while there is no theoretical benefit of utilizing randomly chosen shape parameters on functions in the Gaussian RKHS, there could be a benefit in finite precision. We devised a strategy for choosing these random shape parameters in a safe fashion, and then experimented with more standard ways to generate random parameters to show that they can also provide benefit.

One ideal direction for future work which is afforded by the almost surely invertible theorems of Section 2 would be to attempt to carefully choose the shape parameters associated with each kernel center in some optimal way. Such a strategy could be based on the localized density of points, or perhaps on the desired accuracy in one region more than the rest of the domain. This would be in contrast (or perhaps in concert, with enough analysis) to our randomly chosen shape parameter strategy.

Another important step in this research will be to consider possible connections between the random choice of shape parameters and how random elements of other numerical computations parallel the structure here. One close example arises in the

use of kernels within support vector machines: in an attempt to minimize the cost associated with solving the required quadratic program, [17] suggests that a low dimensional representation can be constructed randomly and accurately solved more quickly. Both random selection combinations of features were discussed in [10] for so-called extreme learning machines (which are trying to deal with a large amount of data). The authors of [7] describe randomly interacting with matrices to produce factorizations.

In all of those random computational circumstances, there is a presumption that analyzing the large amount of data present in a random fashion can produce effective results. Our experimental strategy in Section 6 is similar: if we randomly choose shape parameters and have enough data, then some of the points will be given “appropriate” shape parameters. That statement is logical, so long as appropriateness is defined, but the implicit assumption is that the appropriately assigned kernels will contribute more strongly to the model than those that were given poor choice of shape parameter. While this seems to be the case numerically (which is why superior accuracy was observed in, e.g., Figure 6.3), it is likely that work already done on these nearby topics will help us explore the implications in this research.

Numerically, random shape RBF collocation methods were shown to be successful in solving PDEs [15] adaptively without parameters fine tuning. Although this work focuses on interpolation problems, our invertibility theories can be extended to asymmetric RBF collocation methods, a.k.a. Kansa methods, provided that we have an invertible constant shape Gaussian collocation matrix. In the context of Kansa methods, Hon and Schaback [8] numerically show that singular constant shape Gaussian collocation matrices are rare and difficult to construct. In examples there, sets of singular setup are curves in 2D domains, whose theoretical proof is still missing. Thus, we leave this as a concluding remark instead of a theorem. As constant shape asymmetric RBF collocation matrices do not have structured eigenvalues, we do not expect the variable shape approach will generate any. Yet, we do expect to see reductions in condition number as in the cases of interpolation.

Acknowledgements. This work was supported by Hong Kong Research Grant Council GRF Grants. We thank the two referees for their valuable comments.

REFERENCES

- [1] M. BOZZINI, L. LENARDUZZI, M. ROSSINI, AND R. SCHABACK, *Interpolation with variably scaled kernels*, IMA J. Numer. Anal., 35 (2015), pp. 199–219.
- [2] G. E. FASSHAUER AND M. J. MCCOURT, *Kernel-Based Approximation Methods Using MATLAB*, Interdisciplinary Mathematical Sciences, World Scientific Publishing, 2015.
- [3] B. FORNBERG AND J. ZUEV, *The Runge phenomenon and spatially variable shape parameters*

- in *RBF interpolation*, *Comput. Math. Appl.*, 54 (2007), pp. 379–398.
- [4] A. GELMAN, J. B. CARLIN, H. S. STERN, D. B. DUNSON, A. VEHTARI, AND D. B. RUBIN, *Bayesian data analysis*, vol. 2, CRC press Boca Raton, FL, 2014.
- [5] Q. L. GIA, F. NARCOWICH, J. WARD, AND H. WENDLAND, *Continuous and discrete least-squares approximation by radial basis functions on spheres*, *J. Approximation Theory*, 143 (2006), pp. 124–133.
- [6] A. GOLBABAI, E. MOHEBIANFAR, AND H. RABIEI, *On the new variable shape parameter strategies for radial basis functions*, *comput.. Appl. Math.*, 34 (2015), pp. 691–704.
- [7] N. HALKO, P.-G. MARTINSSON, AND J. A. TROPP, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, *SIAM review*, 53 (2011), pp. 217–288.
- [8] Y. C. HON AND R. SCHABACK, *On unsymmetric collocation by radial basis functions*, *Appl. Math. Comput.*, 119 (2001), pp. 177–186.
- [9] R. A. HORN AND C. R. JOHNSON, *Matrix analysis*, Cambridge university press, 1990.
- [10] G. HUANG, G.-B. HUANG, S. SONG, AND K. YOU, *Trends in extreme learning machines: A review*, *Neural Networks*, 61 (2015), pp. 32–48.
- [11] A. ISKE, *Approximation*, Springer-Lehrbuch Masterclass, Springer Spektrum, 2018.
- [12] E. J. KANSA AND R. E. CARLSON, *Improved accuracy of multiquadric interpolation using variable shape parameters*, *Comput. Math. Appl.*, 24 (1992), pp. 99–120.
- [13] S. LI, K. C. CHEUNG, AND L. LING, *Discrete least-squares radial basis functions approximations*, *Appl. Math. Comput.*, 355 (2019), pp. 542–552.
- [14] L. LING, *A fast block-greedy algorithm for quasi-optimal meshless trial subspace selection*, *SIAM J. Sci. Comput.*, 38 (2016), pp. A1224–A1250.
- [15] L. LING AND S. N. CHIU, *Fully adaptive kernel-based methods*, *Int. J. Numer. Methods Eng.*, 114 (2018), pp. 454–467.
- [16] V. LOMBARDI, M. BOZZI, AND L. PERREGRINI, *An improved meshless method for waveguide eigenvalue problems*, *IEEE Microw. Wirel. Compon. Lett.*, 27 (2017), pp. 1047–1049.
- [17] A. RAHIMI AND B. RECHT, *Random features for large-scale kernel machines*, in *Advances in neural information processing systems*, 2008, pp. 1177–1184.
- [18] C. E. RASMUSSEN AND C. WILLIAMS, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.
- [19] S. A. SARRA AND D. STURGILL, *A random variable shape parameter strategy for radial basis function approximation methods*, *Eng. Anal. Bound. Elem.*, 33 (2009), pp. 1239–1245.
- [20] X. WANG AND F. J. HICKERNELL, *Randomized halton sequences*, *Math. Comput. Modelling*, 32 (2000), pp. 887–899.
- [21] J. D. WARD, *Least squares approximation using radial basis functions: an update*, in *Advances in Constructive Approximation: Vanderbilt2003*, M. Neamtu and E. B. Saff, eds., Brentwood, TN, 2004, Nashboro Press, pp. 499–508.
- [22] S. XIANG, K. M. WANG, Y. T. AI, Y. D. SHA, AND H. SHI, *Trigonometric variable shape parameter and exponent strategy for generalized multiquadric radial basis function approximation*, *Appl. Math. Modelling*, 36 (2012), pp. 1931–1938.
- [23] F. YANG, L. YAN, AND L. LING, *Doubly stochastic radial basis function methods*, *J. Comput. Phys.*, 363 (2018), pp. 87–97.
- [24] K. ZYCZKOWSKI AND M. KUS, *Random unitary matrices*, *J. Phys. A: Math. General*, 27 (1994), p. 4235.