Accurate Emulators for Large-Scale Computer Experiments

Ben Haaland

Joint work with Peter Z. G. Qian and Xu Xu

Office of Clinical Sciences, Centre for Quantitative Medicine
Duke-NUS Graduate Medical School, Singapore
Department of Statistics and Applied Probability
National University of Singapore

Isaac Newton Institute, 9 September 2011

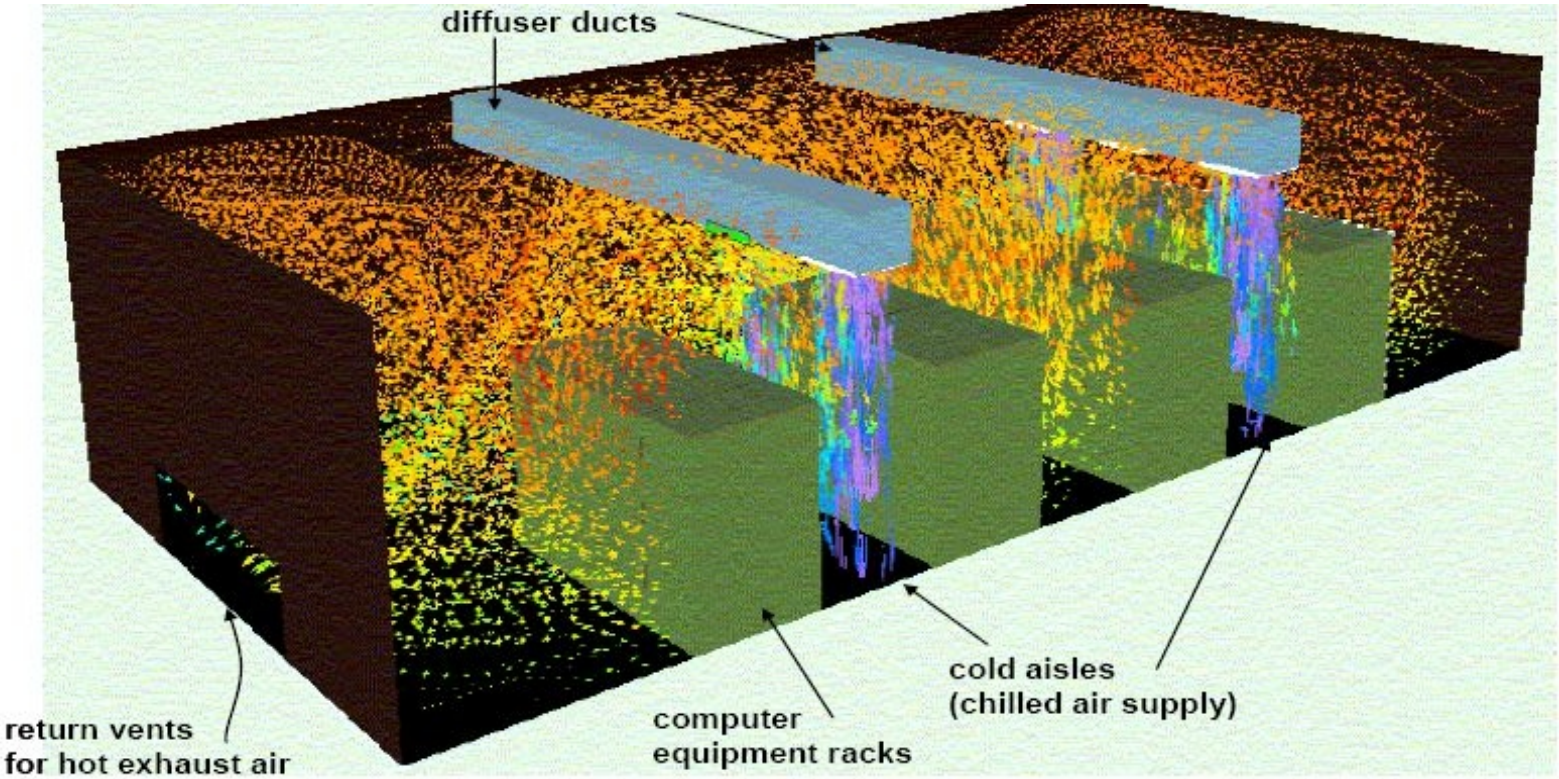
Outline

- Computer experiments
- Emulators for computer experiments
- Error decomposition
- Multi-step interpolator
- Error bounds

Computer Experiments

- **Computer Experiments** use complex mathematical models implemented in large computer codes to study real systems
- In many situations, it is not feasible to perform a physical experiment
- Unethical, impossible, expensive, inconvenient
- Often, the system of interest can be simulated using computer codes
- Input/output pairs can be produced with the help of computers
- Typically, input/output pairs are **deterministic** and **expensive**
- Examples: ecology, systems biology, climatology, engineering, marketing, computational biochemistry, data center efficiency

Computer Model of Data Center Heat Distribution



Based on computational fluid dynamics (CFD).
Implemented in Flotherm.
Each run takes hours to complete.

Courtesy of IBM T.J. Watson Research Center

Emulators for Computer Experiments

- When using a computer experiment to study a real system, a thorough exploration of the unknown surface is typically wanted
- Input/output pairs often too expensive for a complete exploration
- A solution is to evaluate the computer experiment at several well-distributed data sites
- Then, build an **interpolator** which can be used as a stand-in, or **emulator**, for the actual computer experiment
- Thorough exploration can then be carried out on the emulator

Gaussian Process Interpolators

- Gaussian process or reproducing kernel Hilbert space interpolators often used to build emulators for computer experiments
- Building block is a symmetric, positive definite basis function $\Phi : \Omega \times \Omega \to \mathbb{R}$
- **Simple Form:** $\mathcal{P}(x) = \sum_{i=1}^{n} \alpha_i \Phi(x, x_i)$, with $\mathcal{P}(x_i) = f(x_i)$
- Associated with each Φ (and Ω) is a Hilbert space of functions, $\mathcal{N}_\Phi(\Omega)$, whose norm measures size and smoothness
- **Control smoothness:** RKHS interpolator has smallest possible native space norm of any interpolator

Numerical Problems in Interpolation

- Many systems which scientists, engineers, and medical researchers use computer experiments to study exhibit extremely complex behavior in portions of the input space
- Understanding these regions requires many input sites which are potentially very close together
- Most interpolation procedures suffer from increasingly severe numeric problems as the number of data sites becomes larger
- The problem of finding an interpolator becomes ill-conditioned as data sites becomes too near to one another
- Many techniques for numerically stabilizing an interpolator: nugget effect, compactly supported kernels, correlation decomposition, multi-step interpolation

Numeric and Nominal Error

- A computed quantity, $\tilde{x}$, which is subject to floating point rounding, is distinguished from the idealized quantity, x
- The following separation of error is introduced

$$\|f - \tilde{\mathcal{P}}\| \leq \|f - \mathcal{P}\| + \|\mathcal{P} - \tilde{\mathcal{P}}\| \tag{1}$$

- The first and second terms on the right-hand side are called the *nominal* and *numeric* portions of the error, respectively
- The error separation (1) is true for any norm $\|\cdot\|$
- Often, there is a trade-off between the nominal and numeric error

"Emulation of large-scale computer experiments," Haaland, B. and Qian, P. Z. G. (2011), Under Revision for Annals of Statistics.

Nominal and Numeric Trade-Off

- If $f \in \mathcal{N}_\Phi(\Omega)$ and $X_1 \subseteq X_2$, then

$$\|f - \mathcal{P}_2\|_{\mathcal{N}_\Phi(\Omega)} \leq \|f - \mathcal{P}_1\|_{\mathcal{N}_\Phi(\Omega)},$$

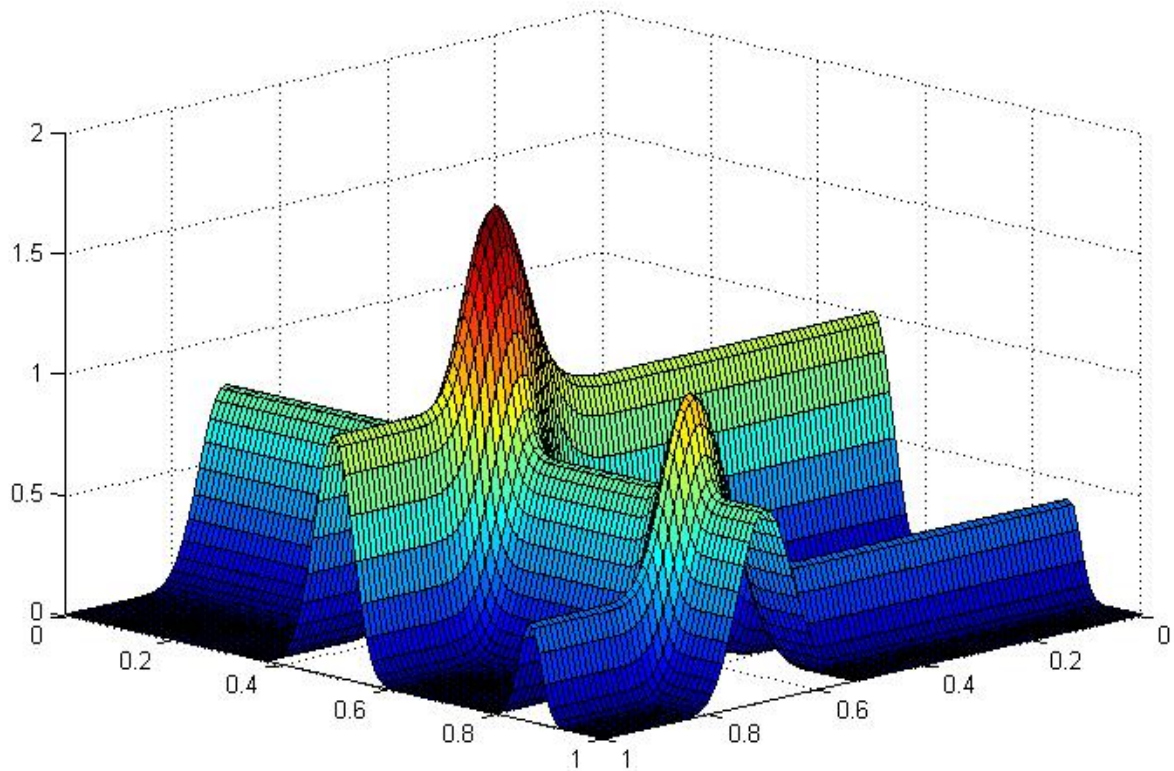
  where $\mathcal{P}_1$ and $\mathcal{P}_2$ denote the RKHS interpolators on the sets X_1 and X_2, respectively

- The nominal error of an interpolator is always reduced (with respect to the native space norm) by the addition of data sites

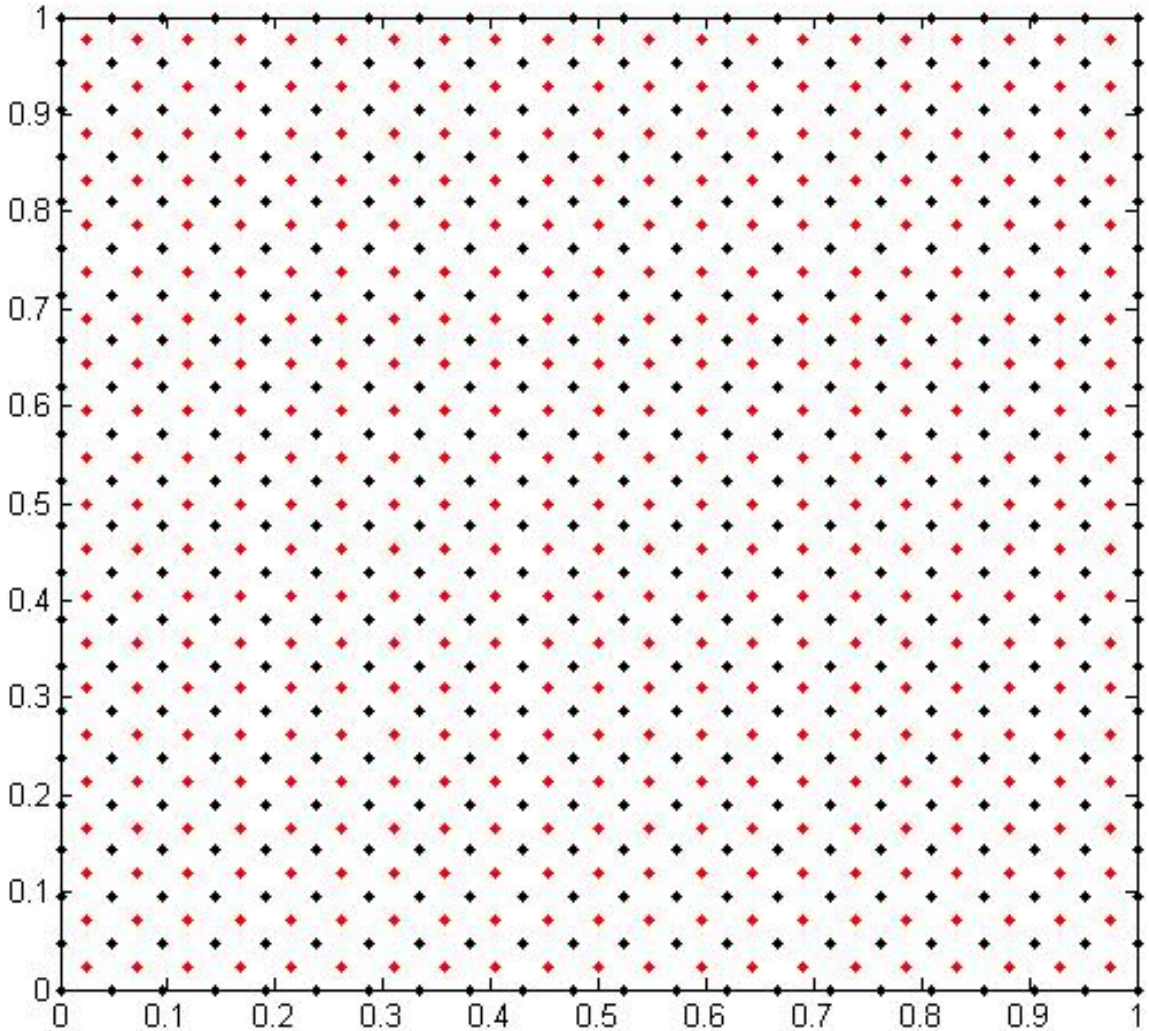
- On the other hand, the numeric error is potential increased to an arbitrary degree by the addition of data sites

Example

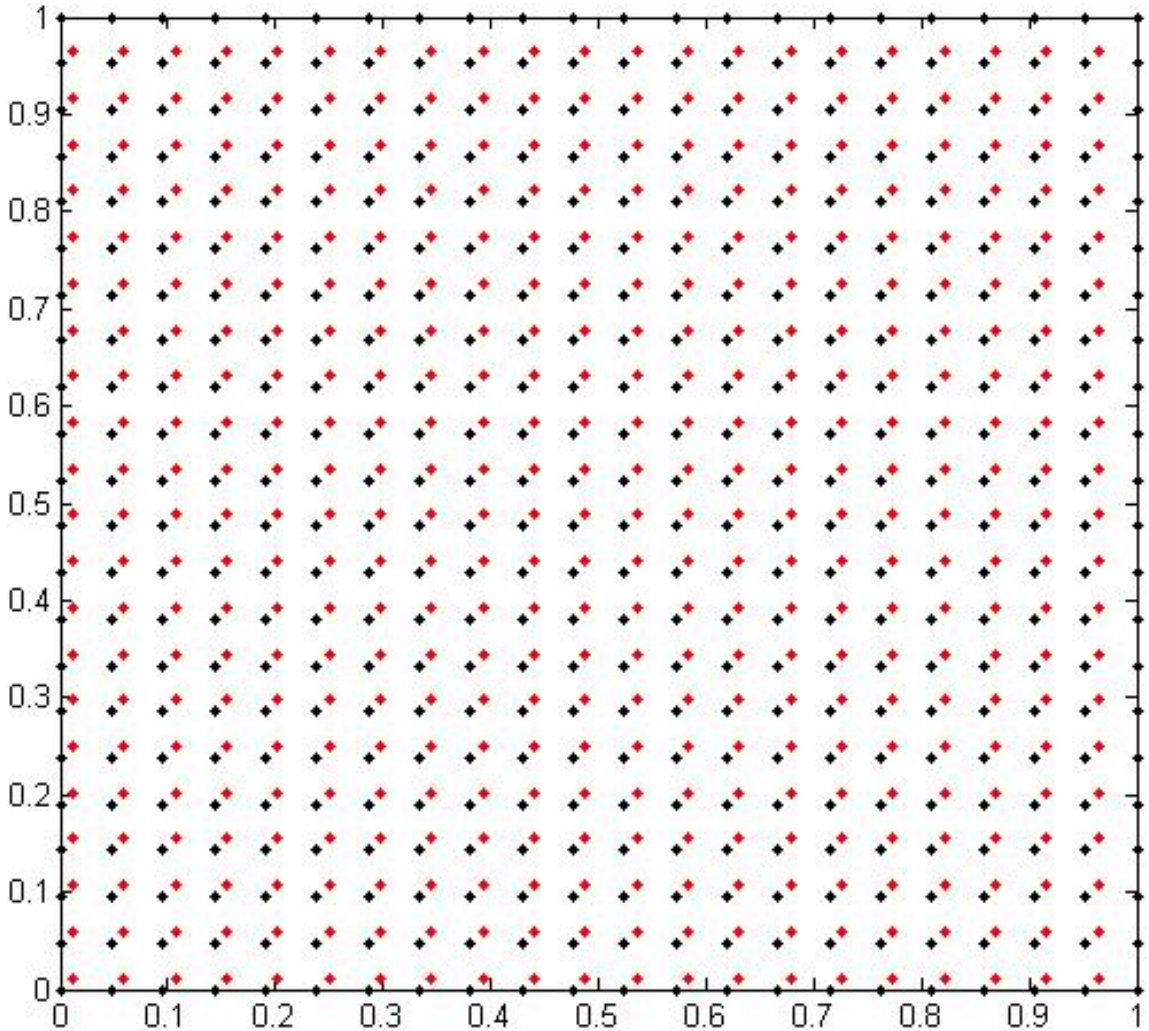
- Consider building an interpolator for the below Michaelewicz function on the third of the following 925 point input sets



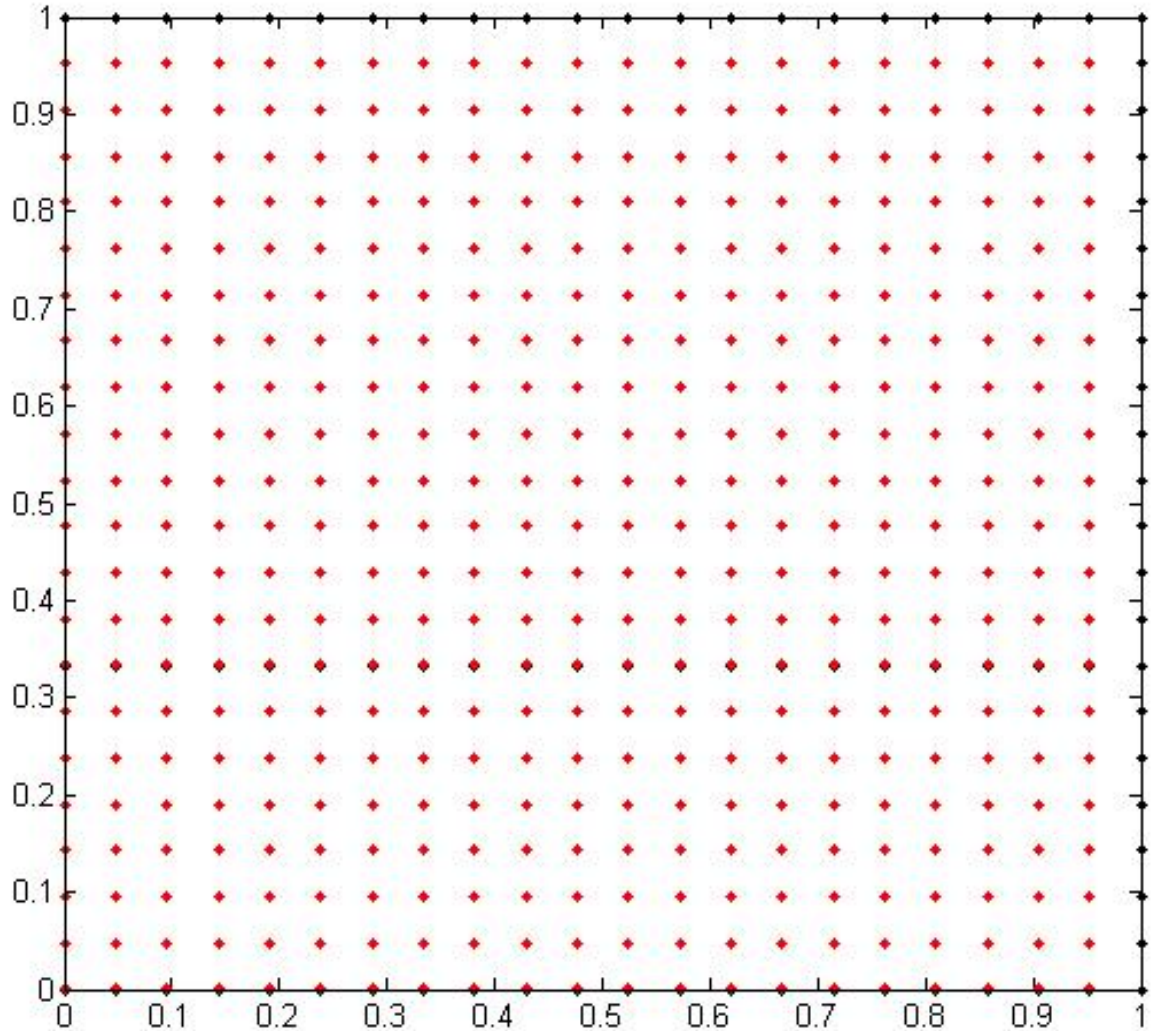
Point Set #1



Point Set #2



Point Set #3



Interpolator with Very Poor Accuracy

- The separation distance is 10^{-10}
- The red points do not contribute much information about the unknown surface, but make the interpolation problem much less stable
- The best possible mean squared error (over all widths of kernels) for traditional interpolation on the third point set is ≈ 0.15, the L_2 norm of the function

Multi-Step Interpolator (Floater and Iske 1996)

- Form well-spread nested subsets of the data
- Interpolate the first subset with a wide and smooth kernel and form residuals on the next subset
- The residuals are then interpolated using a narrower and less smooth kernel
- The residual interpolator is added to the previous stage interpolator to form the current stage interpolator
- Residuals are successively formed and interpolated with narrower and less smooth kernels
- The final interpolator is the sum of the interpolators formed at each stage
- A general theory is developed in Haaland and Qian (2011)

Illustration of Multi-Step Algorithm

- Consider using the multi-level algorithm to interpolate the below function using data at the plotted sites

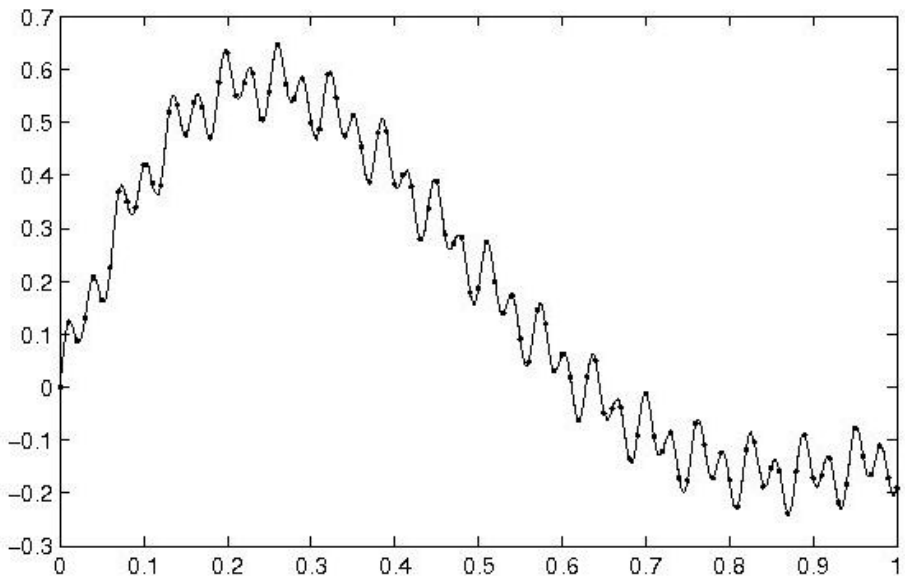


Figure: True function and data sites

Stage 1

- Interpolate at every other point using a wide kernel

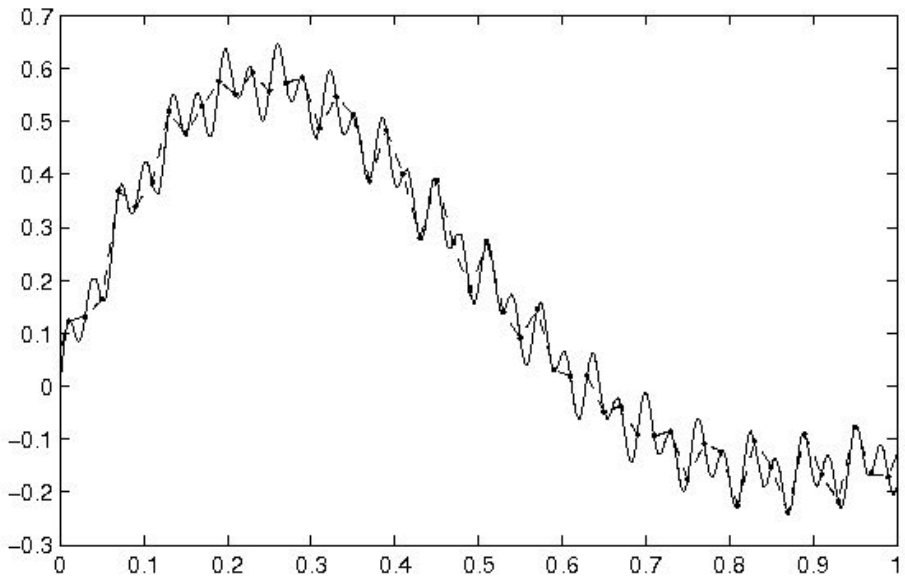


Figure: True function (solid) and first-stage interpolator (dashed)

Stage 2

- Interpolate residuals with narrow kernel

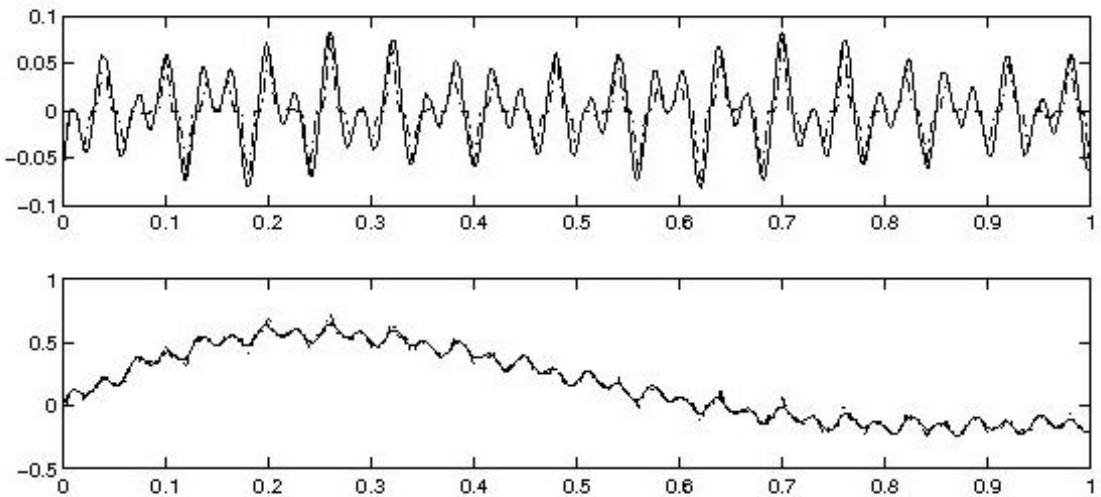


Figure: **Top Panel:** Residuals (solid) and residual interpolator (dashed)
Bottom Panel: True function (solid) and multi-scale interpolator (dashed)

Multi-Step Interpolator Details

- Here, J denotes the number of stages, Φ_j denotes the kernel used in stage j, and $X_1 \subset \cdots \subset X_J = X$. Initialize $\mathcal{P}^0 \equiv 0$
- Then, for $j = 1, \ldots, J$, let

$$\mathcal{P}^j(x) = \sum_{u=1}^{n_j} \alpha_u^j \Phi_j(x - x_u),$$

$$\alpha^j = A_{X_j,\Phi_j}^{-1}(f - \sum_{k=0}^{j-1} \mathcal{P}^k)|_{X_j},$$

$$A_{X_j,\Phi_j} = \{\Phi_j(x_u - x_v)\},\ u, v = 1, \ldots, n_j,$$

$$n_j = \text{card } X_j$$

- Then, the multi-scale interpolator is

$$\mathcal{P}(x) = \sum_{j=1}^{J} \mathcal{P}^j(x)$$

Numeric Accuracy Results

- The only complete and rigorous bound on the numeric error of the multi-step interpolator
- The results make very weak and general assumptions
- Approximations to the terms in the bound can be computed

Numeric Lemma

- The matrix 2-norm $\|\cdot\|_2$ is defined as $\|A\|_2 = \sqrt{\lambda_{\max}(A'A)}$

- Suppose $Ax = b$ and $\tilde{A}\tilde{x} = \tilde{b}$ with $\|A - \tilde{A}\|_2 \leq \delta_A \|A\|_2$, $\|b - \tilde{b}\|_2 \leq \delta_b \|b\|_2$, and $\kappa(A) = r/\delta_A < 1/\delta_A$ for some $\delta_A, \delta_b > 0$. Then, $\tilde{A}$ is non-singular,

$$\frac{\|\tilde{x}\|_2}{\|x\|_2} \leq \frac{1 + r(\delta_b/\delta_A)}{1 - r},$$
$$\frac{\|x - \tilde{x}\|_2}{\|x\|_2} \leq \frac{\delta_A + \delta_b}{1 - r}\kappa(A),$$

where $\kappa(A) = \|A\|_2 \|A^{-1}\|_2$

Numeric Accuracy of Single Stage Interpolator

- Suppose that $\|A_{X,\Phi} - \tilde{A}_{X,\Phi}\|_2 \leq \delta_A \|A_{X,\Phi}\|_2$, $\|f|_X - \tilde{f}|_X\|_2 \leq \delta_f \|f|_X\|_2$, $\kappa(A_{X,\Phi}) = r/\delta_A < 1/\delta_A$, and $\sup_{x,y\in\Omega} |\Phi(x-y) - \tilde{\Phi}(x-y)| < D\delta_A$ for some $\delta_A, \delta_f, D > 0$, then

$$|\mathcal{P}(x) - \tilde{\mathcal{P}}(x)| \leq \|f|_X/\sqrt{n}\|_2 \frac{(\delta_A + \delta_f)}{1-r} g(X,\Phi),$$

$$g(X,\Phi) = \frac{n}{\lambda_{\min}(A_{X,\Phi})}(\kappa(A_{X,\Phi})\Phi(0) + D),$$

  where $\kappa(\cdot)$ is defined in the previous lemma

- To bound error of single stage interpolator, take $\delta_A = \delta_f = 10^{-15}$, restrain $\kappa(A_{X,\Phi})$

Numeric Accuracy of Multi-Step Interpolator

- Suppose that for $j = 1, \ldots, J$, $\|A_{X_j,\Phi_j} - \tilde{A}_{X_j,\Phi_j}\|_2 \leq \delta_j \|A_{X_j,\Phi_j}\|_2$, $\|f|_{X_j} - \tilde{f}|_{X_j}\|_2 \leq \delta \|f|_{X_j}\|_2$, $\kappa(A_{X_j,\Phi_j}) \leq r/\delta_j < 1/\delta_j$, and $\sup_{x,y\in\Omega} |\Phi_j(x-y) - \tilde{\Phi}_j(x-y)| < D\delta$ for some $\delta_j, \delta, D > 0$ with $\delta_j \|(f - \sum_{k=1}^{j-1} \mathcal{P}^k)|_{X_j}/\sqrt{n_j}\|_2 \leq \delta \|f|_{X_j}/\sqrt{n_j}\|_2$, then

$$|\sum_{j=1}^{J} \mathcal{P}^j(x) - \sum_{j=1}^{J} \tilde{\mathcal{P}}^j(x)|$$

$$\leq \delta \|f|_{X_J}/\sqrt{n_J}\|_2 \left[\sum_{M=1}^{J} C^M \sum_{i\in\mathcal{S}_J(M)} \prod_{k=1}^{M} \rho(X_{i_k}, X_{i_{k+1}}) g(X_{i_k}, \Phi_{i_k}) \right],$$

where $C = 2/(1-r)$,
$\mathcal{S}_J(M) = \{i \in \mathbb{N}^{M+1} : 1 \leq i_1 < \cdots < i_M \leq i_{M+1} = J\}$
$\rho(X, Y) = \|f|_X/\sqrt{n_X}\|_2 / \|f|_Y/\sqrt{n_Y}\|_2$, and g is defined in the previous theorem

Bound in Terms of Separation Distance

- Let

$$q_X = \frac{1}{2} \min_{x_i, x_j \in X} \|x_i - x_j\|_2$$

- For $f \in L_1(\mathbb{R}^d)$ define the Fourier transform

$$\hat{f}(\omega) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i\omega' x} \mathrm{d}x$$

- Let $\varphi_*(M, \Phi) = \inf_{\|\omega\|_2 \leq 2M} \hat{\Phi}(\omega)$
- Then, under the assumptions of the previous theorems

$$g(X_j, \Phi_j) \leq \kappa_{\text{upper}}(X_j, \Phi_j)(\kappa_{\text{upper}}(X_j, \Phi_j)\Phi(0) + D),$$

$$\kappa_{\text{upper}}(X_j, \Phi_j) = \frac{n_j q_{X_j}^d}{C_d \varphi_*(M_d / q_{X_j}, \Phi_j)}$$

Point-Wise Nominal Bound

- For a nonsingular Θ, define $\Phi_\Theta(x) = \Phi(\Theta x)$
- Let h_X denote the *fill distance*

$$h_X = \sup_{x \in \Omega} \min_{x_u \in X} \|x - x_u\|_2$$

- Suppose that Ω is bounded and convex, Φ is continuous, positive definite, and has k continuous derivatives, and Θ is non-singular. Then,

$$|f(x) - \mathcal{P}(x)| \leq C_\Phi \|\Theta\|_2^{k/2} h_X^{k/2} \|f\|_{\mathcal{N}_{\Phi_\Theta}(\Omega)}$$

Native Space Error Bound

- Under the assumptions of the previous theorem,

$$\|f - \mathcal{P}\|_{\mathcal{N}_{\Phi_\Theta}(\Omega)} \leq C_\Phi \|\Theta\|_2^{k/2} h_X^{k/2} \|f\|_{\mathcal{N}_{\Phi_\Theta * \Phi_\Theta}(\Omega)}$$

- To allow for different re-scalings at each stage, some additional development and assumptions are necessary

Notation and Assumptions

- Define Ψ_k recursively as

$$\psi^0 = \Phi,$$
$$\psi^k = \psi^{k-1} * \psi^{k-1},$$

  for $k \in \mathbb{N}$

- For the kernel on step j, take

$$\Phi_j = \Psi^{J-j}_{\Theta_j}$$

- Take $c_2 \geq c_1 > 0$ and $\hat{\Upsilon}$ with

$$\omega'\omega \leq \nu'\nu \Longrightarrow \hat{\Upsilon}(\omega) \geq \hat{\Upsilon}(\nu),$$
$$c_1\hat{\Upsilon}(\omega) \leq \hat{\Phi}(\omega) \leq c_2\hat{\Upsilon}(\omega)$$

Norm Relation

- If Φ is continuous and positive definite, the above assumptions are satisfied, and Θ_{j-1}, Θ_j are non-singular with respective inverses Ξ'_{j-1}, Ξ'_j, then

$$\lambda_{\max}(\Theta'_{j-1}\Theta_{j-1}\Xi'_j\Xi_j) \leq 1$$

$$\Longrightarrow \|f\|^2_{\mathcal{N}_{\Phi_j * \Phi_j}(\Omega)} \leq \left(\frac{c_2}{c_1}\right)^{2^{J-(j-1)}} \frac{|\det(\Xi_{j-1})|}{|\det(\Xi_j)|^2} \|f\|^2_{\mathcal{N}_{\Phi_{j-1}}(\Omega)}$$

for $1 \leq j \leq J$

Nominal Error for Multi-Step Interpolator

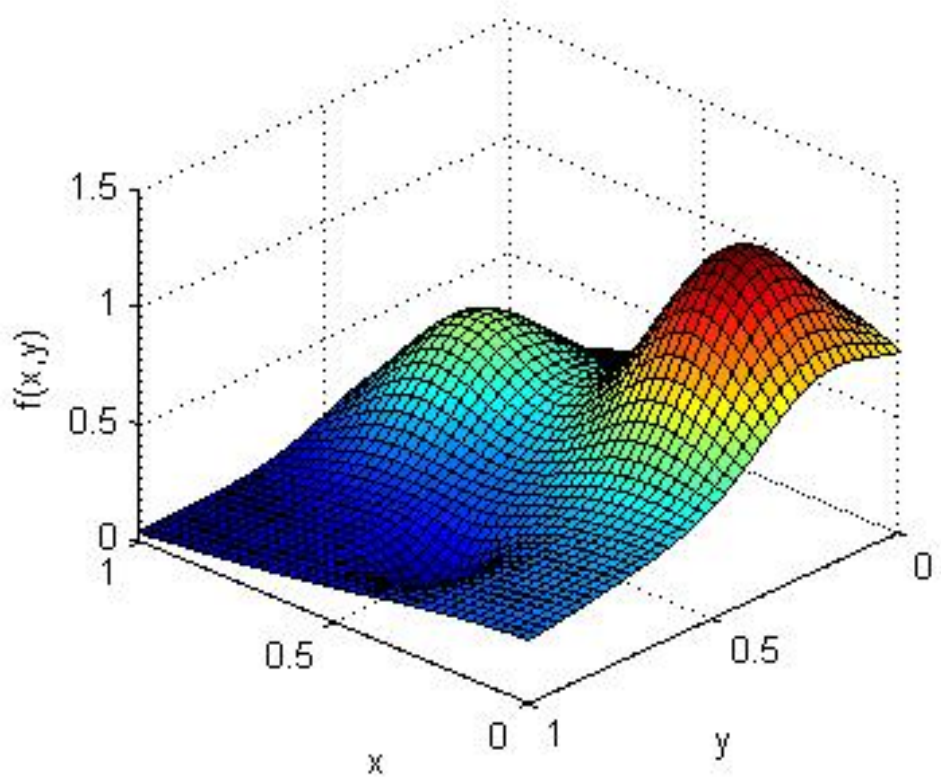
- Under the nominal error assumptions,

$$|f(x) - \sum_{j=1}^{J} \mathcal{P}^j(x)|$$

$$\leq C_{\Phi,J} \|f\|_{\mathcal{N}_{\Phi_0}(\Omega)} \|\Theta_J\|_2^{k/2} h_{X_J}^{k/2} \prod_{j=1}^{J} \left\{ \frac{\sqrt{|\det(\Xi_{j-1})|}}{|\det(\Xi_j)|} \left(\|\Theta_j\|_2^k h_{X_j}^k \right)^{2^{J-j-1}} \right\}$$

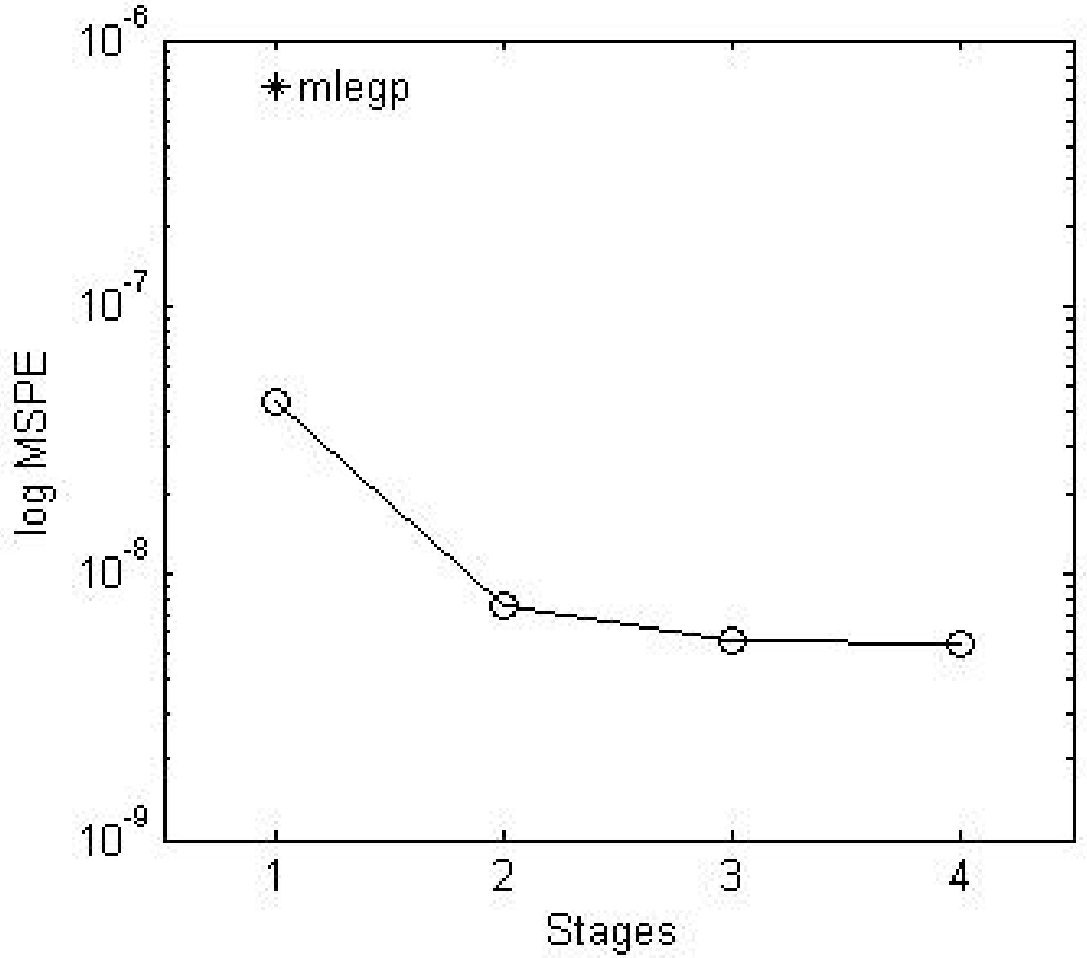
Example

- Consider using the multi-step procedure to interpolate Franke's function
- Nested sequences of data sites generated using randomized (0, 4, 2)-net in base 5, 625 points



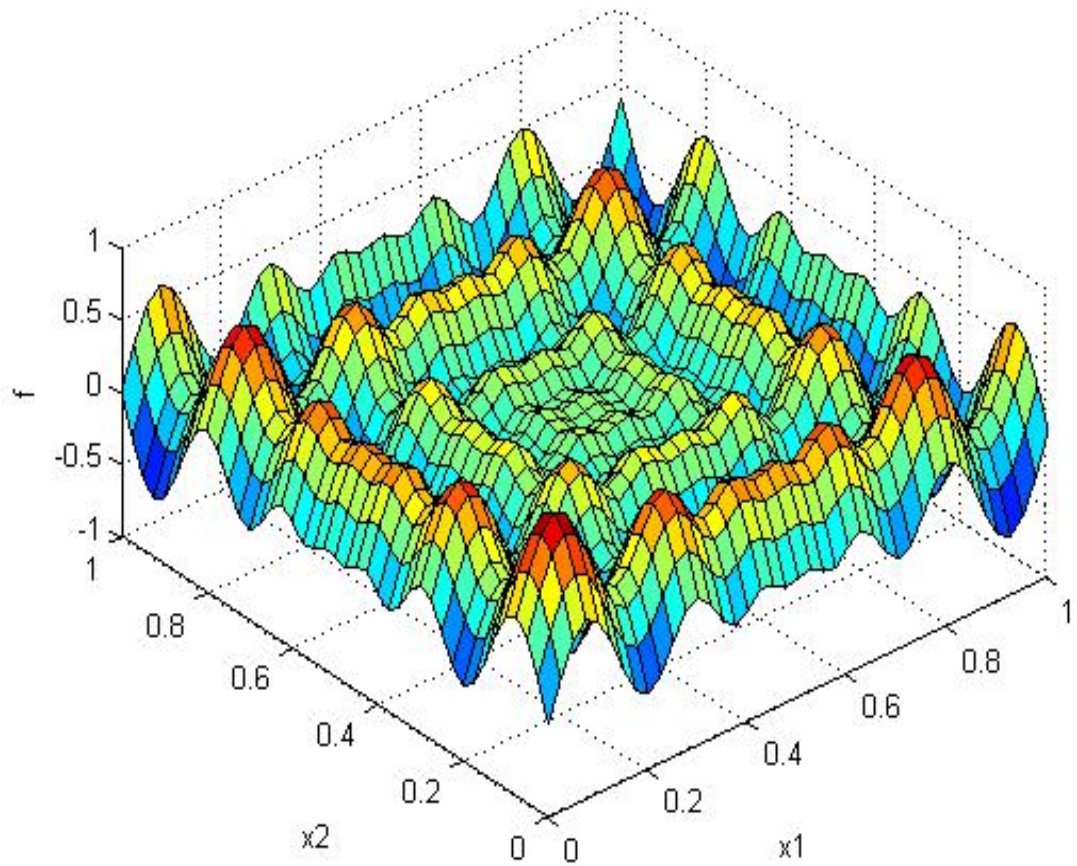
Example (continued)

- log MSPEs vs. number of stages
- Single stage corresponds to usual interpolation



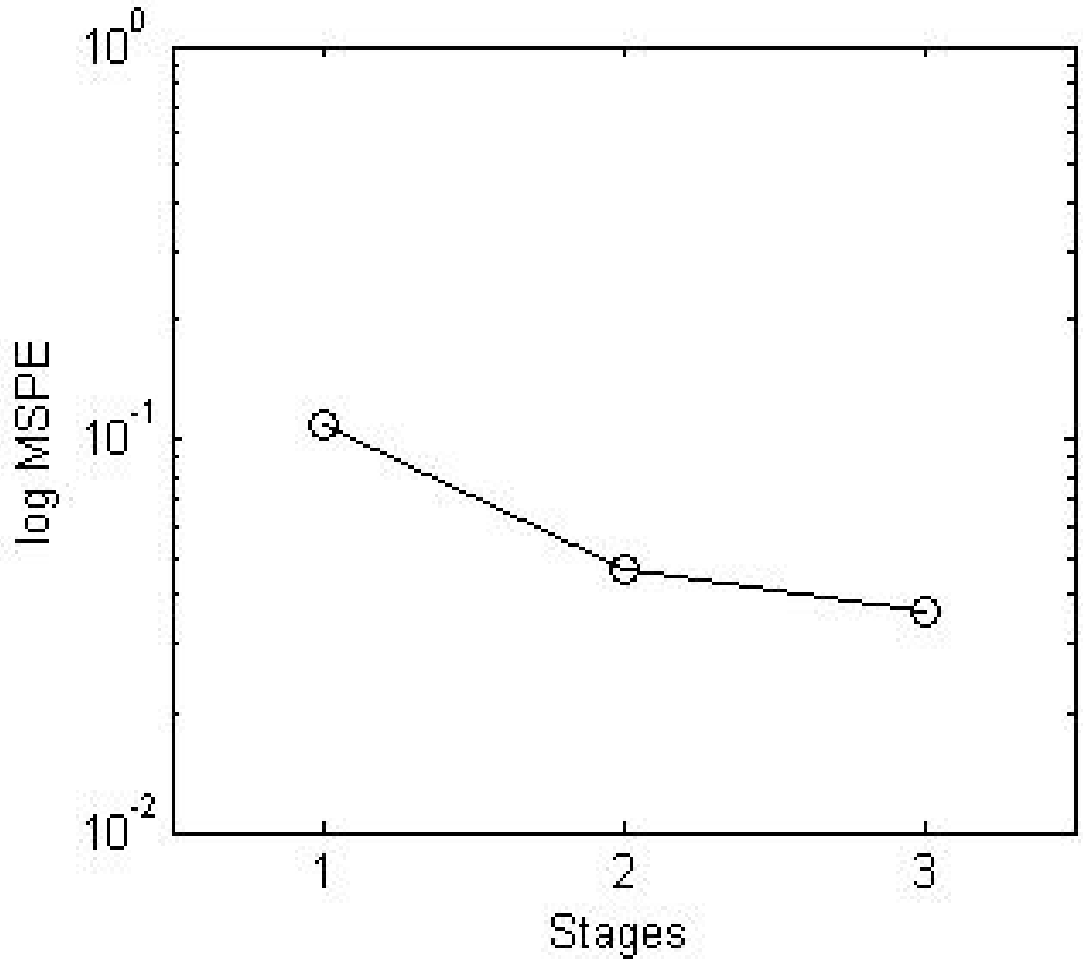
Another Example

- Consider using the multi-step procedure to interpolate Schwefel's function for $d = 5$
- Data are collected from Schwefel's function using a randomized $(0, 8, 5)$-net in base 5 with $5^8 = 390625$ points



Another Example (continued)

- log MSPEs vs. number of stages
- Single stage corresponds to usual interpolation



Confidence Intervals

- If the error at stage j, $f - \sum_{k=0}^{j-1} \mathcal{P}^k$, is modeled as a GP, then confidence intervals on the function's values $f(x)$ can be obtained in much the same manner as a single stage interpolator
- In particular, model the output as

$$f(x) = \sum_{j=1}^{J} Z_j(x),$$

  where the Z_j are mean zero Gaussian processes with $\mathrm{Cov}(Z_j(x_1), Z_j(x_2)) = \sigma_j^2 \Phi_j(x_1 - x_2)$
- Note that the Z_j are *not* independent
- Details in Haaland and Qian (2011)

Conclusions

- Multi-scale procedure can be used to improve the accuracy emulators
- Numeric accuracy requires well separated data points
- Nominal accuracy requires small data-free regions
- Confidence envelopes under Gaussian process assumptions