Function Approximation When Function Values Are Expensive

Fred J. Hickernell[1] Simon Mak[2]

[1]Department of Applied Mathematics, Illinois Institute of Technology, hickernell@iit.edu

[2]School of Industrial and System Engineering, Georgia Institute of Technology

Supported by NSF-DMS-1522687 and DMS-1638521 (SAMSI)

SAMSI-QMC Transition Workshop, May 7, 2018

Thanks to . . .

- SAMSI for sponsoring this program in Quasi-Monte Carlo Methods and High Dimensional Sampling for Applied Mathematics
  - Especially Ilse Ipsen, Karem Jackson, Sue McDonald, Rick Scoggins, Thomas Gehrmann, Richard Smith, and David Banks

Thanks to ...

- SAMSI for sponsoring this program in Quasi-Monte Carlo Methods and High Dimensional Sampling for Applied Mathematics
  - Especially Ilse Ipsen, Karem Jackson, Sue McDonald, Rick Scoggins, Thomas Gehrmann, Richard Smith, and David Banks
- Frances Kuo, Pierre L'Ecuyer, and Art Owen, fellow program leaders
  - Especially Art, who keeps trying to push QMC out of its box

Thanks to . . .

- SAMSI for sponsoring this program in Quasi-Monte Carlo Methods and High Dimensional Sampling for Applied Mathematics
  - Especially Ilse Ipsen, Karem Jackson, Sue McDonald, Rick Scoggins, Thomas Gehrmann, Richard Smith, and David Banks
- Frances Kuo, Pierre L'Ecuyer, and Art Owen, fellow program leaders
  - Especially Art, who keeps trying to push QMC out of its box
- Many of you with whom I have had fruiful discussions, especially WGs 2, 4, and 5

Thanks to . . .

- SAMSI for sponsoring this program in Quasi-Monte Carlo Methods and High Dimensional Sampling for Applied Mathematics
  - Especially Ilse Ipsen, Karem Jackson, Sue McDonald, Rick Scoggins, Thomas Gehrmann, Richard Smith, and David Banks
- Frances Kuo, Pierre L'Ecuyer, and Art Owen, fellow program leaders
  - Especially Art, who keeps trying to push QMC out of its box
- Many of you with whom I have had fruiful discussions, especially WGs 2, 4, and 5
- Mac Hyman, who
  - Promoted QMC to SAMSI behind the scenes
  - Introduced me to problems with expensive function values
  - Co-led WG 5

Thanks to . . .

- SAMSI for sponsoring this program in Quasi-Monte Carlo Methods and High Dimensional Sampling for Applied Mathematics
  - Especially Ilse Ipsen, Karem Jackson, Sue McDonald, Rick Scoggins, Thomas Gehrmann, Richard Smith, and David Banks
- Frances Kuo, Pierre L'Ecuyer, and Art Owen, fellow program leaders
  - Especially Art, who keeps trying to push QMC out of its box
- Many of you with whom I have had fruiful discussions, especially WGs 2, 4, and 5
- Mac Hyman, who
  - Promoted QMC to SAMSI behind the scenes
  - Introduced me to problems with expensive function values
  - Co-led WG 5
- Henryk Woźniakowski, whose work on tractability has inspired some of what I will say

Thanks to . . .

- SAMSI for sponsoring this program in Quasi-Monte Carlo Methods and High Dimensional Sampling for Applied Mathematics
    - Especially Ilse Ipsen, Karem Jackson, Sue McDonald, Rick Scoggins, Thomas Gehrmann, Richard Smith, and David Banks
- Frances Kuo, Pierre L'Ecuyer, and Art Owen, fellow program leaders
    - Especially Art, who keeps trying to push QMC out of its box
- Many of you with whom I have had fruiful discussions, especially WGs 2, 4, and 5
- Mac Hyman, who
    - Promoted QMC to SAMSI behind the scenes
    - Introduced me to problems with expensive function values
    - Co-led WG 5
- Henryk Woźniakowski, whose work on tractability has inspired some of what I will say
- Kai-Tai Fang, who introduced me to experimental design

Approximating Functions When Function Values Are Expensive

- Interested in $f : [-1,1]^d \to \mathbb{R}$, e.g., the result of a climate model, or a financial calculation
- d is dozens or a few hundred
- $\$(f) =$ cost to evaluate $f(\boldsymbol{x})$ for any $\boldsymbol{x} \in [-1,1]^d =$ hours or days or \$1M
- Want to construct a surrogate model, $f_{\text{app}} \approx f$, with $\$(f_{\text{app}}) = \$0.000001$ so that we may quickly explore (plot, integrate, optimize, search for sharp gradients of) f
- f_{app} is constructed using n pieces of information about f
- Want $\left\lVert f - f_{\text{app}} \right\rVert_\infty \leqslant \varepsilon$ for $n = \mathcal{O}(d^p \varepsilon^{-q})$ as $d \uparrow \infty$ or $\varepsilon \downarrow 0$ (with small p and q)
- Assume $\$(f) \gg n^r$ for any practical n and any positive r, so the cost of the algorithm is $\mathcal{O}(\$(f)n)$

Functions Expressed at Series

Let $f : [-1,1]^d \to \mathbb{R}$ have $L^2([-1,1]^d, \varrho)$ an orthogonal series expansion:

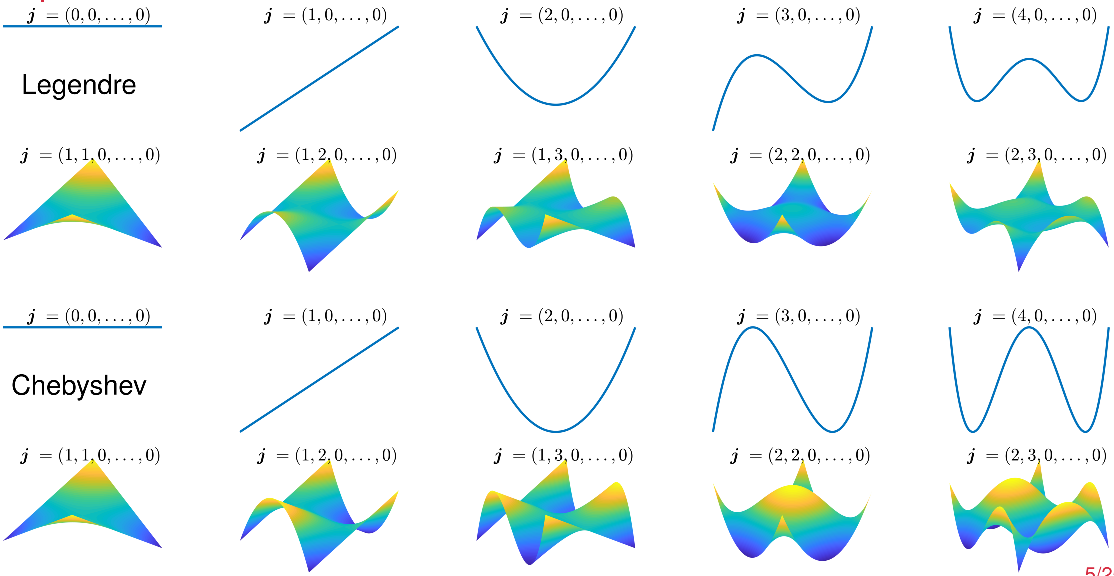
$$f(\mathbf{x}) = \sum_{\mathbf{j} \in \mathbb{N}_0^d} \widehat{f}(\mathbf{j}) \phi_{\mathbf{j}}(\mathbf{x}), \qquad \phi_{\mathbf{j}}(\mathbf{x}) = \phi_{j_1}(x_1) \cdots \phi_{j_d}(x_d), \qquad \left\| \phi_{\mathbf{j}} \right\|_\infty = 1$$

$$\widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}} \rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}} \rangle}, \qquad \langle f, g \rangle := \int_{[-1,1]^d} f(\mathbf{x}) g(\mathbf{x})\, \varrho(\mathbf{x})\, \mathrm{d}\mathbf{x}$$

Legendre polynomials: $\displaystyle\int_{-1}^{1} \phi_j(x) \phi_k(x)\, \mathrm{d}x = c_j \delta_{j,k}$

Chebyshev polynomials: $\phi_j(x) = \cos(j \arccos(x))$, $\displaystyle\int_{-1}^{1} \frac{\phi_j(x)\phi_k(x)}{\sqrt{1-x^2}}\, \mathrm{d}x = c_j \delta_{j,k}$

Example Bases



Approximation by Series Coefficients

$$f(\mathbf{x}) = \sum_{\mathbf{j} \in \mathbb{N}_0^d} \widehat{f}(\mathbf{j}) \phi_{\mathbf{j}}(\mathbf{x}), \qquad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}} \rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}} \rangle}, \qquad \| \phi_{\mathbf{j}} \|_\infty = 1$$

Suppose that we may observe the series coefficients $\widehat{f}(\mathbf{j})$ at a cost of \$1M each. (Eventually we want to consider the case of observing function values.) For any vector of non-negative constants, $\boldsymbol{\gamma} = (\gamma_{\mathbf{j}})_{\mathbf{j} \in \mathbb{N}_0^d}$, define the norm

$$\| \widehat{f} \|_{q,\boldsymbol{\gamma}} := \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j} \in \mathbb{N}_0^d} \right\|_q, \qquad 0/0 = 0, \quad \gamma_{\mathbf{j}} = 0 \;\&\; \| \widehat{f} \|_{\infty,\boldsymbol{\gamma}} < \infty \implies \widehat{f}(\mathbf{j}) = 0$$

Order the wavenumbers $\mathbf{j}$ such that $\gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots$. The optimal approximation why? to f given the choice of n series coefficients is

$$f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^n \widehat{f}(\mathbf{j}_i) \phi_{\mathbf{j}_i}, \quad \| f - f_{\text{app}} \|_\infty = \left\| \sum_{i=n+1}^\infty \widehat{f}(\mathbf{j}_i) \phi_{\mathbf{j}_i} \right\|_\infty \overset{\text{loose}}{\leqslant} \| \widehat{f} - \widehat{f}_{\text{app}} \|_1 \underset{\text{optimal}}{\overset{\text{tight}}{\leqslant}} \| \widehat{f} \|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^\infty \gamma_{\mathbf{j}_i}$$

How Quickly Does Error Decay?

$$f(\boldsymbol{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\boldsymbol{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

$$\gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \quad f_{\text{app}}(\boldsymbol{x}) = \sum_{i=1}^n \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i}, \quad \|f - f_{\text{app}}\|_\infty \overset{\text{loose}}{\leqslant} \|\widehat{f} - \widehat{f}_{\text{app}}\|_1 \underset{\text{optimal}}{\overset{\text{tight}}{\leqslant}} \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^\infty \gamma_{\mathbf{j}_i}$$

An often used trick $(q > 0)$:

$$\gamma_{\mathbf{j}_{n+1}} \leqslant \left[\frac{1}{n}\left(\gamma_{\mathbf{j}_1}^{1/q} + \cdots + \gamma_{\mathbf{j}_n}^{1/q}\right) \right]^q \leqslant \frac{1}{n^q}\|\boldsymbol{\gamma}\|_{1/q}, \qquad \|\boldsymbol{\gamma}\|_{1/q} = \left[\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}}^{1/q} \right]^q$$

$$\sum_{i=n+1}^\infty \gamma_{\mathbf{j}_i} \leqslant \|\boldsymbol{\gamma}\|_{1/q} \sum_{i=n}^\infty \frac{1}{i^q} \leqslant \frac{\|\boldsymbol{\gamma}\|_{1/q}}{(q-1)(n-1)^{q-1}}$$

rate controlled by finiteness of $\|\boldsymbol{\gamma}\|_{1/q}$

Recap

$$f(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

dependence of f on d is hidden $\qquad \gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \qquad f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i},$

$$\|f - f_{\text{app}}\|_\infty \leqslant \|\widehat{f} - \widehat{f}_{\text{app}}\|_1 \leqslant \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i} \leqslant \frac{\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_{1/q}}{(q-1)(n-1)^{q-1}} \overset{\text{Want}}{\leqslant} \varepsilon$$

$$n = \mathcal{O}\left(\left[\frac{\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_{1/q}}{\varepsilon} \right]^{1/(q-1)} \right) \text{ is sufficient}$$

To succeed with $n = \mathcal{O}(d^p)$[1], we need $\|\boldsymbol{\gamma}\|_{1/q} = \mathcal{O}(d^{p'})$

[1] Novak, E. & Woźniakowski, H. *Tractability of Multivariate Problems Volume I: Linear Information. EMS Tracts in Mathematics* **6** (European Mathematical Society, Zürich, 2008), Kühn, T. *et al.* Approximation numbers of Sobolev embeddings—Sharp constants and tractability. *J. Complexity* **30,** 95–116 (2014).

Recap

$$f(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \left\|\phi_{\mathbf{j}}\right\|_\infty = 1, \quad \left\|\widehat{f}\right\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

dependence of f on d is hidden $\qquad \gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \qquad f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i},$

$$\left\|f - f_{\text{app}}\right\|_\infty \leqslant \left\|\widehat{f} - \widehat{f}_{\text{app}}\right\|_1 \leqslant \left\|\widehat{f}\right\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i} \leqslant \frac{\left\|\widehat{f}\right\|_{\infty,\boldsymbol{\gamma}} \left\|\boldsymbol{\gamma}\right\|_{1/q}}{(q-1)(n-1)^{q-1}} \overset{\text{Want}}{\leqslant} \varepsilon$$

$$\left\|\boldsymbol{\gamma}\right\|_{1/q} = \mathcal{O}(d^{p'}) \implies n = \mathcal{O}(d^{p})$$

What remains?

Recap

$$f(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

dependence of f on d is hidden $\qquad \gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \qquad f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i},$

$$\|f - f_{\text{app}}\|_\infty \leqslant \|\widehat{f} - \widehat{f}_{\text{app}}\|_1 \leqslant \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i} \leqslant \frac{\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_{1/q}}{(q-1)(n-1)^{q-1}} \overset{\text{Want}}{\leqslant} \varepsilon$$

$$\|\boldsymbol{\gamma}\|_{1/q} = \mathcal{O}(d^{p'}) \implies n = \mathcal{O}(d^{p})$$

What remains?

- How do we infer γ in practice? Tradition fixes something convenient.

Recap

$$f(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

dependence of f on d is hidden $\qquad \gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \qquad f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i},$

$$\|f - f_{\text{app}}\|_\infty \leqslant \|\widehat{f} - \widehat{f}_{\text{app}}\|_1 \leqslant \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i} \leqslant \frac{\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_{1/q}}{(q-1)(n-1)^{q-1}} \overset{\text{Want}}{\leqslant} \varepsilon$$

$$\|\boldsymbol{\gamma}\|_{1/q} = \mathcal{O}(d^{p'}) \implies n = \mathcal{O}(d^p)$$

What remains?

- How do we infer $\boldsymbol{\gamma}$ in practice? Tradition fixes something convenient.
- How do we infer a bound on $\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}$?

Recap

$$f(\mathbf{x}) = \sum_{\mathbf{j} \in \mathbb{N}_0^d} \widehat{f}(\mathbf{j}) \phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}} \rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}} \rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j} \in \mathbb{N}_0^d} \right\|_q$$

dependence of f on d is hidden $\quad \gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \quad f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i) \phi_{\mathbf{j}_i},$

$$\|f - f_{\text{app}}\|_\infty \leqslant \|\widehat{f} - \widehat{f}_{\text{app}}\|_1 \leqslant \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i} \leqslant \frac{\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_{1/q}}{(q-1)(n-1)^{q-1}} \overset{\text{Want}}{\leqslant} \varepsilon$$

$$\|\boldsymbol{\gamma}\|_{1/q} = \mathcal{O}(d^{p'}) \implies n = \mathcal{O}(d^p)$$

What remains?

- How do we infer $\boldsymbol{\gamma}$ in practice? Tradition fixes something convenient.
- How do we infer a bound on $\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}$?
- How do we approximate using function values, not series coefficients?

Recap

$$f(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

dependence of f on d is hidden $\qquad \gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \qquad f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i},$

$$\|f - f_{\text{app}}\|_\infty \leqslant \|\widehat{f} - \widehat{f}_{\text{app}}\|_1 \leqslant \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i} \leqslant \frac{\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_{1/q}}{(q-1)(n-1)^{q-1}} \overset{\text{Want}}{\leqslant} \varepsilon$$

$$\|\boldsymbol{\gamma}\|_{1/q} = \mathcal{O}(d^{p'}) \implies n = \mathcal{O}(d^{p})$$

What remains? Assume that the function is nice enough to allow this inference.

- How do we infer $\boldsymbol{\gamma}$ in practice? Tradition fixes something convenient.
- How do we infer a bound on $\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}$?
- How do we approximate using function values, not series coefficients?

Main New Ideas

It is assumed that the f is nice enough to justify the following:

Inferring γ Assume a structure informed by experimental design principles. Infer coordinate importance from a pilot sample with wavenumbers

$$\begin{aligned}\mathcal{J} &:= \{(0,\ldots,0,j,0\ldots,0) : j = 0,\ldots,n_0\}\\ &= \{j\boldsymbol{e}_k : j = 0,\ldots,n_0,\ k = 1,\ldots,d\}\end{aligned}$$

Inferring $\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}$ Iteratively add wavenumber with largest $\gamma_{\mathbf{j}}$ to $\mathcal{J}$. Inflate the norm that is observed so far and assume

$$\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \leqslant C\left\|\left(\widehat{f}_{\mathbf{j}}\right)_{\mathbf{j}\in\mathcal{J}}\right\|_{\infty,\boldsymbol{\gamma}}$$

Function values Let the new wavenumber, $\mathbf{j}$, pick the next design point via a shifted van der Corput sequence. Use interpolation to estimate $\left(\widehat{f}_{\mathbf{j}}\right)_{\mathbf{j}\in\mathcal{J}}$.

Product, Order, and Smoothness Dependent (POSD) Weights

$$f(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(|\widehat{f}(\mathbf{j})| / \gamma_{\mathbf{j}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

$$\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}}^{1/q} = \mathcal{O}(d^{p'}) \implies \|f - f_{\text{app}}\|_\infty \leqslant \varepsilon \text{ for } n = \mathcal{O}(d^p) \qquad \text{if } \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} < \infty$$

Experimental design assumes[2]

Effect sparsity: Only a small number of effects are important

Effect hierarchy: Lower-order effects are more important than higher-order effects

Effect heredity: Interaction is active only if both parent effects are also active

Effect smoothness: Coarse horizontal scales are more important than fine horizontal scales

Consider product, order, and smoothness dependent (POSD) weights:

$$\gamma_{\mathbf{j}} = \Gamma_{\|\mathbf{j}\|_0} \prod_{\substack{\ell=1 \\ j_\ell>0}}^{d} w_\ell s_{j_\ell}, \qquad \Gamma_0 = s_1 = 1, \qquad \begin{cases} w_\ell = \text{coordinate importance} \\ \Gamma_r = \text{order size} \\ s_j = \text{smoothness degree} \end{cases}$$

[2] Wu, C. F. J. & Hamada, M. *Experiments: Planning, Analysis, and Parameter Design Optimization.* (John Wiley & Sons, Inc., New York, 2000).

Product, Order, and Smoothness Dependent Weights

Effect sparsity: Only a small number of effects are important

Effect hierarchy: Lower-order effects are more important than higher-order effects

Effect heredity: Interaction is active only if both parent effects are also active

Effect smoothness: Coarse horizontal scales are more important than fine horizontal scales

Consider product, order and smoothness dependent weights:

$$\gamma_{\mathbf{j}} = \Gamma_{\|\mathbf{j}\|_0} \prod_{\substack{\ell=1 \\ j_\ell > 0}}^{d} w_\ell s_{j_\ell}, \qquad \Gamma_0 = s_1 = 1, \qquad \begin{cases} w_\ell = \text{coordinate importance} \\ \Gamma_r = \text{order size} \\ s_j = \text{smoothness degree} \end{cases}$$

$$\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}}^{1/q} = \sum_{\mathfrak{u}\subseteq 1:d} \left[\Gamma_{|\mathfrak{u}|}^{1/q} \left(\prod_{\ell\in\mathfrak{u}} w_\ell^{1/q} \right) \left(\sum_{j=1}^{\infty} s_j^{1/q} \right)^{|\mathfrak{u}|} \right] = \mathcal{O}(d^{p'})$$

$$\Longrightarrow \left\| f - f_{\text{app}} \right\|_\infty \leqslant \varepsilon \text{ for } n = \mathcal{O}(d^p) \qquad \text{if } \left\| \widehat{f} \right\|_{\infty,\boldsymbol{\gamma}} < \infty$$

Special Cases of Weights

$$\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}}^{1/q} = \sum_{\mathfrak{u}\subseteq 1:d} \left[\Gamma_{|\mathfrak{u}|}^{1/q} \left(\prod_{\ell\in\mathfrak{u}} w_\ell^{1/q} \right) \left(\sum_{j=1}^{\infty} s_j^{1/q} \right)^{|\mathfrak{u}|} \right] \overset{\text{Want}}{=} \mathcal{O}(d^{p'})$$

Coordinates, orders equally important
$\Gamma_r = w_\ell = 1$

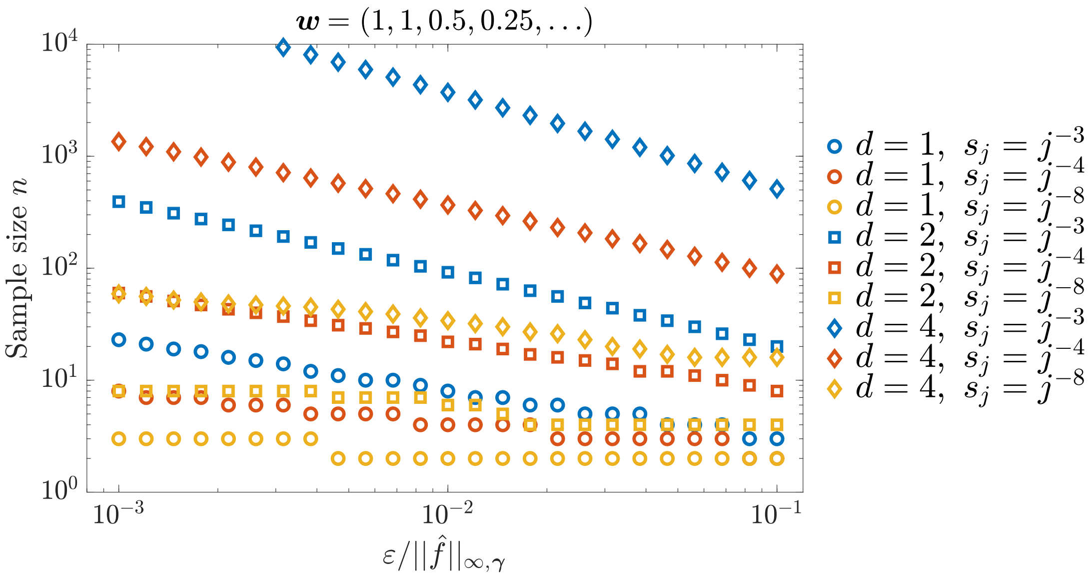
$$\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}}^{1/q} = \left[1 + \sum_{j=1}^{\infty} s_j^{1/q} \right]^{d}$$ Fail

Coordinates equally important
No interactions
$w_\ell = \Gamma_1 = 1, \ \Gamma_r = 0 \ \forall r > 1$

$$\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}}^{1/q} = 1 + d \sum_{j=1}^{\infty} s_j^{1/q}$$ Success

Coordinates differ in importance
Interactions equally important
$\Gamma_r = 1$

$$\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}}^{1/q} \leqslant \exp\left(\sum_{k=1}^{\infty} w_k^{1/q} \sum_{j=1}^{\infty} s_j^{1/q} \right)$$ Success


$\boldsymbol{w} = (1, 1, 0.5, 0.25, \ldots)$
Sample size n
10^0
10^1
10^2
10^3
10^4
10^{-3}
10^{-2}
10^{-1}
$\varepsilon / \|\hat{f}\|_{\infty,\gamma}$
$d = 1,\ s_j = j^{-3}$
$d = 1,\ s_j = j^{-4}$
$d = 1,\ s_j = j^{-8}$
$d = 2,\ s_j = j^{-3}$
$d = 2,\ s_j = j^{-4}$
$d = 2,\ s_j = j^{-8}$
$d = 4,\ s_j = j^{-3}$
$d = 4,\ s_j = j^{-4}$
$d = 4,\ s_j = j^{-8}$

Algorithm When Both $\boldsymbol{\gamma}$ and $\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}$ Are Inferred

Require:
- Γ = vector of order sizes
- $\boldsymbol{s}$ = vector of smoothness degrees
- $w^* = \max_k w_k$
- n_0 = minimum number of wavenumbers in each coordinate
- C = inflation factor
- $\widehat{f}$ = a black-box series coefficient generator for the function of interest, f, where $\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \leqslant C\|(\widehat{f}_{\mathbf{j}})_{\mathbf{j}\in\mathcal{J}}\|_{\infty,\boldsymbol{\gamma}}$, $\mathcal{J} := \{(0,\ldots,0,j,0\ldots,0) : j = 0,\ldots,n_0\}$ for all $\boldsymbol{\gamma}$
- ε = positive absolute error tolerance

Ensure: $\|f - f_{\text{app}}\|_\infty \leqslant \varepsilon$

1: Evaluate $\widehat{f}(\mathbf{j})$ for $\mathbf{j} \in \mathcal{J}$

2: Define $\boldsymbol{w} = \min \underset{w_\ell \leqslant w^*}{\operatorname{argmin}} \|(\widehat{f}_{\mathbf{j}})_{\mathbf{j}\in\mathcal{J}}\|_{\infty,\boldsymbol{\gamma}}$

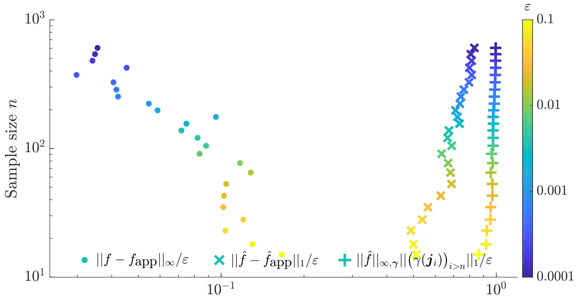
3: Let $n = \min\left\{n' : \sum_{i=n'+1}^{\infty} \gamma_{\mathbf{j}_i} \leqslant \frac{\varepsilon}{C\|(\widehat{f}_{\mathbf{j}})_{\mathbf{j}\in\mathcal{J}}\|_{\infty,\boldsymbol{\gamma}}}\right\}$

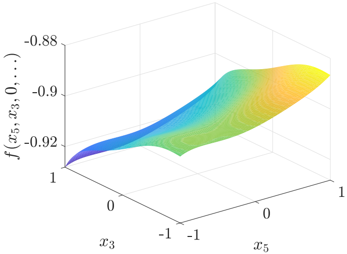
4: Compute $f_{\text{app}} = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i}$

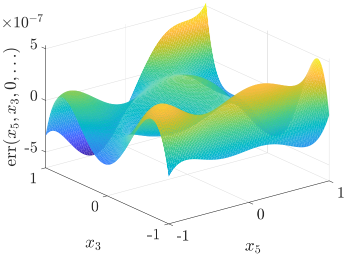
Computational cost is $n = \mathcal{O}\left(\left[\varepsilon^{-1}C\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}\|\boldsymbol{\gamma}\|_{1/q}\right]^{1/(q-1)}\right)$

Example

f manufactured in terms of random series coefficients







A Gap Between Theory and Practice

Theory using series coefficients

Practice using function values

Photo Credit: Xinhua

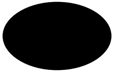
A Very Sparse Grid on $[-1,1]^d$

j	0	1	2	3	4	$\cdots$
van der Corput t_j	0	$1/2$	$1/4$	$3/4$	$1/8$	$\cdots$
$\psi(t_j) := 2(t_j + 1/3 \bmod 1) - 1$	$-1/3$	$2/3$	$1/6$	$-5/6$	$-1/12$	$\cdots$
$\psi(t_j) := -\cos(\pi(t_j + 1/3 \bmod 1))$	-0.5	0.8660	0.2588	-0.9659	-0.1305	$\cdots$

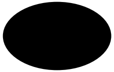
To estimate $\widehat{f}(\mathbf{j})$, $\mathbf{j} \in \mathcal{J}$, use the design $\{(\psi(t_{j_1}), \ldots, \psi(t_{j_d}) : \mathbf{j} \in \mathcal{J}\}$. E.g., for

$$\mathcal{J} = \{(0,0,0,0), (1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1), (2,0,0,0), (3,0,0,0), (1,1,0,0)\}$$

Even Points



ArcCos Points



Algorithm Using Function Values When Both γ and $\left\|\widehat{f}\right\|_{\infty,\gamma}$ Are Inferred

Require: ■ $\boldsymbol{\Gamma}$ = vector of order sizes ■ $\boldsymbol{s}$ = vector of smoothness degrees ■ $w^* = \max_k w_k$

■ n_0 = minimum number of wavenumbers in each coordinate ■ C = inflation factor

■ f = a black-box function value generator ■ ε = positive absolute error tolerance

Ensure: $\|f - f_{\text{app}}\|_\infty \leqslant \varepsilon$

1: Approximate $\widehat{f}(\mathbf{j})$ for $\mathbf{j} \in \mathcal{J} := \{(0,\ldots,0,j,0\ldots,0) : j = 1,\ldots,n_0\}$ by interpolating the function data $\{(\mathbf{x_j}, f(\mathbf{x_j})) : \mathbf{x_j} = \psi(t_{j_1}),\ldots,\psi(t_{j_d}),\ \mathbf{j} \in \mathcal{J}\}$

2: Define $\boldsymbol{w} = \min \operatorname{argmin}_{w_\ell \leqslant w^*} \left\| \left(\widehat{f}_{\mathbf{j}}\right)_{\mathbf{j}\in\mathcal{J}} \right\|_{\infty,\gamma}$

3: **while** $C\left\| \left(\widehat{f}_{\mathbf{j}}\right)_{\mathbf{j}\in\mathcal{J}} \right\|_{\infty,\gamma} \sum_{\mathbf{j}\notin\mathcal{J}} \gamma_{\mathbf{j}} > \varepsilon$ **do**

4: Add $\operatorname{argmin}_{\mathbf{j}\notin\mathcal{J}} \gamma_{\mathbf{j}}$ to $\mathcal{J}$

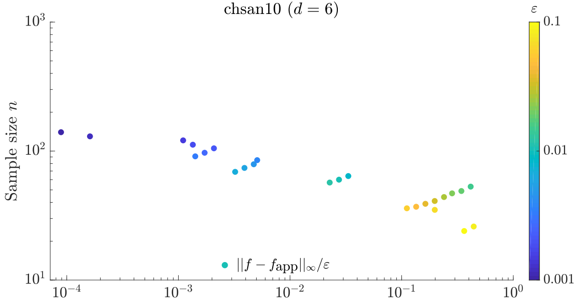
5: Approximate $\widehat{f}(\mathbf{j})$ for $\mathbf{j} \in \mathcal{J}$ by interpolating the function data $\{(\mathbf{x_j}, f(\mathbf{x_j})) : \boldsymbol{x} = \psi(t_{j_1}),\ldots,\psi(t_{j_d}),\ \mathbf{j} \in \mathcal{J}\}$

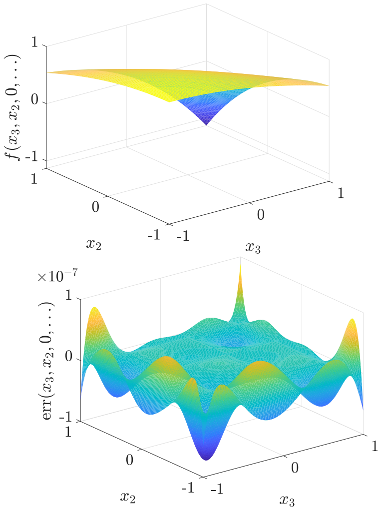
6: **end while**

7: Compute $f_{\text{app}} = \sum_{\mathbf{j}\in\mathcal{J}} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}$

Example[3]

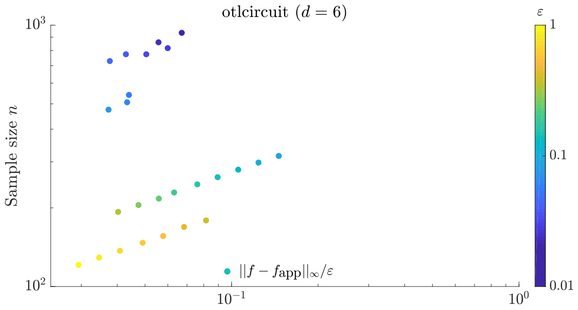
$$f(\mathbf{x}) = \exp((x_2+1)(x_3+1)/4)\cos((x_2+1)/2+(x_3+1)/2), \quad d = 6$$

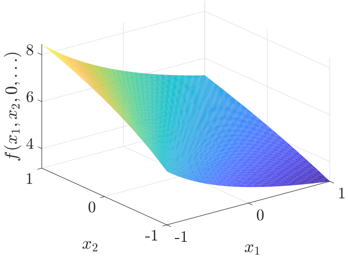


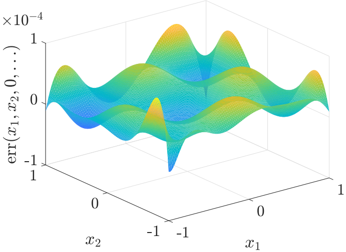


[3] Bingham, D. & Surjano, S. *Virtual Library of Simulation Experiments.* 2013. https://www.sfu.ca/~ssurjano/.

OTL CircuitExample[4]







[4] Bingham, D. & Surjano, S. *Virtual Library of Simulation Experiments.* 2013. https://www.sfu.ca/~ssurjano/.

Summary

- Functions must be nice to succeed with few function values
- Ideas underlying experimental design and tractability show us how to define "nice"
    - Effect sparsity, hierarchy, heredity, and smoothness
    - Product, order, and smoothness dependent (POSD) weighted function spaces
- Infer properties of f from limited data ($\boldsymbol{\gamma}$, $\|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}$, $\widehat{f}$)
- Must assume some structure on weights to make progress at all
- Design determined by wavenumbers included in approximation via van der Corput, preserves low condition number of the design matrix
- Gap in theory when sampling function values versus series coefficients
- Sample size seems to be larger than necessary
- Can we also infer the smoothness weights?

Thank you

These slides are available at

`speakerdeck.com/fjhickernell/samsi-qmc-transition-2018-may`

Novak, E. & Woźniakowski, H. *Tractability of Multivariate Problems Volume I: Linear Information.* *EMS Tracts in Mathematics* **6** (European Mathematical Society, Zürich, 2008).

Kühn, T., Sickel, W. & Ullrich, T. Approximation numbers of Sobolev embeddings—Sharp constants and tractability. *J. Complexity* **30,** 95–116 (2014).

Wu, C. F. J. & Hamada, M. *Experiments: Planning, Analysis, and Parameter Design Optimization.* (John Wiley & Sons, Inc., New York, 2000).

Bingham, D. & Surjano, S. *Virtual Library of Simulation Experiments.* 2013. https://www.sfu.ca/~ssurjano/.

In What Sense Is This Optimal?

$$f(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} \widehat{f}(\mathbf{j})\phi_{\mathbf{j}}(\mathbf{x}), \quad \widehat{f}(\mathbf{j}) = \frac{\langle f, \phi_{\mathbf{j}}\rangle}{\langle \phi_{\mathbf{j}}, \phi_{\mathbf{j}}\rangle}, \quad \|\phi_{\mathbf{j}}\|_\infty = 1, \quad \|\widehat{f}\|_{q,\boldsymbol{\gamma}} = \left\| \left(\frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} \right)_{\mathbf{j}\in\mathbb{N}_0^d} \right\|_q$$

$$\gamma_{\mathbf{j}_1} \geqslant \gamma_{\mathbf{j}_2} \geqslant \cdots, \quad f_{\text{app}}(\mathbf{x}) = \sum_{i=1}^{n} \widehat{f}(\mathbf{j}_i)\phi_{\mathbf{j}_i}, \quad \|f - f_{\text{app}}\|_\infty \overset{\text{loose}}{\leqslant} \|\widehat{f} - \widehat{f}_{\text{app}}\|_1 \underset{\text{optimal}}{\overset{\text{tight}}{\leqslant}} \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i}$$

For any other approximation, g, based on series coefficients, $\{\widehat{f}(\mathbf{j})\}_{\mathbf{j}\in\mathcal{J}}$ with $|\mathcal{J}| = n$,

$$\sup_{\substack{h:\|\widehat{h}\|_{\infty,\boldsymbol{\gamma}} = \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \\ \widehat{h}(\mathbf{j}) = \widehat{f}(\mathbf{j})\ \forall \mathbf{j}\in\mathcal{J}}} \|\widehat{h} - \widehat{g}\|_1 = \left\| \left(|\widehat{f}(\mathbf{j}) - \widehat{g}(\mathbf{j})| \right)_{\mathbf{j}\in\mathcal{J}} \right\|_1 + \sup_{h:\|\widehat{h}\|_{\infty,\boldsymbol{\gamma}} = \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}} \left\| \left(|\widehat{h}(\mathbf{j}) - \widehat{g}(\mathbf{j})| \right)_{\mathbf{j}\notin\mathcal{J}} \right\|_1$$

$$\geqslant \sup_{h:\|\widehat{h}\|_{\infty,\boldsymbol{\gamma}} = \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}} \left\| \left(\widehat{h}(\mathbf{j}) \right)_{\mathbf{j}\notin\mathcal{J}} \right\|_1 = \sup_{h:\|\widehat{h}\|_{\infty,\boldsymbol{\gamma}} = \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}}} \|\widehat{h}\|_{\infty,\boldsymbol{\gamma}} \sum_{\mathbf{j}\notin\mathcal{J}} \gamma_{\mathbf{j}}$$

$$\geqslant \|\widehat{f}\|_{\infty,\boldsymbol{\gamma}} \sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i}$$

back

Inferring γ from Data

Given (estimates of) series coefficients, $\widehat{f}(\mathbf{j})$ for $\mathbf{j} \in \mathcal{J} := \{(0,\dots,0,j,0\dots,0) : j = 1,\dots,n_0\}$, and fixed $\{\Gamma_r\}_{r=0}^{d}$, and $\{s_j\}_{f=0}^{\infty}$, note that

$$\left\| \left(\widehat{f}(\mathbf{j})\right)_{\mathbf{j}\in\mathcal{J}} \right\|_{\infty,\boldsymbol{\gamma}} = \max_{\mathbf{j}\in\mathcal{J}} \frac{|\widehat{f}(\mathbf{j})|}{\gamma_{\mathbf{j}}} = \frac{1}{\Gamma_1} \max_{k=1,\dots,d} \frac{\widehat{f}_{k,\max}}{w_k}, \qquad \widehat{f}_{k,\max} := \sup_{j=1,\dots,n_0} \frac{|\widehat{f}(j\boldsymbol{e}_k)|}{s_j}$$

We choose

$$w_k = \frac{\widehat{f}_{k,\max}}{\max_{\ell} \widehat{f}_{\ell,\max}}, \qquad \left\| \left(\widehat{f}(\mathbf{j})\right)_{\mathbf{j}\in\mathcal{J}} \right\|_{\infty,\boldsymbol{\gamma}} = \frac{\max_{\ell} \widehat{f}_{\ell,\max}}{\Gamma_1}$$

Tail Sum of γ

The term

$$\sum_{i=n+1}^{\infty} \gamma_{\mathbf{j}_i} = \sum_{i=1}^{\infty} \gamma_{\mathbf{j}_i} - \sum_{i=1}^{n} \gamma_{\mathbf{j}_i}$$

appears in the error bound. For certain γ of PSD form, we can compute the first sum on the right:

$$\sum_{\mathbf{j}\in\mathbb{N}_0^d} \gamma_{\mathbf{j}} = \sum_{\mathfrak{u}\subseteq 1:d} \left[\left(\prod_{\ell\in\mathfrak{u}} w_\ell \right) \left(\sum_{j=1}^{\infty} s_j \right)^{|\mathfrak{u}|} \right] = \prod_{\ell=1}^{d} (1 + w_\ell s_{\mathsf{sum}}), \qquad s_{\mathsf{sum}} = \sum_{j=1}^{\infty} s_j$$