Sparse Grids

& "A Dynamically Adaptive Sparse Grid Method for Quasi-Optimal Interpolation of Multidimensional Analytic Functions" from MK Stoyanov, CG Webster

Léopold Cambier

ICME, Stanford University

February 17, 2017

The Problem

A Recurring Question

How to interpolate

$$f : X \subset \mathbb{R}^d \to \mathbb{R}$$

where

$$X = \bigotimes_{i=1}^{d} [-1, 1]$$

and d is "large" (> 1)

Naive Solution

- Simply use repeated *one-dimensional* rules
- We have a tensor of rules

$$\begin{aligned} T_n^d f &= \bigotimes_{i=1}^{d} I_{n_i}^i f \\ &= I_{n_1}^1 \otimes I_{n_2}^2 \otimes \cdots \otimes I_{n_d}^d f \end{aligned}$$

- In practice

$$(T_n^d f)(x) = \sum_{k_1=1}^{m(n_1)} \cdots \sum_{k_d=1}^{m(n_d)} \underbrace{f(\bar{x}_{1,k_1}, \ldots, \bar{x}_{d,k_d})}_{\text{interpolation nodes}} \underbrace{T_{k_1}^1(x_1) \ldots T_{k_d}^d(x_d)}_{\text{Lagrange basis functions}}$$

The Issue ?

Curse of Dimensionality

m nodes per dimension $\Rightarrow$ m^d nodes total

Outline

1. Curse of Dimensionality
2. Sparse Grids
3. How to Choose Θ ?
   - Quasi-Optimal Polynomial Space
   - Quasi-Optimal Interpolation
   - Estimating the Parameters
   - Optimal Sparse Grids Interpolant
4. Numerical Results
5. Conclusion

Sparse Grids

The Main Idea

- Consider the tensor rule

$$T_n^d = \bigotimes_{i=1}^{d} I_{n_i}^i$$

- Define

$$\Delta_j^i = I_j^i - I_{j-1}^i$$

- Rewrite

$$T_n^d = \sum_{\alpha \leq n} \bigotimes_{i=1}^{d} \Delta_{\alpha_i}^i$$

Sparse Grids

Sparse Grids Interpolator

$$Q^d_\Theta = \sum_{\alpha \in \Theta} \bigotimes_{i=1}^{d} \Delta^i_{\alpha_i}$$

with $\Theta \subset \mathbb{N}^d$ and where ideally

$$|\Theta| \ll n^d$$

If

$$\Theta = \{\alpha \in \mathbb{N}^d : \alpha \leq n\} \Rightarrow Q^d_\Theta = T^d_n$$

Back to the Interpolation Formula

- We want an expression like

$$(If)(x) = \sum_k T_k(x) f(x_k)$$

- We have (if Θ is *admissible*[1])

$$Q^d_\Theta = \sum_{\alpha \in \Theta} \bigotimes_{i=1}^{d} \Delta^i_{\alpha_i}$$

$$= \sum_{\alpha \in \Theta} c_\alpha \bigotimes_{i=1}^{d} I^i_{\alpha_i}$$

[1] $\forall \alpha \in \Theta$, $\{\beta : \beta \leq \alpha\} \subseteq \Theta$

Back to the Interpolation Formula

- Since

$$\left(\bigotimes_{i=1}^{d} I_{\alpha_i}^{i}\right)[f](x) = \sum_{k_1=1}^{m(\alpha_1)} \cdots \sum_{k_d=1}^{m(\alpha_d)} f(\bar{x}_{\alpha_1,k_1}, \ldots, \bar{x}_{\alpha_d,k_d}) U_{\alpha_1,k_1}(x_1) \ldots U_{\alpha_d,k_d}(x_d)$$
$$= \sum_{k \leq m(\alpha)} f(\bar{x}_{\alpha,k}) U_{\alpha,k}(x)$$

- We have

$$Q_{\Theta}^{d} = \sum_{\alpha \in \Theta} c_\alpha \bigotimes_{i=1}^{d} I_{\alpha_i}^{i}$$
$$\Rightarrow Q_{\Theta}^{d}[f](x) = \sum_{\alpha \in \Theta} c_\alpha \sum_{k \leq m(\alpha)} f(\bar{x}_{\alpha,k}) U_{\alpha,k}(x)$$

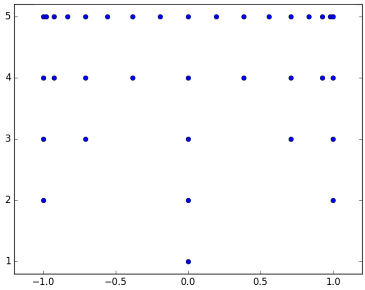
Back to the Interpolation Formula

- If nodes are *nested* $\Rightarrow$ factor $f(\bar{x}_{\alpha,k})$ and find $\mathcal{K}(\alpha)$ such that

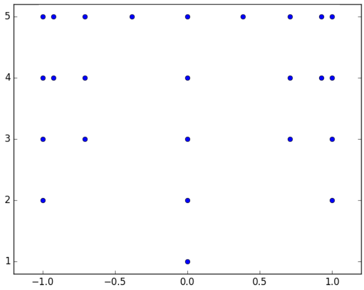
$$Q^d_\Theta[f](x) = \sum_k f(x_k) \sum_{\alpha\in\Theta : k\in\mathcal{K}(\alpha)} c_\alpha U_{\alpha,k}(x)$$
$$= \sum_k f(x_k) U_k(x)$$

Univariates Rules

- It is preferable to have nested interpolation nodes
- Let $m(l)$ be the number of nodes as a function of the order l
- Different rules are possible: Clenshaw-Curtis ($m(l) = 2^{l-1} + 1$), R-Leja 2 ($m(l) = 2l - 1$), etc.



Clenshaw-Curtis

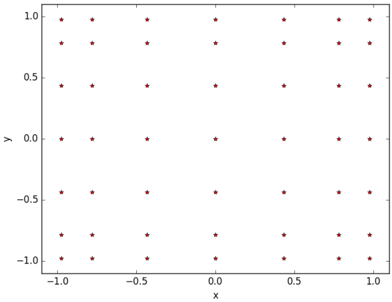


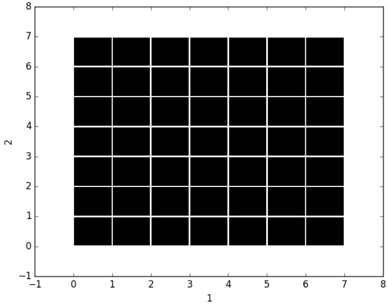
R-Leja 2

Example 1

$$f(x, y) = x^6 + y^6$$

With tensor method ($\epsilon = 10^{-16}$).

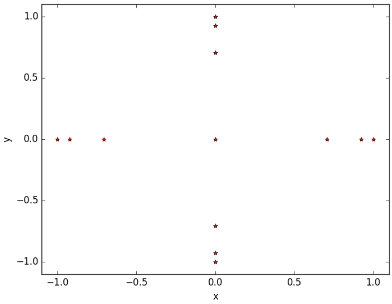


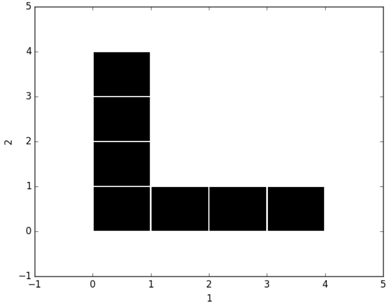


Example 1

$$f(x, y) = x^6 + y^6$$

With sparse grid method ($\epsilon = 10^{-16}$) using R-Leja 2 rule.

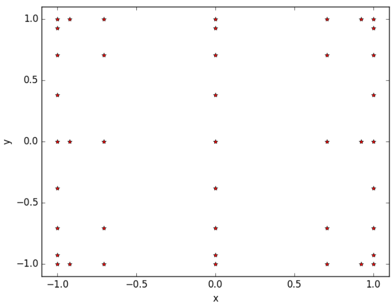


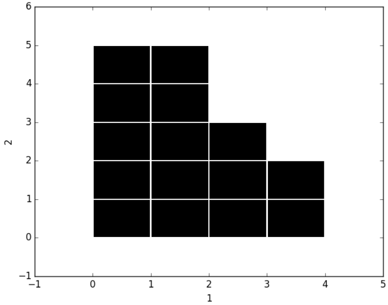


Example 2

$$f(x, y) = \sqrt{1 + x^2 + y^2}$$

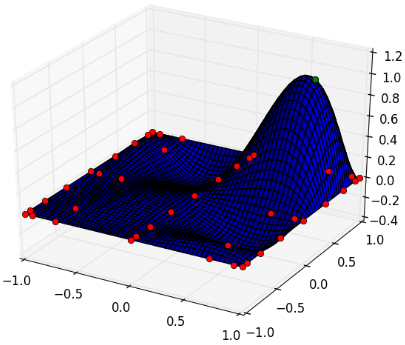
With sparse grid method ($\epsilon = 10^{-4}$) using R-Leja 2 rule.



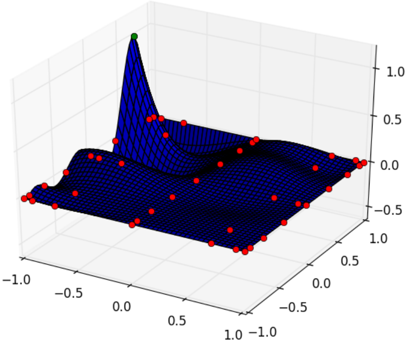


Example 2: Basis Functions

$$Q_{\Theta}^{d}[f](x) = \sum_{k} f(x_k) U_k(x)$$



$U_2(x)$

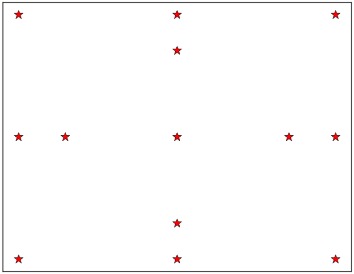


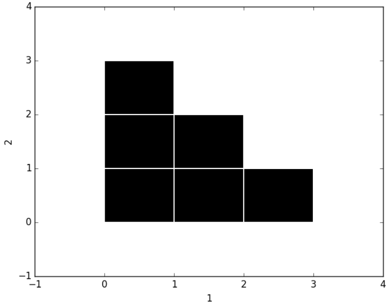
$U_5(x)$

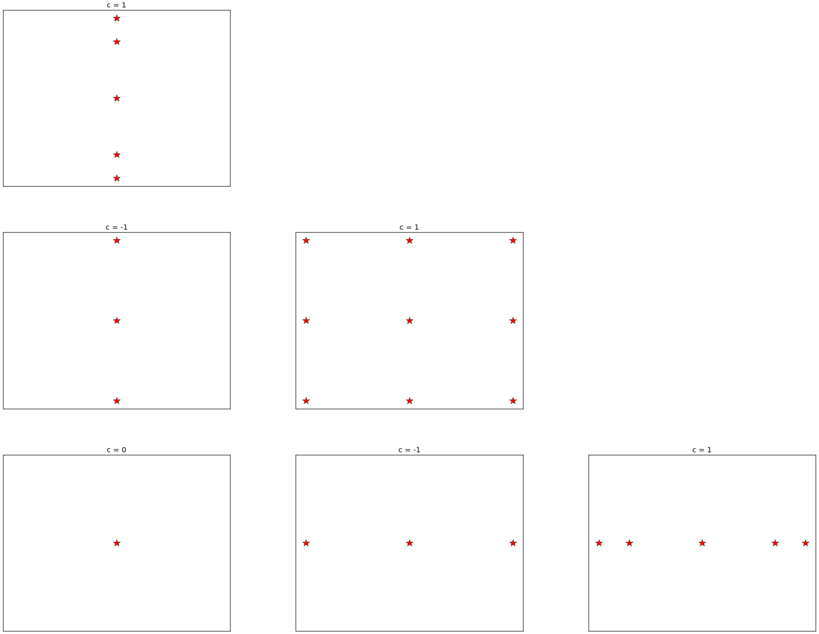
Example 3: Sum of Tensor Rules

$$f(x, y) = x^4 + x^2y^2 + y^4$$

$$Q_{\Theta}^{d} = \sum_{\alpha \in \Theta} c_{\alpha} \bigotimes_{i=1}^{d} I_{\alpha_i}^{i}$$






c = 1
c = -1
c = 1
c = 0
c = -1
c = 1

A Question Remains

How to choose Θ ?

How to Choose Θ ?

How to Choose Θ

M. Stoyanov, C. Webster, "A Dynamically Adaptive Sparse Grid Method for Quasi-Optimal Interpolation of Multidimensional Analytic Functions"

Let's Take a Step Back

There are actually 2 questions:

- How well can we expand a function f in terms of mixed-order polynomials. This is *projection*
- How good/bad is the *interpolation* versus *projection*.

Projection

- Working on $\Gamma = [-1, 1]^d$
- $\Lambda \subset \mathbb{N}^d$ is the *degrees* space
- Project in

$$\mathcal{P}_\Lambda = \text{span}\left\{x \to x^\nu : \nu \in \Lambda\right\}$$

- Legendre polynomials is a good basis
- Project as

$$f \approx f_\Lambda = \sum_{\nu \in \Lambda} c_\nu L_\nu$$

where L_ν are Legendre polynomials of degree at most x^ν and

$$c_\nu = \int_\Gamma f(x) L_\nu(x) \mathrm{d}x.$$

Projection

- Projection is

$$f \approx f_\Lambda = \sum_{\nu \in \Lambda} c_\nu L_\nu$$

- L_2 error is

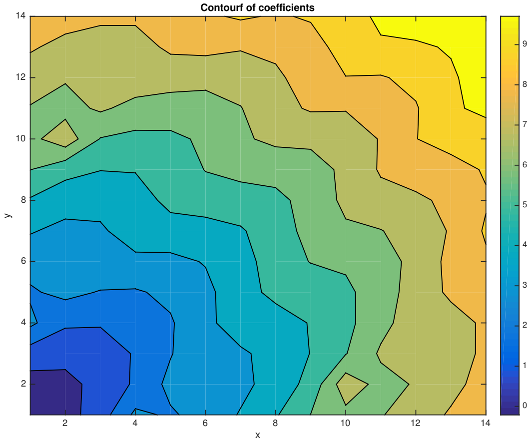
$$\|f - f_\Lambda\|_{L_2}^2 = \sum_{\nu \notin \Lambda} |c_\nu|^2$$

Question

How fast does $|c_\nu|$ decrease ?

Example

$$f(x, y) = \frac{1}{\sqrt{(x-2)^2 + (y-2)^2}} = \sum_{\nu} c_\nu L_\nu(x, y)$$



Quasi-Optimal Projection Space

Assumption 1

f is holomorphic[a] in a poly-ellipse

$$\mathcal{E}_\rho = \bigcup_{\theta\in[0,2\pi]} \bigotimes_{i=1}^{d} \left\{ z_k \in \mathbb{C} : |\Re(z_k)| \leq \frac{\rho_k + \rho_k^{-1}}{2}\cos\theta, |\Im(z_k)| \leq \frac{\rho_k - \rho_k^{-1}}{2}\sin\theta \right\}$$

[a] Differentiable in the neighborhood of every point $\Rightarrow$ infinitively differentiable

Result 1

Under this assumption,

$$|c_v| \leq C \exp(-\alpha\cdot\nu) \prod_{i=1}^{d} \sqrt{2\nu_k + 1}$$

with $\alpha = \log(\rho)$

Proof : Taylor Integral Theorem and Formula

Quasi-Optimal Projection Space

The optimal projection space of level p is then defined as

$$\Lambda^{\alpha}(p) = \left\{ \nu : \alpha \cdot \nu - \frac{1}{2} \sum_{i=1}^{d} \log(\nu_i + 0.5) \leq p \right\}$$

Interpolation

- Interpolation $\neq$ Projection
- It depends on the set of nodes
- Can be arbitrarily bad even for "nice" function (Runge Phenomenon)
- Quantified by the Lebesgue constant

Lebesgue Constant

- I is the interpolation operator,

$$I : \{f \text{ bounded}\} \to \mathcal{P}_\Lambda$$

- p^* is the projection of f on $\mathcal{P}_\Lambda$
- Then,

$$\|f - If\|_\infty \leq (1 + C_\Lambda)\|f - p^*\|_\infty$$

- With

$$\|I\|_\infty = C_\Lambda = \sup_g \frac{\|Ig\|_\infty}{\|g\|_\infty}$$

and g bounded.

Lebesgue Constant

Assumption 2

For $\Lambda_\nu = \{\alpha : \alpha \leq \nu\}$ (tensor),

$$C_{\Lambda_\nu} \leq C_\gamma \prod_{i=1}^{d} (\nu_k + 1)^{\gamma_k}$$

i.e., polynomial growth.

- True for Chebyshev-like polynomial interpolation
- False for equispaced interpolation.

Lebesgue Constant for Sparse Grids

Lebesgue Constant

If

$$\lambda_l = \|I_l^i\|_\infty \le C_\gamma (l+1)^\gamma$$

then

$$\|Q_\Theta^d\|_\infty \le C_\gamma^d |\Theta|^{\gamma+1}$$

- Usually not sharp
- Polynomial

Lebesgue Constant for Clenshaw-Curtis

- Roots of Chebyshev Polynomials

$$x_k = \cos\left(\frac{\pi k}{n}\right) \quad k = 0, \ldots, n$$

where

$$m(l) = 2^{l-1} + 1$$

- One can show

$$\lambda_l \approx \log(m(l)) \propto l$$

Lebesgue Constant for Other Rules

- R-Leja 2 with

$$m(l) = 2l - 1$$

  and

$$\lambda_l \approx m(l) \propto l$$

- Selecting the nodes by minimizing $\|I_l\|_\infty$

$$m(l) = l + 1, \quad \lambda_l \approx 4\sqrt{l+1}$$

Quasi-Optimal Interpolation

Combine both results to get level sets like

$$C_\nu C \exp(-\alpha \cdot \nu) \prod_{i=1}^{d} \sqrt{2\nu_k + 1} \leq \tilde{C} \exp(-\alpha \cdot \nu) \prod_{i=1}^{d} (\nu_k + 1)^{\gamma_k + 0.5}$$

Optimal Interpolation Space of level L

$$\Lambda^{\alpha,\beta}(L) = \{\nu : \alpha \cdot \nu + \beta \log(\nu + 1) \leq L\}$$

for unknown α (projection error) and β (interpolation error).

Estimating α and β

- We do not know α and β
- We can estimate them from the interpolant $\bar{f}_\Lambda$
  1. Get $\bar{c}_\nu$
  $$f(x) \approx \bar{f}_\Lambda(x) = \sum_{\nu \in \Lambda} \bar{c}_\nu L_\nu(x)$$
  2. Assume
  $$|\bar{c}_\nu| \propto \exp(-\alpha \cdot \nu)(\nu + 1)^{-\beta}$$
  3. Solve
  $$\min_{\alpha,\beta} \sum_{\nu \in \Lambda} (C + \alpha \cdot \nu + \beta \log(\nu + 1) + \log |\bar{c}_\nu|)^2$$

Let's Summarize

- Sparse Grids

$$Q^d_\Theta = \sum_{\alpha\in\Theta} \bigotimes_{i=1}^{d} \Delta^i_{\alpha_i} = \sum_{\alpha\in\Theta} c_\alpha \bigotimes_{i=1}^{d} I^d_{\alpha_i}$$

with Θ the *order* (*level*) space and I^i_l a sequence of 1-d rules of order l with $m(l)$ nodes and polynomial Lebesgue constant λ_l.

- Optimal level-p interpolant has a *degree* space as

$$\Lambda(p) = \{\nu : \alpha\cdot\nu + \beta\log(\nu+1) \leq p\}$$

where α and β can be approximated based on a *current* interpolant

Minimal Polynomial Interpolant

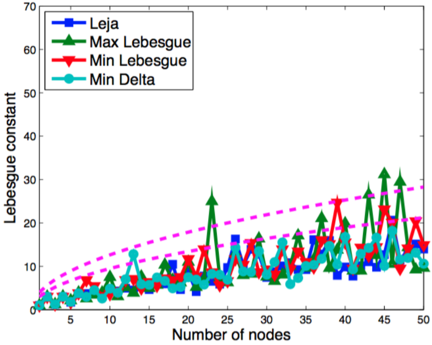
For given p, the smallest Θ that covers $\Lambda(p)$ (unique) is optimal.

Algorithm

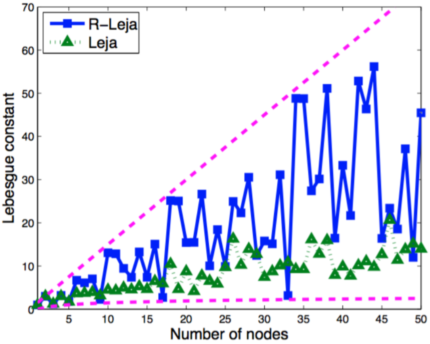
- Given univariate rules,
- Select initial Λ^0 and Θ^0 optimal
- Repeat for $n = 0, 1, \ldots$
  - Compute $f_{\Lambda^n} = Q^d_{\Theta^n} f$
  - Compute $\bar{c}_\nu$
  - Estimate α and β
  - Update Λ^{n+1} and Θ^{n+1}

Numerical Results

Estimating Lebesgue's Constant



$3\sqrt{l+1}$ and $4\sqrt{l+1}$



$\frac{3}{2}(l+1)$ and $\frac{2}{\pi}\log(2^l+1)$

Experiment 1 : Parametrized Elliptic PDE

- Solve in $x \in D$

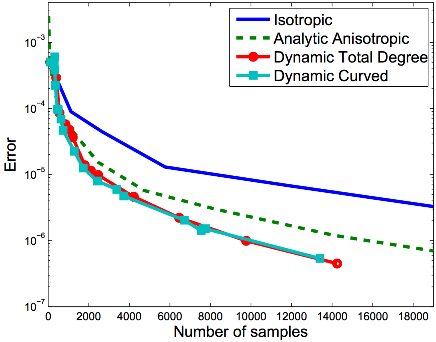
$$-\nabla_x (a(x,y) \nabla_x u(x,y)) = b(x)$$

- Evaluate for $y \in \Gamma \in \mathbb{R}^7$

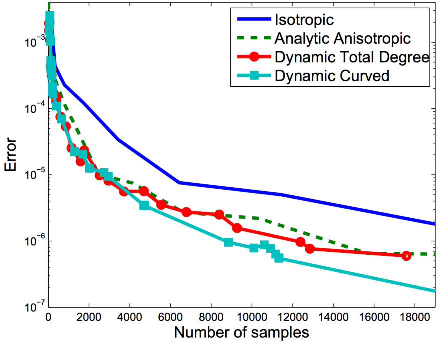
$$f(y) = \|u(x,y)\|_{L_2}$$

- Goal : interpolate f

Experiment 1 : Parametrized Elliptic PDE

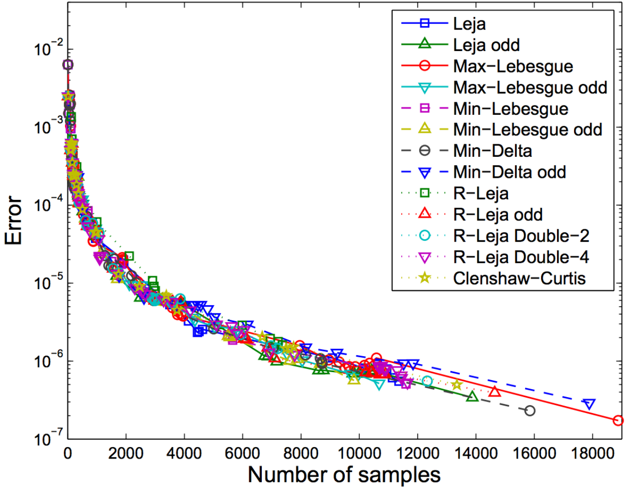


Clenshaw-Curtis



R-Leja

Experiment 1 : Parametrized Elliptic PDE



Experiment 2 : Steady-State Burger's Equation

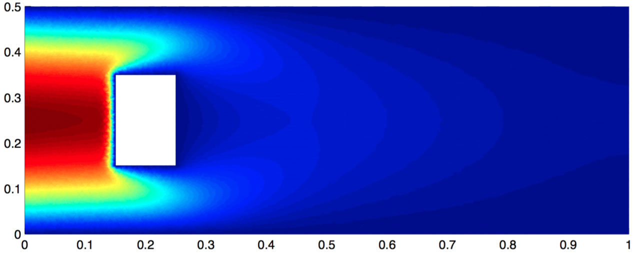
- Solve in $x \in D \in \mathbb{R}^2$

$$-\nabla_x \cdot (a(y)\nabla_x u(x,y)) + (v(y) \cdot \nabla_x u(x,y))u(x,y) = 0$$

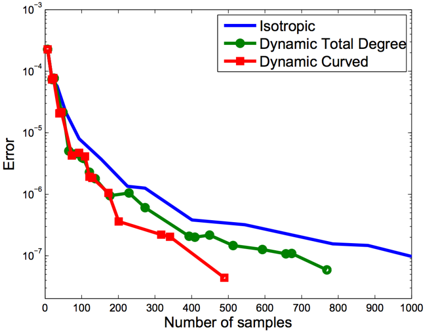
- Evaluate for $y \in [-1,1]^3$

$$f(y) = \int_{\tilde{D}} u(x,y)\mathrm{d}x$$

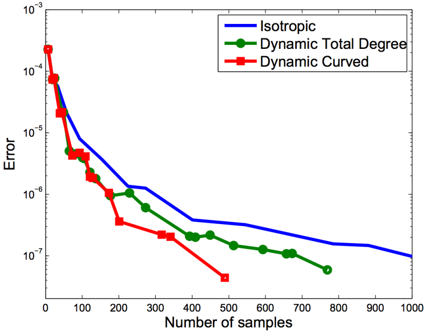
- Goal : interpolate f



Experiment 2 : Steady-State Burger's Equation



Clenshaw-Curtis



R-Leja

Conclusion

- Adaptive sparse grids algorithm
- Source of error : projection (α - "best M terms") and (β - Lebesgue's constant)
- Coefficients can be estimated from previous interpolator
- Several new univariate rules based on minimization of Lebesgue's constant

References

All results are from [1]. [2, 3, 4] are interesting references.

- Stoyanov, Miroslav K., and Clayton G. Webster. "A dynamically adaptive sparse grid method for quasi-optimal interpolation of multidimensional analytic functions." arXiv preprint arXiv:1508.01125 (2015).
- Kaarnioja, Vesa. "Smolyak quadrature." (2013).
- Beck, Joakim, et al. "Convergence of quasi-optimal stochastic Galerkin methods for a class of PDEs with random coefficients." Computers & Mathematics with Applications 67.4 (2014): 732-751.
- Nobile, Fabio, Lorenzo Tamellini, and Raúl Tempone. "Convergence of quasi-optimal sparse grid approximation of Hilbert-valued functions: application to random elliptic PDEs." Mathicse report 12 (2014).