A Hyperspherical Approach for Discontinuity Detection in Uncertainty Quantification

Guannan Zhang, Clayton Webster, Max Gunzburger, John Burkardt

....

http://people.sc.fsu.edu/∼jburkardt/presentations/...
hypersphere_2014_siam_seas.pdf

....

SIAM SEAS Conference
28-30 March 2014

Motivation: Discontinuities

Smooth models are powerful and ubiquitous in computation; they fail when applied unknowingly to problems in which **discontinuities** arise.

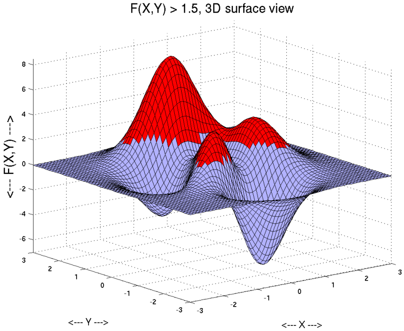
Given a mathematical model exhibiting discontinuity, we may wish to:

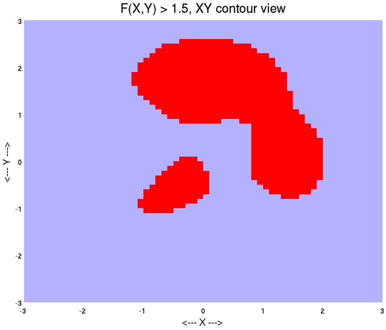
- identify the points of discontinuity;
- subdivide the geometry into subregions of smooth behavior;
- construct piecewise interpolants which are smooth over a subregion;
- estimate the volume of a given subregion;
- estimate integrals over a given subregion;

As special cases of such models, we may include:

- the characteristic function $\chi(x)$ of some region; $\{x : \chi(x) = 1\}$;
- the points where a scalar function $f(x)$ exceeds a threshold: $\{x : f(x) > f_{min}\}$;

While our original function might be smooth, using a threshold creates a discontinuous function, and one or more implicitly defined regions (red).





Motivation: Detecting and Measuring Discontinuities

We are interested in discontinuities because they arise occasionally during the analysis of multidimensional functions associated with stochastic problems, particularly when thresholding is applied by the analyst, or when the physical system itself actually includes essentially discontinous behavior.

A typical problem might include spatial variables $\mathbf{x}$ and stochastic parameters ω; although it is easy to visualize functions which are discontinuous in space, we are more interested in the harder problems in which discontinuity occurs with respect to the nonphysical parameters.

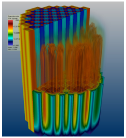
The dimension of the stochastic parameter set may be large, and it may be difficult simply to detect that a discontinuity has arisen, let alone to deal with it. We term this general task High Dimensional Discontinuity Detection (HDDD).

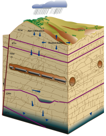
Motivation: Uncertainty Quantification, Rare Events

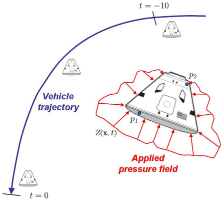
In Uncertainty Quantification (UQ), we may be given a random variable $\omega \in \mathbb{R}^N$ with the PDF $\rho(\omega)$, a state variable $u(\mathbf{x}, \omega)$ governed by a PDE, and a quantity of interest $q(u(:, \omega)$; we seek the probability of the event $\{q(u(:, \omega)) \geq q_0\}$, i.e.

$$\mathbb{P}(q(u(:, \omega) \geq q_0)) = \int \mathcal{I}_{\{q(u(:, \omega)) \geq q_0\}}(\omega) d\rho(\omega)$$

High dimensional discontinuity detection can be used in quantifying probabilities of rare events and risk assessment analysis for turbulence, nuclear reactors, groundwater flow, or the re-entry of space vehicles, in which the physical parameters are only known approximately or are modeled probabilistically.

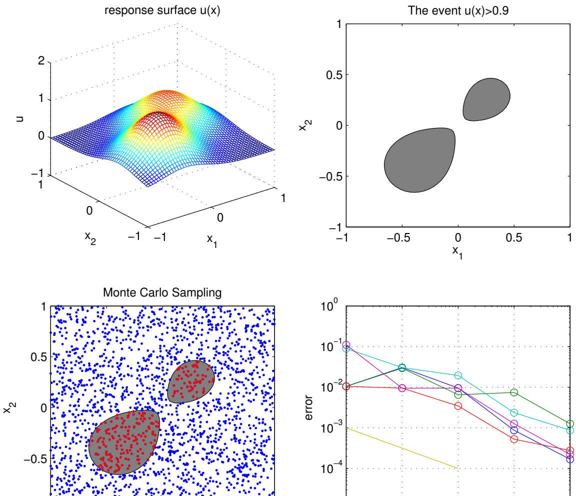






Motivation: Estimating Probability Integrals

To estimate the probability that $q(\omega)$ exceeds a threshold, extract sample values ω from parameter space, solve for q, and record the proportion of times $q(\omega)$ exceeds the threshold. The error decreases at the usual Monte-Carlo rate.



Motivation: PDE's with random input

In the UQ setting, our mathematical model might look like:

$$\begin{cases} -\nabla \cdot (a(\mathbf{x},\omega)\nabla \mathbf{u}(\mathbf{x},\omega)) &= f(\mathbf{x},\omega) \quad \text{in } D \times \Omega, \\ u(\mathbf{x},\omega) &= 0 \quad \text{on } \partial D \times \Omega, \end{cases}$$

with $\mathbf{x} \in \mathbf{D}$ and $\omega \in \Omega$ where $(\Omega, \mathcal{F}, P)$ is a complete probability space.

We assume the random fields $a(\mathbf{x},\omega)$ and $f(\mathbf{x},\omega)$ depend on a finite number of random variables $\mathbf{y}(\omega)$ so that

$$a(\mathbf{x},\omega) = \mathbf{a}(\mathbf{x},\mathbf{y}(\omega)), \quad \mathbf{f}(\mathbf{x},\omega) = \mathbf{f}(\mathbf{x},\mathbf{y}(\omega))$$

Estimating the effect of uncertainty means evaluating the model over the probability domain: $\Gamma = \prod_{i=1}^{N} \Gamma_i \subset \mathbb{R}^N$. In the simplest cases, our investigation involves integration over the entire space with a weight, or over a subset defined by some hypercube.

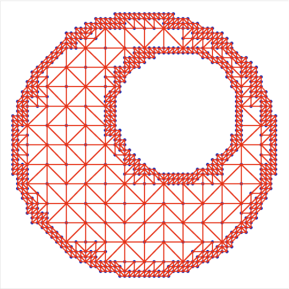
Motivation: Fundamental Problems

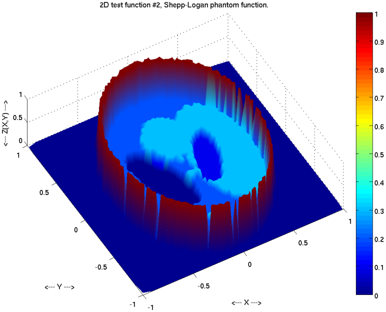
We often begin with a bounded domain $\Gamma \subset \mathbb{R}^N$ which is a product region, but it may be the case that we are interested in a subdomain D which can only be described implicitly. We may choose to represent D using a characteristic function $f(\mathbf{x}) : \Gamma \to \mathbb{R}$ defined by

$$f(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in D \subset \Gamma, \\ 0, & \text{otherwise} . \end{cases}$$

- Can we detect regions in Γ that are near the boundary ∂D. of the discontinuity of $f(\mathbf{x})$?
- Can we detect and handle regions in Γ forming components of D?
- Can we accurately and efficiently estimate the integral: $\int_\Gamma f(\mathbf{x}) g(x) d\mathbf{x} = \int_D g(x) d\mathbf{x}$?

Motivation: Detecting and Measuring Discontinuities





Existing methods for detecting discontinuities include:

- adaptive triangle mesh refinement;
- adaptive Centroidal Voronoi Tessellation (CVT);
- ENO / WENO (weakly) essentially non oscillatory methods;
- discontinuous Galerkin methods;
- polynomial annihilation.

Motivation: Detecting the Discontinuity Interface

Given no other information, we can search for the discontinuity interface by repeatedly evaluating the characteristic function $f(x)$.

Even though we think of the result as simply 0 or 1, each evaluation of the characteristic function may require an expensive simulation.

A **Monte Carlo** or **random sampling** approach is easy to set up, but suffers from slow convergence does not take advantage of any structure in the problem, and can return unstructured results.

A method which is adaptive, and produces organized information, is the **hierarchical adaptive sparse grid** (HASG) method [Zenger 1991, Griebel 1998]; this produces a sparse grid around the discontinuity surface, which can be refined as desired, and provides a good convergence rate.

HASG: Hierarchical Adaptive Sparse Grid

The HASG method uses 1d hierarchical subspaces:

- For each dimension $n = 1, \ldots, N$, define $V_n := L^2(\Gamma_n)$
- The desired approximation is based on a sequence of subspaces $\{V_{i_n}\}_{i_n=0}^{\infty}$ of V_n of increasing dimension M_{i_n} which is dense in V_n, i.e., $\cup_{i_n=0}^{\infty} V_{i_n} = V_n$, and nested:

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_{i_n} \subset V_{i_n+1} \subset \cdots \subset V_n$$

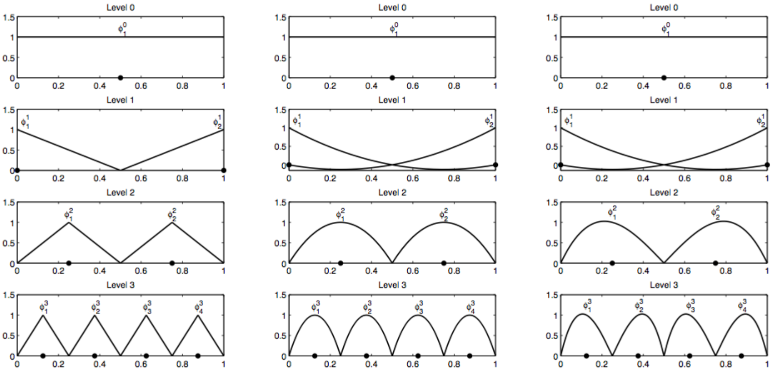
Take V_{i_n} to be the span of a nodal piecewise polynomial basis of order p:

$$V_{i_n} = \text{span}\{\phi_{j_n}^{i_n}(x_n) \mid 0 \leq j_n \leq 2^{i_n}\},$$

where i_n is the scaling level of the basis functions $\phi_{j_n}^{i_n}$ with compact support, i.e., $\text{supp}(\phi_{j_n}^{i_n}) = O(2^{-i_n})$ and $\phi_{j_n}^{i_n}(x_n)$ is a polynomial of degree p.

HASG: 1D Basis Function Hierarchy

Piecewise linear, quadratic, and cubic subspace sequences:



These functions are discussed in [Bungartz, Griebel 2004].

HASG: Multidimensions

In the multi-dimensional case, define $\mathcal{V}^N := L^2(\Gamma)$.
Construct a sequence of subspaces $\{\mathcal{V}_L^N\}_{L=0}^{\infty}$ of $\mathcal{V}^N$ using

$$\mathcal{V}_L^N = \bigcup_{g(\mathbf{i}) \le L} \bigotimes_{n=1}^{N} V_{i_n} = \bigcup_{g(\mathbf{i}) \le L} \operatorname{span}\left\{ \prod_{n=1}^{N} \phi_{j_n}^{i_n}(x_n) \;\middle|\; 0 \le j_n \le 2^{i_n} \right\},$$

where

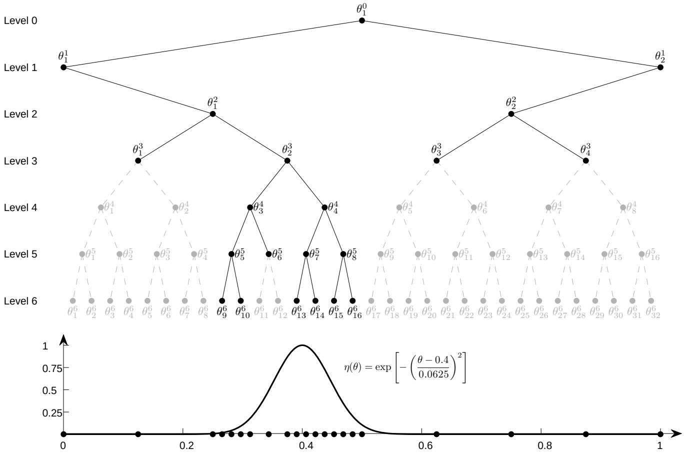
- $\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{N}_+^N$ is a multi-index;
- $g(\mathbf{i}) \le L$ defines the resolution of the approximation in $\mathcal{V}_L^N$.

Our choice of $g(\mathbf{i})$ can select various subspaces:

- Full tensor product hierarchical subspace: $g(\mathbf{i}) = \max_{n=1\ldots,N} i_n$
- Isotropic sparse hierarchical subspace: $g(\mathbf{i}) = |\mathbf{i}| \equiv i_1 + \cdots + i_N$

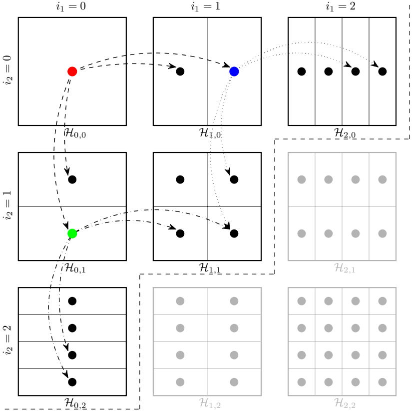
HASG 1D Example: Level 6 Adaptive Interpolation

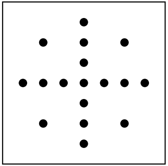
This adaptive grid has 21 points; a full grid would have used 65 points in order to achieve the same resolution (**all** points on level 6)



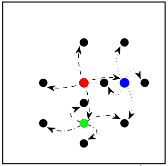
HASG 2D Example: Level 0, 1, 2 Grids

In a 2D adaptive sparse grid, points where a large surplus is detected (red, then green and blue) spawn 2 "children" in X and 2 in Y. The adaptive grid uses 12 points, where the isotropic sparse grid uses 17.





Isotropic sparse grid $\mathcal{H}_2^2$



Adaptive sparse grid $\widehat{\mathcal{H}}_2^2$

HASG: Convergence

Given a function $f(\mathbf{x})$ with bounded mixed derivatives up to order p, the convergence rate is:

$$\mathcal{O}(h^{p+1}(\log h^{-1})^{N-1})$$

where p the is order of the polynomial basis [Bungartz, Griebel 2004].

For discontinuous functions, HASG can incur very high costs, even in dimensions as low as 4, because:

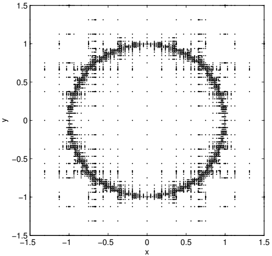
- The sparse-grid interpolant does not converge in L^∞ norm, which means the surplus does not decay to zero.
- The adaptivity generates a dense grid around the surface.
- Many grid points do not contribute much to the approximation.
- We don't use high-order hierarchical basis functions.

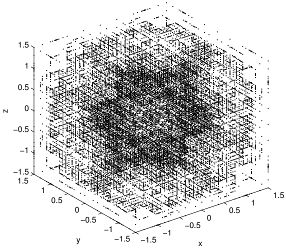
HASG: Examples in 2D and 3D

We approximate a characteristic function $f(\mathbf{x})$ with $\mathbf{x} = (x_1, \ldots, x_N)$ as

$$f(\mathbf{x}) = \begin{cases} 1, & \sqrt{x_1^2 + \cdots + x_N^2} \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The 2D adaptive sparse grid requires 5,925 grid points; the 3D grid uses 21,501 grid points.





HASG: 2D/3D/4D Error Comparison with Monte Carlo

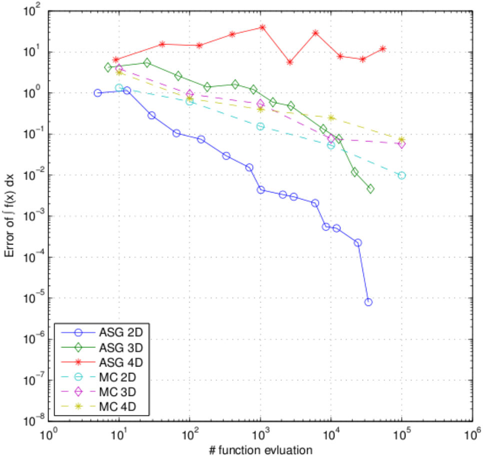


Figure: The error is measured by $\left|\int f(x)dx - \int \mathcal{I}_L^N(f)(x)dx\right|$

HYPER: A Hyperspherical Approach to HDDD

Consider again the characteristic function $f(\mathbf{x})$ of an N-dimensional argument restricted to some hyperrectangle.

- We seek to capture the discontinuity surface, an N-1-dimensional manifold, by building upper and lower bound surfaces.
- We want to estimate the integral $\int f(\mathbf{x})d\mathbf{x}$ accurately.

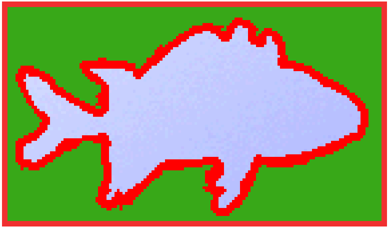
A Possible Approach:

- If the discontinuity surface of $f(\mathbf{x})$ is smooth or at least continuous, building an approximate manifold could be more efficient than approximating $f(\mathbf{x})$ directly.
- Then we could transform the manifold in N-dimensional space to a function in an N-1-dimensional subspace by a hyper-spherical transformation, build an adaptive sparse grid approximation for the manifold, then transform it back to the original Cartesian system.

HYPER: Assumptions

We make assumptions about the characteristic function $f(\mathbf{x})$:

- The discontinuity surface is an N-1 dimensional closed manifold, or can be approximated so;
- The discontinuity surface consists of a single connected component;
- The manifold is convex, or at least a substantial portion of it forms a convex set.
- An "observation" point is given, or can be found, which lies inside the manifold, and from which all, or substantially all, of the manifold can be "seen" directly. (Such a point could be encountered during the analysis of the full region, or else determined by random sampling.)



HYPER: Hyperspherical Coordinate System

A hyperspherical coordinate system is a generalization of the 2D polar and 3D spherical coordinate systems:

- One radial coordinate r;
- One angular coordinate θ_1 ranging over $[0, 2\pi)$;
- $N-2$ angular coordinates $\theta_2, \ldots, \theta_{N-1}$ ranging over $[0, \pi)$.

We can convert hyperspherical coordinates to Cartesian coordinates:

$$
\begin{aligned}
x_1 &= r\cos(\theta_1) \\
x_2 &= r\sin(\theta_1)\cos(\theta_2) \\
x_3 &= r\sin(\theta_1)\sin(\theta_2)\cos(\theta_3) \\
&\vdots \\
x_{N-1} &= r\sin(\theta_1)\cdots\sin(\theta_{N-2})\cos(\theta_{N-1}) \\
x_N &= r\sin(\theta_1)\cdots\sin(\theta_{N-2})\sin(\theta_{N-1})
\end{aligned}
$$

HYPER: Partition of the Domain

Now consider a bounded domain $\Gamma \subset \mathbb{R}^N$ and a characteristic function $f(\mathbf{x}) : \Gamma \to \mathbb{R}$ defined by:

$$f(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in D \subset \Gamma, \\ 0, & \text{otherwise} , \end{cases}$$

where D is a closed, convex domain, and ∂D is the discontinuity surface.

Our goal is to find two bounded domains D_{up} and D_{low} such that

- $D_{low} \subset D \subset D_{up} \subset \Gamma$
- $\text{dist}\left(\partial D_{up}, \partial D_{low}\right) \leq \varepsilon$

where ε is a prescribed accuracy.

It is easy to see that $f(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial D_{up}$ and $f(\mathbf{x}) = 1$ for $\mathbf{x} \in \partial D_{low}$.

HYPER: Algorithm (part 1)

- Random sampling in Γ to find $\mathbf{x}_0 \in D \subset \Gamma$ such that $f(\mathbf{x}_0) = 1$.
- Transform the Cartesian coordinates $x_1, \ldots, x_N$ to the hyperspherical coordinates $r, \theta_1, \ldots, \theta_{N-1}$ with origin $\mathbf{x}_0$.
- Due to the convexity of the domain D, ∂D can be represented by a function $r = g(\theta)$ on the bounded N-1 dimensional sub-domain

$$\Gamma_\theta = \prod_{n=1}^{N-1} [0, \pi] \times [0, 2\pi]$$

where for any $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{N-1}) \in \Gamma_\theta$, $(g(\theta), \theta)$ is on ∂D.

- Build an L-level sparse grid $\mathcal{H}_L^{N-1}$ on Γ_θ with M grid points.

$$\mathcal{H}_L^{N-1} = \{\theta_i \in \Gamma_\theta,\ \text{for } i = 1, \ldots, M\}$$

HYPER: Algorithm part 2

- At grid point $\theta_i \in \mathcal{H}_L^{N-1}$ for $i = 1, \ldots, M$, i.e. along the direction corresponding to θ_i, find two values g_i^{low} and g_i^{up} using the bisection method with accuracy tolerance ε, such that

$$g_i^{low} \leq g(\theta_i) \leq g_i^{up} \quad \text{and} \quad |g_i^{low} - g_i^{up}| \leq \varepsilon$$

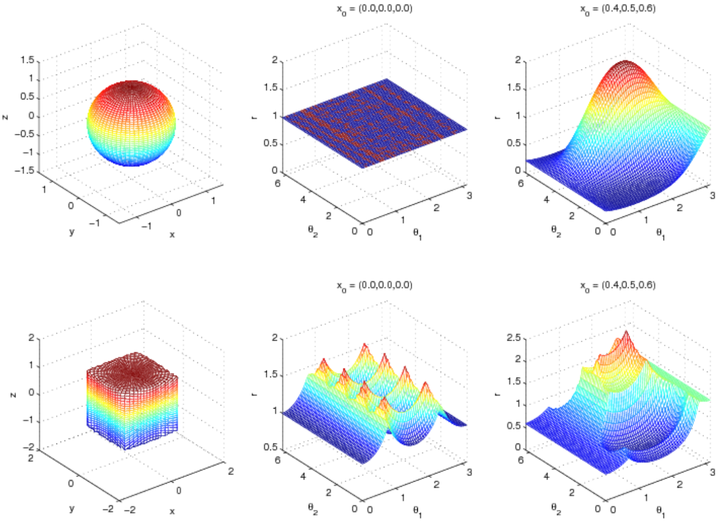
- Build sparse-grid interpolants $g^{low}(\theta)$ and $g^{up}(\theta)$ based on $\{g_i^{low}, i = 1, \ldots, M\}$ and $\{g_i^{up}, i = 1, \ldots, M\}$, respectively. Then we have

$$(g^{up}(\theta), \theta) \quad \Longrightarrow \quad \partial D_{up}$$
$$(g^{low}(\theta), \theta) \quad \Longrightarrow \quad \partial D_{low}$$

- # function evaluations = # sparse-grid points × # bisection trials

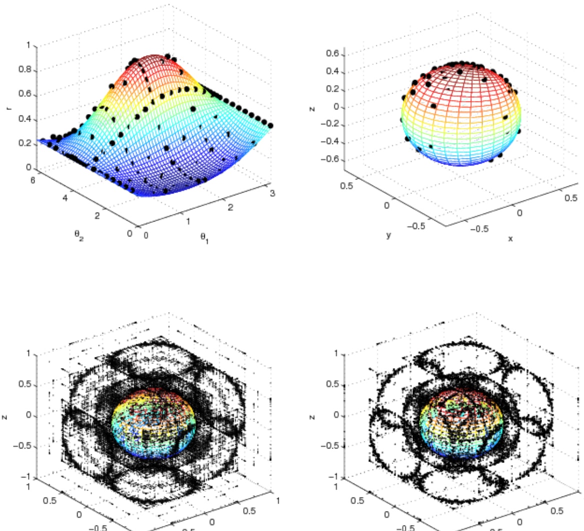
- According to smoothness of the interface ∂D, different types of basis functions can be used.

Surface in 3D, manifold with central or off-centered origin:



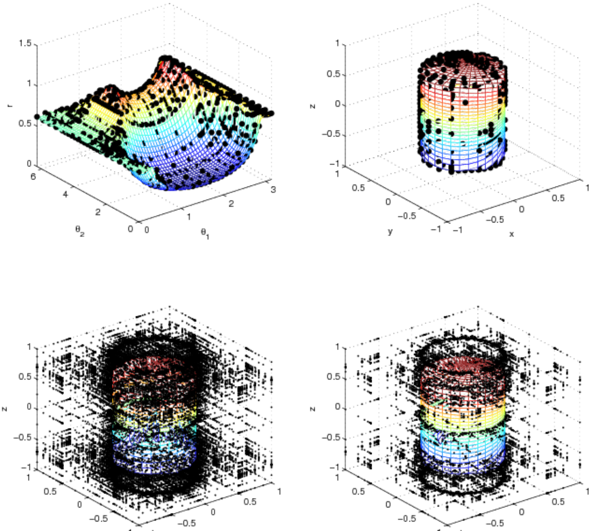
HYPER: Compare HS-SG and HASG for Unit Ball

For the unit ball, the hyperspherical approach use 137 grid points, the HASG approach 44,921 grid points, or 22,631 in the compressed version.



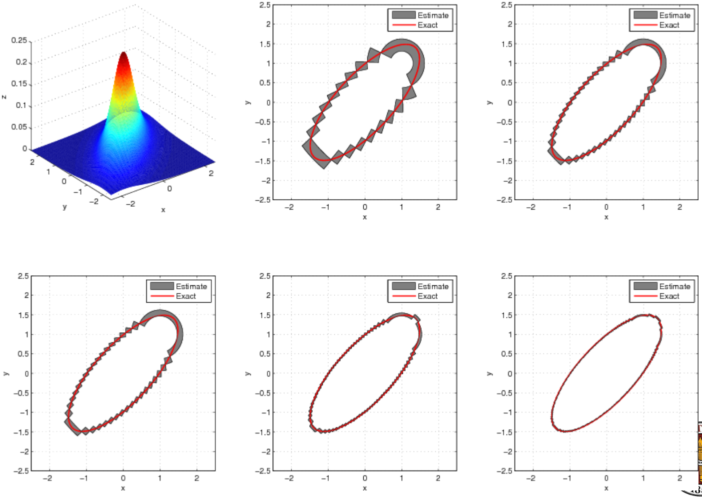
HYPER: Compare HS-SG and HASG for Cylinder

For a cylinder, the hyperspherical approach use 283 grid points, the HASG approach 37,713 grid points, or 15,407 in the compressed version.



HYPER: Seeking the region where PDF(X) $>$ tol

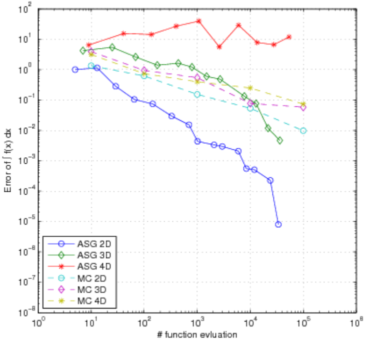
For a PDF of a 2D argument, we seek the region where pdf(x,y) $>$ 0.4. By tightening the tolerance of the bisection solver, we refine our estimate of the “discontinuity” region.

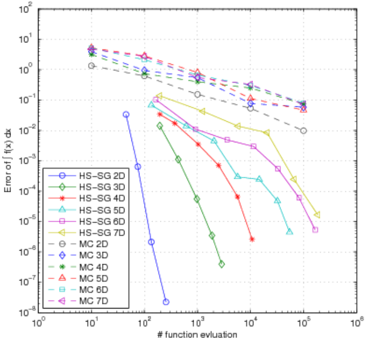


We approximate a characteristic function $f(\mathbf{x})$ with $\mathbf{x} = (x_1, \ldots, x_N)$ as

$$f(\mathbf{x}) = \begin{cases} 1, & \sqrt{x_1^2 + \cdots + x_N^2} \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

where $x_0 = (0.3, 0.3, \ldots, 0.3)$.





Summary

- Our study is motivated by UQ applications, such as quantifying rare events and risk assessment.
- Conventional adaptive sparse grid methods incur an explosive cost growth for discontinuity detection, even in low dimensions;
- Under some assumptions, we build sparse grid approximations in a hyperspherical coordinate system that characterize the high-dimensional discontinuity surface.
- Bounding surfaces capture the shape of the discontinuity surface.
- The convergence rate is significantly improved.
- G. Zhang C. Webster, M. Gunzburger, J. Burkardt,
  A hyper-spherical sparse grid approach for high-dimensional discontinuity detection,
  ORNL Tech Report, submitted: Mathematics of Computation, 2014.

THANK YOU!

References

- G. Zhang C. Webster, M. Gunzburger, J. Burkardt, *A hyper-spherical sparse grid approach for high-dimensional discontinuity detection*, ORNL Technical Report, submitted to Mathematics of Computation, 2014.
- H. Bungartz and M. Griebel, *Sparse grids*, Acta Numerica, 2004.
- M. Griebel, *Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences*, Computing, 1998.
- H. Bungartz and S. Dirnstorfer *Multivariate quadrature on adaptive sparse grids*, Computing, 2003
- M. Hegland, *Adaptive sparse grids*, ANZIAM Journal, 2003
- M. Gunzburger, C. Webster and G. Zhang, *An adaptive wavelet stochastic collocation method for irregular solutions of PDEs with random input data,* ORNL technical report, ORNL/TM-2012/186, 2012
- M. Gunzburger, C. Webster and G. Zhang, *An adaptive high-order stochastic collocation method for high-dimensional discontinuity detection,* Int. J. of Comp. Math, 2013.
- F. Nobile, R. Tempone and C. Webster, *A sparse grid stochastic collocation method for PDEs with random input data*, SIAM J. Numer. Anal., 2008
- D. Pflüger, B. Peherstorfer, H. Bungartz, *Spatially adaptive sparse grids for high-dimensional data-driven problems*, Journal of Complexity, 2010