Extending adaptive sparse grids for stochastic collocation to hybrid architectures

Rick Archibald[1](archibaldrk@ornl.gov), Ralf Deiterding[1], Cory Hauck[1], John Jakeman[2], Dongbin Xiu[2]
[1]Computer Science and Mathematics Division, Oak Ridge National Laboratory, [2]Department of Mathematics, Purdue University

We are developing an adaptive sparse grid library tailored for emerging architectures that will allow the solution of very large stochastic problems. In here, we give a brief overview of the problem at hand and present first results for a GPU-based small cluster.

Stochastic collocation

In a stochastic simulation, one is interested in the relationship between the input random variables $\mathbf{Z}$ and the outputs g of the simulation state $u = u(\mathbf{x}, t; \mathbf{Z})$ that depends on $\mathbf{Z}$ and the deterministic variables x and (possibly) time t. The mapping from $\mathbf{Z}$ to $g(u)$ can be given by

$$\mathbf{Z} \mapsto G(\mathbf{Z}; \mathbf{x}, t) \triangleq g(u(\mathbf{x}, t; \mathbf{Z})). \quad (1)$$

The key idea behind stochastic collocation (SC) is to select a set of nodes in the random space and then conduct repetitive *deterministic simulation* at each node.
Denoting with $\{G(Z_{i,j})_{j=1}^{m_i}\}$ the deterministic approximation at discrete points, we can approximate the one-dimensional component of the solution u over the range of Z_i by

$$\mathcal{G}_i[G] = \sum_{j=1}^{m_i} G(Z_{i,j}) \cdot \psi_{i,j}(Z_i) \quad (2)$$

where $\psi_{i,j}$ is the interpolating basis. Using tensor products, the entire space can be constructed as

$$\mathcal{G}_l = \mathcal{G}_{i_1} \otimes \cdots \otimes \mathcal{G}_{i_d}. \quad (3)$$

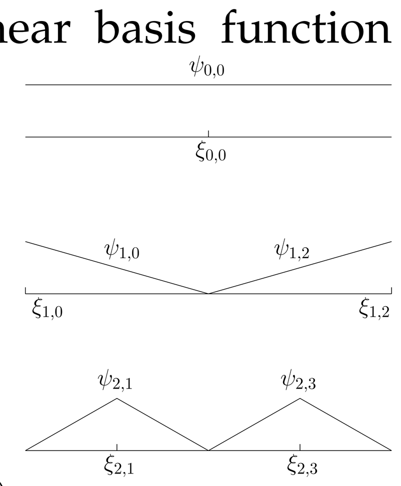
Yet, Eq. (3) suffers from the *curse of dimensionality*. It can be delayed by employing sparse grids that are based on the Smolyak construction

$$\mathcal{G}_l = \sum_{l-d+1 \le |\mathbf{i}|_1 \le l} (-1)^{l-|\mathbf{i}|_1} \cdot \binom{d-1}{l-|\mathbf{i}|_1} \cdot (\mathcal{G}_{i_1} \otimes \cdots \otimes \mathcal{G}_{i_d}) . \quad (4)$$

Sparse grids with adaptation

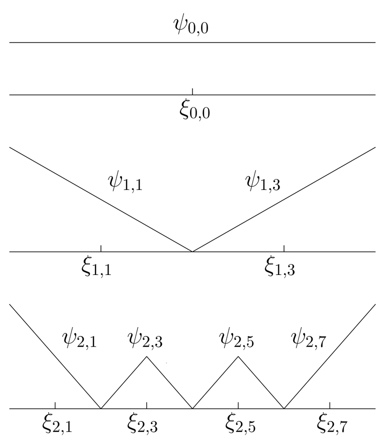
The types of basis functions $\psi_{i,j}(\xi)$ are dependent on the type of one-dimensional grids employed. Using $\psi(\xi) = \max(1-|\xi|, 0)$, we define a linear basis function

$$\psi_{i,j}(\xi) = \begin{cases} 1 & \text{if } i = 0, \\ \max(1-2\xi, 0) & \text{if } i = 1 \wedge j = 0, \\ \max(2\xi-1, 0) & \text{if } i = 1 \wedge j = 2, \\ \psi(2^i\xi - j) & \text{otherwise,} \end{cases} \quad (5)$$



and a modified linear basis function $\psi_{i,j}(\xi)$

$$= \begin{cases} 1 & \text{if } i = 0, \\ \max(2-2^{i+1}\xi, 0) & \text{if } i > 0 \wedge j = 1, \\ \max(2^{i+1}\xi - j + 1, 0) & \text{if } i > 0 \wedge j = 2^{i+1}-1, \\ \psi(2^{i+1}\xi - j) & \text{otherwise.} \end{cases} \quad (6)$$



A higher-order basis function can be defined with Lagrange polynomials with local support as

$$\psi_{i,j}^{(p)}(\xi) = \begin{cases} \Pi_{k=0}^{p} \frac{\xi-\xi_k}{\xi_{i,j}-\xi_k} & \text{if } |\xi - \xi_{i,j}| < h_i, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The 1D basis functions are used to form a set of d-dimensional basis functions

$$\psi_{\mathbf{i},\mathbf{j}}(\mathbf{Z}) = \prod_{n=1}^{d} \psi_{i_n,j_n}(Z_n) . \quad (8)$$

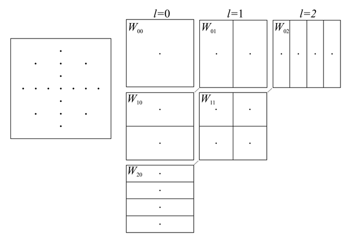
Utilizing these basis functions, an equivalent formulation to the one by Smolyak, Eq. (4), is to construct an interpolant hierarchically, that is

$$G_{l,d}(\mathbf{Z}) = \sum_{|\mathbf{i}|_1 \le l} g_{\mathbf{i}}(\mathbf{Z}), \quad g_{\mathbf{i}}(\mathbf{Z}) = \sum_{\mathbf{j} \in B_{\mathbf{i}}} v_{\mathbf{i},\mathbf{j}} \cdot \psi_{\mathbf{i},\mathbf{j}}(\mathbf{Z}) \in W_{\mathbf{i}} , \quad (9)$$

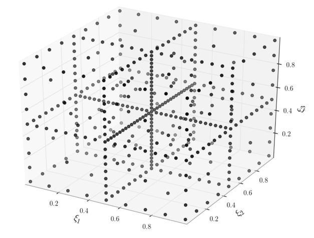
where

$$B_{\mathbf{i}} = \{j_n = 1, \ldots, m_{i_n}, \, j \text{ odd}, \, n = 1, \ldots, d\}, \quad (10)$$

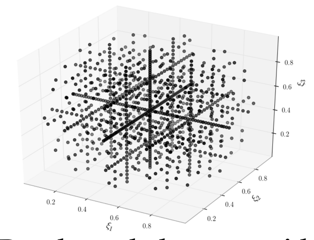
and $v_{\mathbf{i},\mathbf{j}}$ is known as the hierarchical surplus.



2D unbounded sparse grid and corresponding subspaces.



3D bounded sparse grid based on Eq. (5)



3D unbounded sparse grid based on Eq. (6)

The sparse grid approach splits up the interpolant into contributions from hierarchical difference spaces

$$W_{\mathbf{i}} = \text{span}\{\psi_{\mathbf{i},\mathbf{j}} \,|\, \mathbf{j} \in B_{\mathbf{i}}\} . \quad (11)$$

Since the interpolation subspaces are constructed hierarchically, adaptation is provided naturally for sparse grids. Local adaptation is implemented by manipulating the sets $B_{\mathbf{i}}$. An intrinsic refinement indicator is $|v_{\mathbf{i},\mathbf{j}}|$.

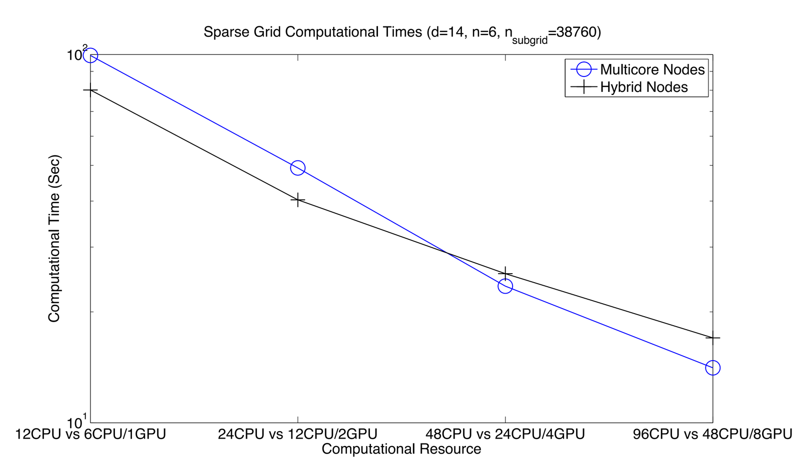
Parallelization

A coefficient transformation and interpolation code for non-adapted sparse grids was implemented as a first prototype.
Hybrid parallelization approach (MPI plus CUDA):

1. Distribute entire hierarchical subgrids $W_{\mathbf{i}}$ to nodes and thread blocks
2. Allow multiple threads to evaluate $v_{\mathbf{i},\mathbf{j}}$ according to Eq. (9) for individual points
3. Communicate $G(Z_{i,j})$ in suitable data chunks successively to all nodes and thread blocks

Configuration studied:

- Basis function (7) with Chebychev node distribution (bounded case)
- Dimension $d = 14$, levels $l = 6$
- 38,760 subgrids $W_{\mathbf{i}}$ and 1,009,905 grid points used in total
- Testbed: 16-node cluster at OLCF, each node has one 6-core Opteron 2435 CPU plus one NVIDIA Fermi GPU C2050

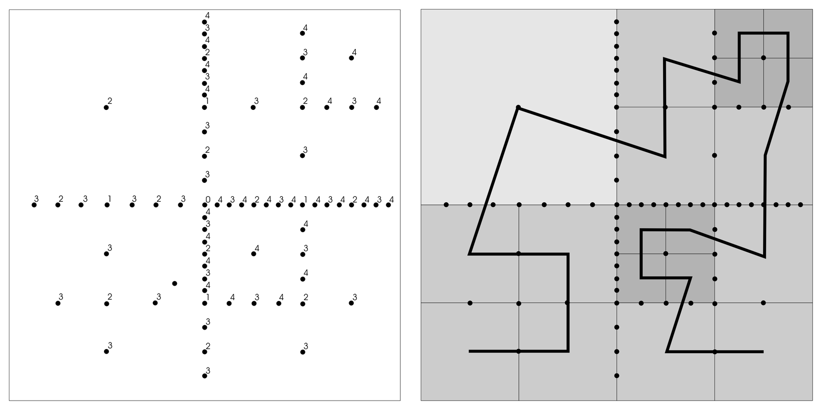


Strong scaling of sparse grid construction on the hybrid test cluster. MPI only vs. hybrid MPI/CUDA parallelization.

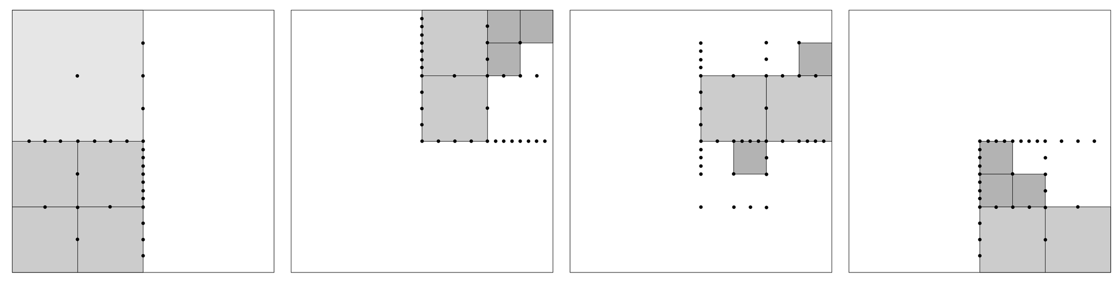
Approach for hybrid parallelization of developed prototype implementation of serial adaptive sparse grid library:

- Use geometric decomposition to ensure always strictly local evaluation of Eq. (9)
- Distribute geometry regions to nodes and thread blocks
- Again, use multiple threads to evaluate surpluses $v_{\mathbf{i},\mathbf{j}}$ and interpolate with Eq. (9) at individual points
- Use locality-preserving distribution algorithm. Suggested: generalized d-dimensional space-filling curve (SFC)

+ Cyclical communication of $G(Z_{i,j})$ and intermediate results in (9a) to all nodes is elegantly avoided
+ Serial implementation based on hash tables, including all approximation rules, can be effectively reused
- Partitioning algorithm and synchronization are complex to implement
- Determination of hierarchical points with an influence on local subregions is non-trivial



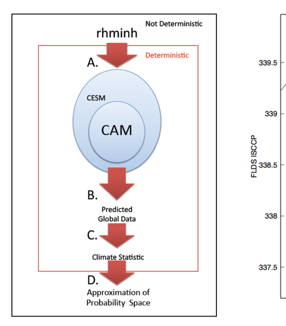
2D unbounded adapted sparse grid using 5 subspace levels. Left: each point indexed with subspace level used. Right: domains of the respective highest level and a generalized SFC used for decomposition.

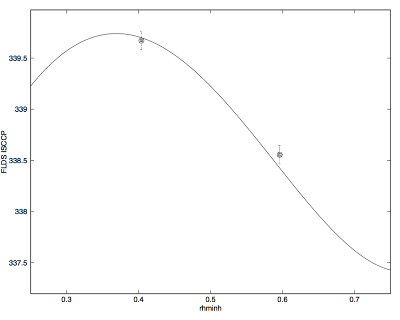


Domain decomposition of above adaptive sparse grid to four nodes and/or thread blocks based on a generalized SFC.

Application

We demonstrate how uncertainty error estimates can be derived from sparse grids to characterize parameter distributions in the community climate earth systems model (CESM).





Left: Process for UQ in the CESM. Right: UQ error estimates of CESM statistics.

Specifically, the Newton form of the residual for a m^{th} order polynomial interpolation has an equivalent form based on the local polynomial annihilation method,

$$\begin{aligned} f(y_k) - p_m(y_k) &= f[y_0, \ldots, y_k, \ldots, y_{m+1}](y_k - y_0) \ldots (y_k - y_{m+1}) \\ &= c_k q_m L_m f(y_k) \end{aligned} \quad (12)$$

This relationship allows the reconstruct of high-order global error estimator by re-purposing the edge detector which operates locally at a relatively small computational cost.

This work is sponsored by the Office of Advanced Scientific Computing Research, Applied Mathematics Research Program; U.S. Department of Energy (DOE) and was performed at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725.

References

[1] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 1–123, 2004.

[2] R. Archibald, A. Gelb, R. Saxena, and D. Xiu. Discontinuity detection in multivariate space for stochastic simulations. *J. Comput. Phys.*, 228(7):2676–2689, 2009.

[3] J. Jakeman, R. Archibald, and D. Xiu. Characterization of discontinuities in high-dimensional stochastic problems on adaptive sparse grids. *J. Comput. Phys.*, 230(10):3977–3997, 2011.

[4] M. Chakoumakos and Z. Zhuang. Characterizing the elements of earth's radiative budget: applying uncertainty quantication to climate models. *Siemens Competition*, ORHS Thesis project, 2011