This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Sandia National Laboratories

SAND2019-8116C

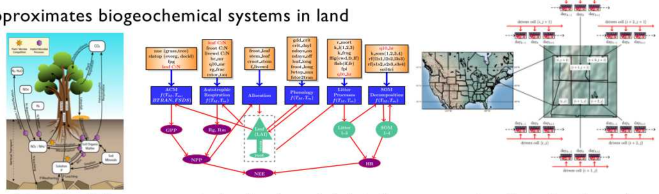
Approximating Data with Stochastic and Physical Dependence using Functional Tensor Train Models

Tim Reid[1,2], Cosmin Safta[2], Alex Gorodetsky[3], John Jakeman[4], Khachik Sargsyan[2]

1. North Carolina State University. 2. Sandia National Laboratories, Livermore, CA. 3. University of Michigan. 4. Sandia National Laboratories, Albuquerque, NM.

Motivation

- E3SM Land Model is land component of DOE Earth system model [3]
- Land Model has 47 stochastic inputs and is evaluated separately for each location and date
- Approximates biogeochemical systems in land



- Goal is using FT as surrogate for land model data for many sets of stochastic and physical inputs
- We need to incorporate physical dependence into the FT and develop optimization algorithms to fit this FT to the data.

Surrogate Modeling With Functional Tensor Train

- The Functional Tensor-Train (FT) approximates data, y, that depends on stochastic inputs, x, using the product of matrix valued functions called Tensor Cores [1]:

$$f(x) = \mathbf{F}^{(1)}(x_1)\mathbf{F}^{(2)}(x_2)\cdots\mathbf{F}^{(d)}(x_d)$$

- x has d dimensions, there is one tensor core for each dimension of x
- Dimensions of k^{th} tensor core are $r_k \times r_{k+1}$, called ranks of tensor core
- Tensor cores, $\mathbf{F}^{(k)}(x_k)$, are made of univariate functions
- Univariate functions are Polynomial Chaos Expansions (PCE) of the corresponding input [2]

$$f_{i,j}^{(k)} = \sum_{p=0}^{P_k} \theta_{k,i,l,j}\phi_p(x_k)$$

- $\theta_{k,i,l,j}$ is the coefficient and $\phi_p(x_k)$ is the degree p basis polynomial

Incorporating Physical Dependence in FT

- We want the FT to take physical and stochastic inputs separately
- Two ways to incorporate physical inputs:
  1. Treat physical input as another stochastic input
  2. Add physical dependence to coefficients in PCEs
- First method adds another tensor core for physical input p
- Second method replaces coefficients with polynomial expansions in terms of the physical input p:

$$\theta_{k,i,l,j}(p) := \sum_{s=1}^{S} \theta_{k,i,l,j,s}\psi_s(p)$$

- $\psi_s(p)$ is basis polynomial with corresponding coefficient $\theta_{k,i,l,j,s}$
- Not necessary to put physical dependence on all coefficients
- If physical dependence on all coefficients, PCE degree is $p_k = P$, and all ranks are equal to R, then there are dR^2SP coefficients to solve for

Determining Coefficients in FT

- To find the best coefficients we minimize the objective function

$$J(\theta) = \frac{1}{2}\sum_{u=1}^{V}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(y_{m,n,u} - f(x^{(m)},p^{(n)})_u\right)^2$$

- Data is for V values at M instances of the stochastic input and N instances of the physical input
- Objective function is minimized using Alternating Least Squares or gradient based methods L-BFGS and Gradient Descent

Gradient

- Partial derivative of objective function used to compute gradient is

$$\frac{\partial J}{\partial \theta_{k,i,l,j,s,v}} = \sum_{u=1}^{V}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(y_{m,n,u} - f(x^{(m)},p^{(n)})_u\right)\frac{\partial f(x^{(m)},p^{(n)})_u}{\partial \theta_{k,i,l,j,s,v}}$$

- Where

$$\frac{\partial f(x^{(m)},p^{(n)})_u}{\partial \theta_{k,i,l,j,s,v}} = \mathbf{F}^{(1)}\cdots\mathbf{F}^{(k-1)}\mathbf{G}\mathbf{F}^{(k+1)}\cdots\mathbf{F}^{(d)}$$

- And

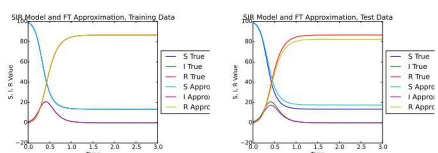
$$g_{w,z} = \begin{cases}\phi_l^{(k)}(x^{(m)})\psi_s(p^{(n)}) & (w,z)=(i,j)\\ 0 & (w,z)\neq(i,j)\end{cases}$$

Test Functions

- Test the FT by approximating the output of two types of test functions:
  1. Generating a random FT and using a different FT to approximate it
  2. Approximating system of ODEs (SIR Model and Lorenz System[4,5]) with parameters as stochastic input and time as physical input
- Stochastic inputs for ODE tests are sampled from a normal distribution centered at a nominal parameter value
- We fit the FT to data from the function and then test how well the FT approximates data with different stochastic and physical inputs

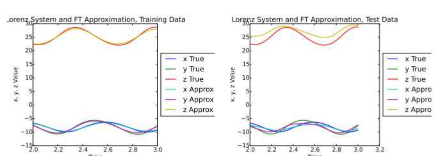
Approximating a Random FT
Trained and Tested with
100 Stochastic Samples, 100 Time Samples

	2 Norm Error
Training Data	9.30×10^{-4}
Test Data	3.41×10^{-3}



SIR Model
Trained and Tested with
25 Stochastic Samples, 100 Time Samples

	2 Norm Error
Training Data	3.12×10^{-3}
Test Data	7.56×10^{-2}



Lorenz System
Trained and Tested with
50 Stochastic Samples, 1000 Time Samples

	2 Norm Error
Training Data	1.90×10^{-2}
Test Data	1.51×10^{-1}

- Error on test data is higher than error on training data

Summary and Future Work

- FT can approximate a set of data dependent on stochastic and physical inputs
- Analytic derivative of objective function allows use of gradient based optimization
- We will apply the FT to approximate data from the E3SM Land Model
- Add more types of functions (radial basis functions, Fourier series) as options for univariate functions in tensor cores

Refrences

[1] Gorodetsky, A. A., & Jakeman, J. D. (2018). Gradient-based optimization for regression in the functional tensor-train format. *Journal of Computational Physics*, 374, 1219-1238. https://doi.org/10.1016/j.jcp.2018.08.010

[2] Najm, H. N. (2009) Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, 41, 35-52, doi:10.1146/annurev.fluid.010908.165248

[3] https://e3sm.org

[4] Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* 115(772), 700-721. https://doi.org/10.1098/rspa.1927.0118

[5] Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2), 130-141. doi:0.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. ___________

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), Office of Biological and Environmental Research (BER), and Scientific Discovery through Advanced Computing (SciDAC) program.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.