The following is a summary of the my work for week 6 after meeting with Shantenu on Wednesday June 24th

General:
- Skimmed The Structure of Scientific Revolutions [1]
- Read (50%) Wikipedia article on Paradigms [2]
- Read (70%) Wikipedia article on Philosophy Of Science [3]
- Read (70%) Wikipedia article on Falsifiability [4]

Smart surrogates:
My goal is to understand the different types, properties and challenges of smart surrogates.
- Read(100%) about surrogate models in order to better understand surrogacy[14]
- Read (80%) the paper "KERNEL: Enabler to build smart surrogates for online optimization and knowledge discovery"
- Read (50%) the paper "Artificial Neural Network based surrogate modelling for multi- objective optimisation of geological CO2 storage operations"
- Read (50%) the paper "A Deep Neural Network Surrogate for High-Dimensional Random Partial Differential Equations"
- Read (50%) the paper "Improved surrogates in inertial confinement fusion with manifold and cycle consistencies" (use of inverse surrogates)

EnTK:
- Understood strong scaling and weak scaling

A-Strong scaling:

Amdahl’s law: speedup is limited by the fraction of the serial part of the software

speedup = 1 / (s + p/ N)

B- Weak Scaling:

Gustafson’s law: the parallel part scales linearly with the amount of resources, and the serial part does not increase with respect to the size of the problem.

scaled speedup =s+p*N

s: proportion of execution time spent on the serial part,

p: proportion of execution time spent on the part that can be parallelized,

N: number of processors

- Designed an experiment to test the strong and weak scaling properties of EnTK. Please see the code in strongScaling.py and weakScaling.py

A- Strong Scaling: To investigate strong scaling, I run 7 applications, each consists of 1 pipeline containing two stages at 5 tasks/stage. Each task sleeps for 100s. The number of cores increases for every application (1,2,4,8,16,32,64).

B- Weak Scaling: To investigate weak scaling, I run 7 applications, each consists of 1 pipeline containing two stages at variant number of tasks/stage. Each task sleeps for 100s. The number of cores increases for every application and equals the number of tasks per stage. The number of cores (and tasks/stage) for the applications are 1,2,4,8,16,32,64.

For the strong scaling experiment, I expect TTX to decrease linearly with the increase in the number of cores. This is because there are more available resources for the fixed number of tasks.

For the weak scaling experiment, I expect TTX to increase non-linearly and gradually because EnTK does not have ideal weak scaling. This is due to delays in the Executor module of the RTS Agent[6]

After plotting the number of cores vs speadup for both experiments, do curve fitting and find the fitted value for the serial fraction s. I expect s to be 0.2 for Amdahl's law and [0.015-0.5] for Gustafson's law

Resources:

[1] https://www.lri.fr/~mbl/Stanford/CS477/papers/Kuhn-SSR-2ndEd.pdf

[2] https://en.wikipedia.org/wiki/Paradigm

[3] https://en.wikipedia.org/wiki/Philosophy_of_science

[4] https://en.wikipedia.org/wiki/Falsifiability

[5] https://www.kth.se/blogs/pdc/2018/11/scalability-strong-and-weak-scaling/#:~:text=Strong%20scaling%20concerns%20the%20speedup,is%20governed%20by%20Gustafson's%20law.

[6] Harnessing the Power of Many: Extensible Toolkit for Scalable Ensemble Applications: https://arxiv.org/pdf/1710.08491.pdf

[7] https://www.geeksforgeeks.org/julia-fractal-python/

[8] https://www.sharcnet.ca/help/index.php/Measuring_Parallel_Scaling_Performance#Table

[9] https://www.pnas.org/content/117/18/9741
[10] https://arxiv.org/pdf/1806.02957.pdf
[11] https://www.sciencedirect.com/science/article/pii/S1876610214021924
[12] https://www.tandfonline.com/doi/abs/10.1080/10426914.2016.1269918
[13]https://en.wikipedia.org/wiki/Surrogate_model#Types_of_surrogate_models