Matthew Ragoza
Dr. Hessam Babaee
ME 2063 - Data-driven Modeling for Engineers
April 23, 2023

Final Project Report

In this project, we analyzed a set of transient temperature simulations on a 2D rectangular plate with a cylindrical hole. The forward heat conduction problem was solved a number of times using FEM to generate different temperature profiles with varying forcing terms. The forcing terms were determined by the angle of the heat source relative to the hole.

Part a.

After generating the full-order data set of temperature profiles, the first objective was to obtain a set of basis functions that could be used to efficiently represent the temperature simulations. I performed proper orthogonal decomposition (POD) to identify the singular values and vectors of the temperature data. The data were first reshaped from the dimensions (Nnode, Ntime, Nsample) to a matrix of shape (Nnode, Ntime $\times$ Nsample) which was then decomposed through SVD to identify POD modes and singular values common across all of the temperature profiles.

Figure 2 displays the four most dominant POD modes of the temperature simulations and Figure 1 shows the first 50 singular values, plotted on a log scale. We can see that the temperature simulations have most of their power concentrated in the most dominant modes. Specifically, we can capture over 99.9% of the power by using only the first 8 modes.

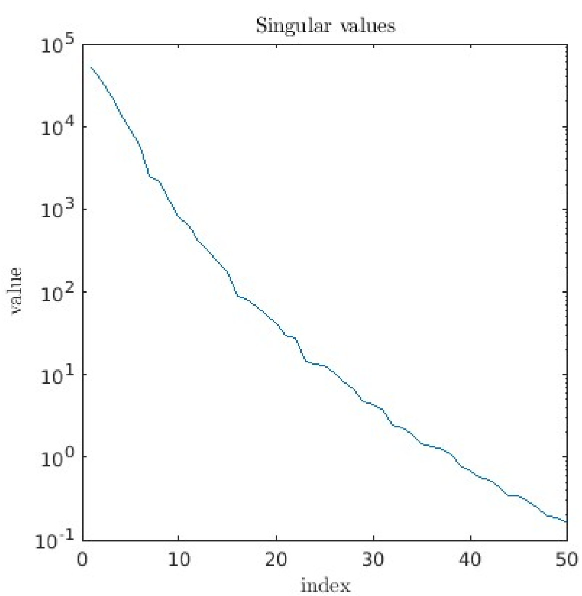

Figure 1. Top 50 singular values of the temperature simulations.

Part b.

After selecting the 8 most dominant POD modes as the basis for further modeling, we wanted to further reduce the order of the problem by identifying the locations in the spatial domain that are most informative for reconstructing the simulations using the POD basis. I applied QDEIM in order to identify the near-optimal sensor locations. This involves using pivoted QR decomposition on the transposed matrix of the POD basis, then converting the returned permutation matrix into a set of indices that represent the nodes in the FEM mesh that are most important.

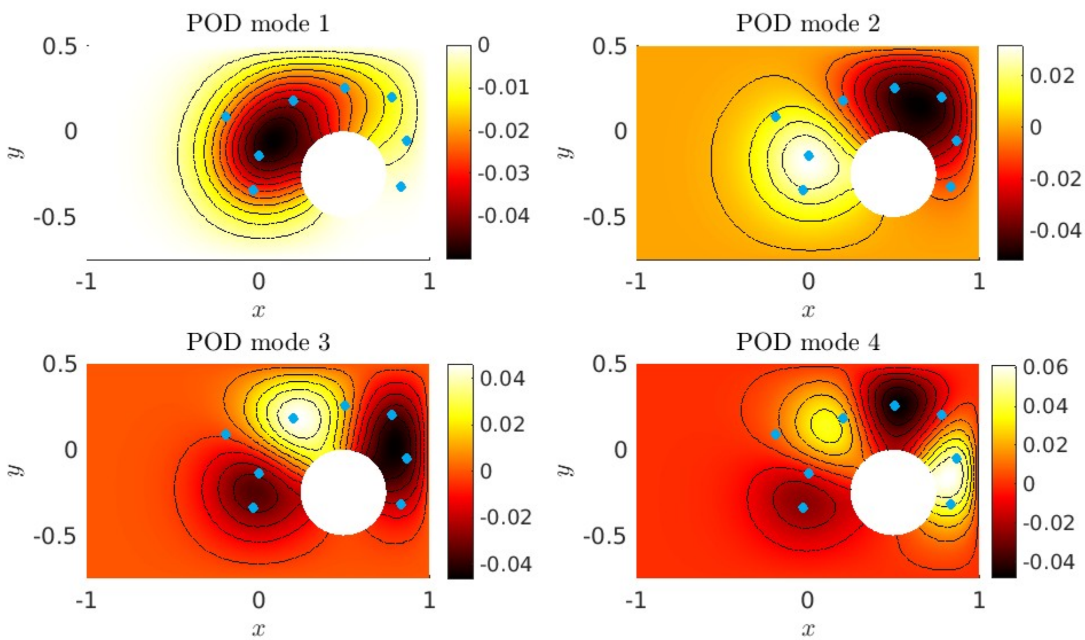

Figure 2. Four most dominant POD modes of the temperature simulations. QDEIM sensor locations are overlaid on POD modes.

Figure 2 shows the four most dominant POD modes overlaid with the optimal sensor locations identified by QDEIM. We can see that the sensors are scattered around the cylindrical hole near the locations that the heat source appears in the training data set. The sensor points also tend to have high variability across the dominant POD modes.

After identifying these sensor points, I used them to interpolate the last snapshot of the last temperature profile in the dataset into the POD basis. The true temperature distribution, interpolated distribution, and the error between them are shown in Figure 3. The mean squared error of the interpolation was around 6,950 (0.05%).

Part c.

The next task was to test the QDEIM interpolation method on novel temperature snapshots that were not used to construct the POD basis. I generated a test set by simulating the forward heat conduction problem at a new heat source position (with angle $\theta = 120$). Then, I extracted the temperature data at the previously identified QDEIM sensor locations. I interpolated the QDEIM data points of the test simulation at three different time points (t = 0.01, 0.03, 0.05) using the POD basis functions to predict temperature distributions.

The resulting interpolations are shown in Figure 4. The relative error of the interpolation for each of the three test snapshots ere 1.7%. 0.3%, and 0.21%. There was slighly more error in these interpolations than was present in the interpolation of the snapshot from Part b. This is expected because the previous snapshot was part of the training data set, which was used to extract the POD basis functions. The relative error of the method on these previously unseen temperature snapshots is still surprisingly low given that only 8 sensor locations were used to create the interpolations. This indicates that the combination of POD basis functions extracted from the full-order model with the use of QDEIM interpolation points can result in highly accurate field reconstruction from sparse data.

Part d.

In the final part of this project, we compare to an entirely different approach to surrogate modelling. The overall objective is the same: We would like to create a model than can predict the temperature distribution on the given domain for previously unseen heat sources, based on observed instances of simulated temperature profiles. This time, we want to learn the relationship between the angle of the heat source as an input, and the temperature profile as an output, as a linear combination of Legendre polynomial basis functions:

$$T(\theta; x, y, t) = \sum_i \hat{T}_i(x, y, t)\varphi_i(\theta)$$

I accomplished this by evaluating the polynomial basis functions $\varphi_i(\theta)$ at the set of angles in the training data, then solving the above system for the unknown parameters, $\hat{T}_i(x, y, t)$. In this case, I used the same number of basis functions as temperature profiles in the training data, so the linear system was exactly determined. Therefore, this corresponds to an interpolation of the training data using the polynomial basis functions, as opposed to a regression or Galerkin projection. The resulting coefficients for the first four polynomial basis functions are shown in Figure 5. Interestingly, the coefficient distributions look similar to the dominant POD modes of the temperature simulation.

Next, I tested the polynomial surrogate model by evaluating the basis functions at a test angle not present in the training data ($\theta = 120$). Then I predicted a temperature distribution by taking the linear combination of the evaluated basis functions with the polynomial coefficients. The test temperature snapshot, predicted snapshot, and error for this method are shown in Figure 6. The mean squared error was around 18,900 (0.18%) in this evaluation.

The polynomial surrogate model was more accurate than the QDEIM model in terms of relative error. However, each method has different data and compute requirements for the "offline" and "online" phases. In the offline phase, the POD + QDEIM model performs SVD on the full-order snapshot matrix (Nnodes x Ntime x Nsamples), then QR decomposition on the low-rank basis matrix (Nnodes x Nbasis). On the other hand, the polynomial model solves a (Nsamples x Nbasis) linear system for the (Nnodes x Ntime x Nbasis) polynomial basis coefficients. These amount to similar asymptotic computational complexity, as both cases are dominated by computing a matrix decomposition on the full-order simulation.

In the online phase, the goal is to compute a temperature profile for some arbitrary, previously unseen heat source angle. In the QDEIM method, this requires sparse measurements from a temperature profile from this heat source, then interpolation is performed using the POD basis to reconstruct the full field. The interpolation solves an (Npoints x Nbasis) linear system for the interpolation coefficients. In contrast, the polynomial model does not have to solve a linear system at test time, but must compute a large matrix multiplication to predict the full temperature field using the polynomial coefficients with the evaluate polynomial basis functions. Therefore, QDEIM is less computationally expensive in the online stage, but requires access to sparse measurements from the test data at the sensor locations, while the polynomial surrogate model requires no test measurements.

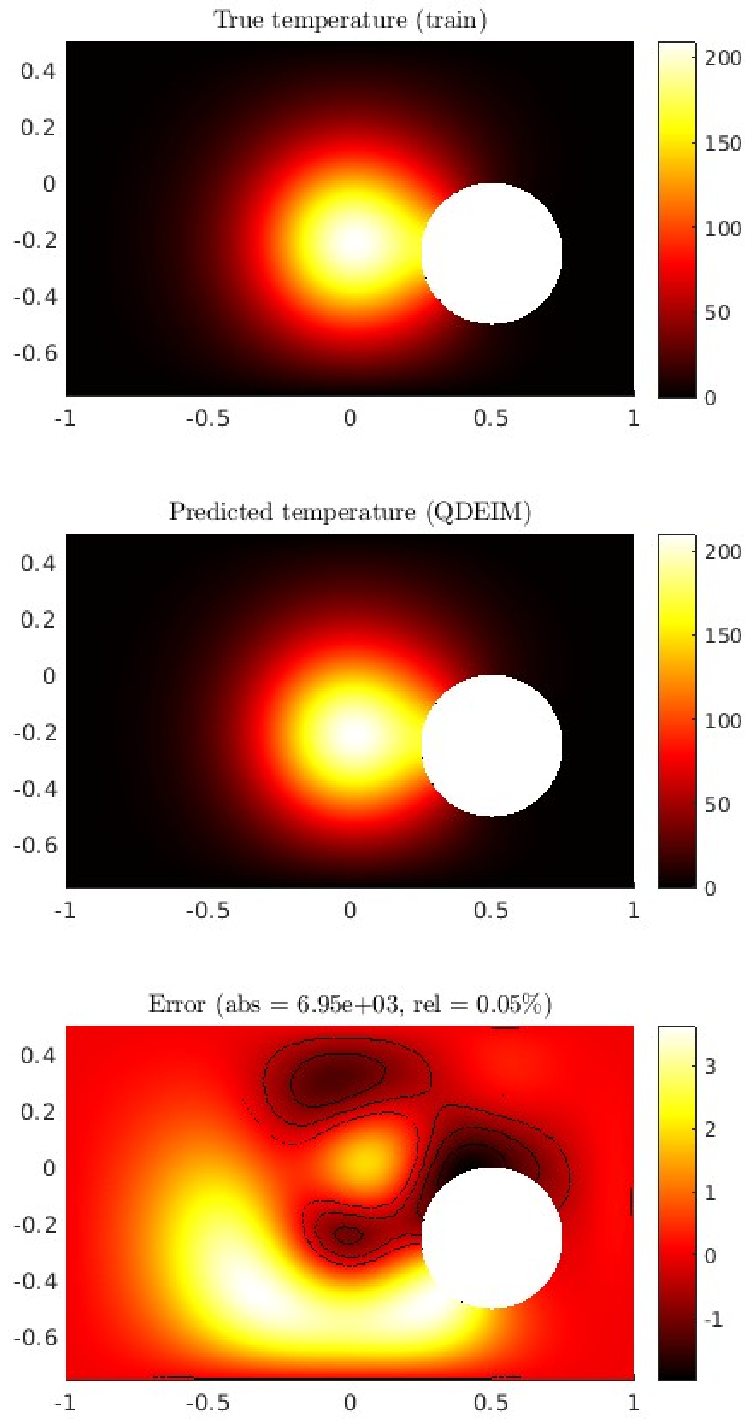

Figure 3. QDEIM interpolation of temperature snapshot from training data set.

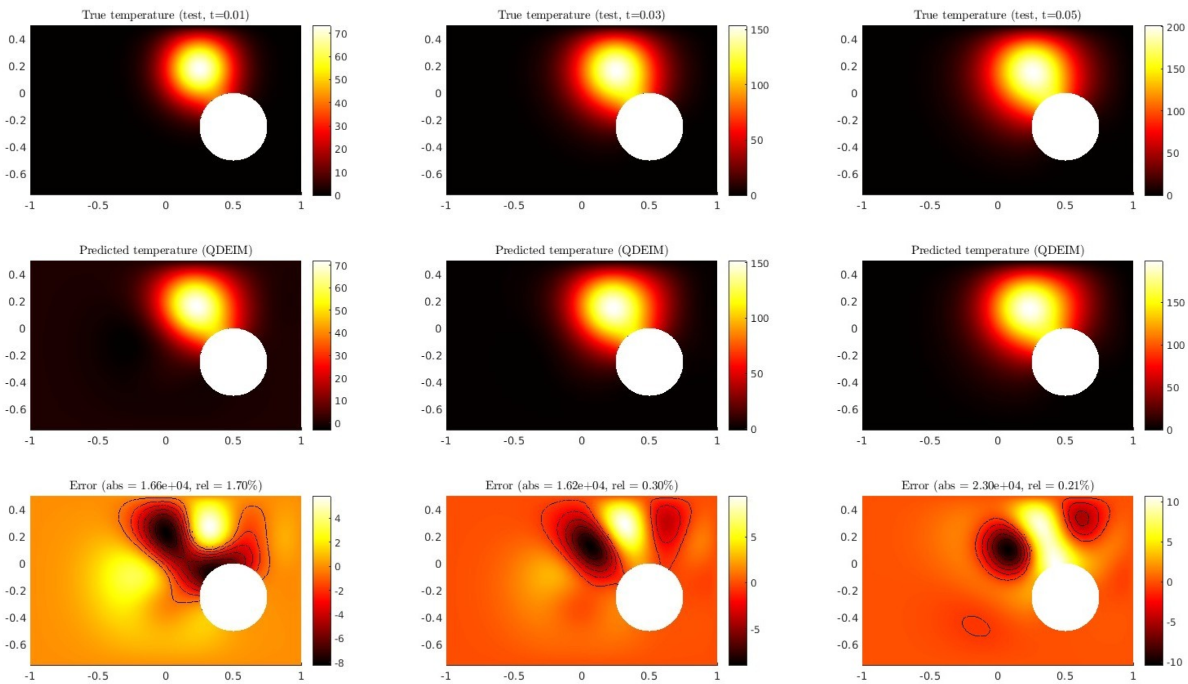

Figure 4. QDEIM interpolation of temperature snapshots from test data set.

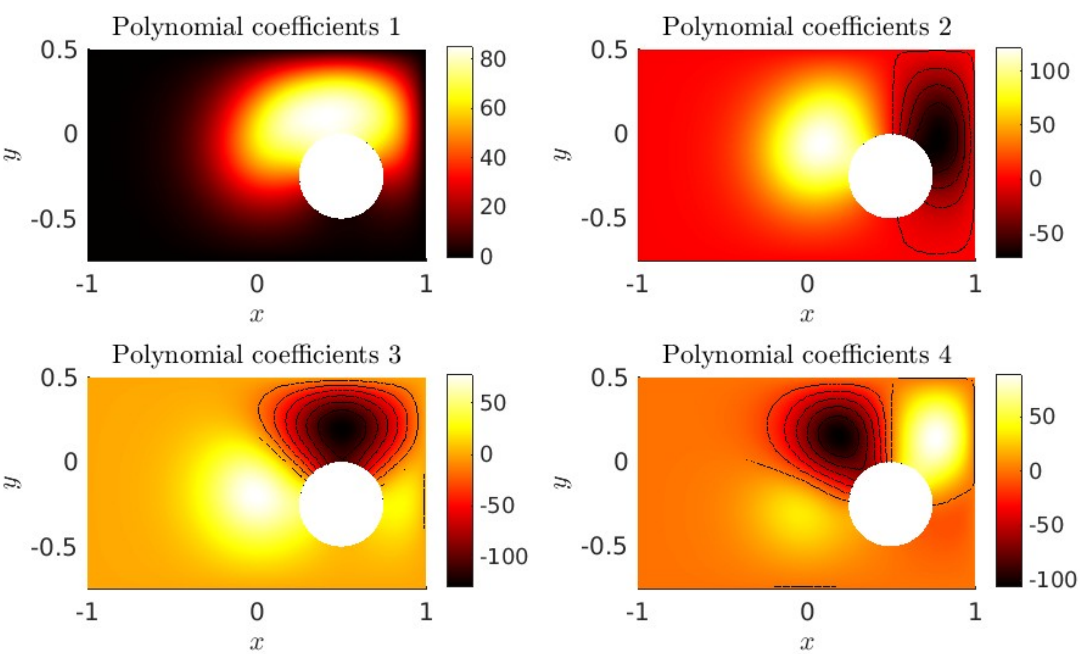

Figure 5. Coefficients of first four Legendre polynomial basis functions.

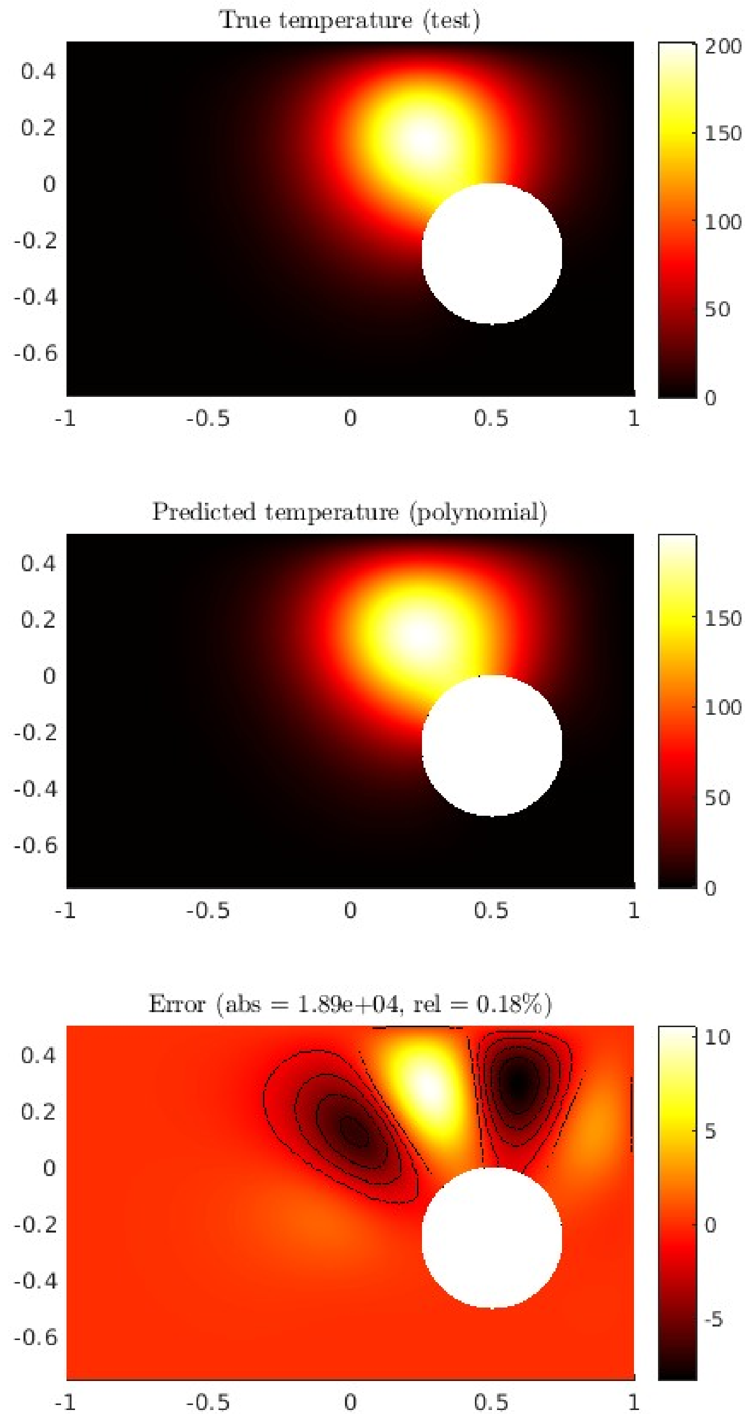

Figure 6. Evaluation of polynomial model at test heat source angle.