CANSSI / SAMSI UQ SUMMER SCHOOL
Worksheet 2 :: Exploratory Model Analysis

Consider the following model of average fluid velocity u from the Hartmann problem, which is a simplified model of a magnetohydrodynamics system:

$$u \;=\; -\frac{\partial p_0}{\partial x}\,\frac{\eta}{B_0^2}\left(1 - \frac{B_0}{\sqrt{\eta\mu}}\coth\left(\frac{B_0}{\sqrt{\eta\mu}}\right)\right) \tag{1}$$

The five parameters have interpretation, units, and ranges from the following table.

Parameter	Interpretation	Lower bound	Upper bound
μ	fluid viscosity	0.05	2.0
ρ	fluid density	1.0	5.0
$\frac{\partial p_0}{\partial x}$	pressure gradient	0.5	3.0
η	resistivity	0.5	3.0
B_0	magnetic field	0.1	1.0

1. Write a function that (i) takes a point from the *normalized* parameter space $[-1,1]^5$ (i.e., five inputs with ranges each $[-1,1]$), (ii) unnormalizes the inputs to their natural ranges (see the table), and (iii) evaluates the average velocity. The *nominal* value of the inputs is the origin.
2. Perform axis-aligned coordinate sweeps (or slices) for the five input parameters using 11 evaluations per sweep. What can you say about smoothness and monontonicity of the function? Was 11 points enough to be confident in these assessments?
3. Algorithm 1 provides a method for finding points of a random sweep. Run five random sweeps ($M = 5$) with ten points per sweep ($N = 10$). Are the insights from the random sweeps consistent with the coordinate sweeps?
4. Draw 100 points uniformly at random from the normalized domain $[-1,1]^5$. Evaluate the function at each point to create a data set of 100 (5d) input / (1d) output pairs.
5. Make 5 coordinate scatter plots of each the output versus the samples of each input. What can you see from these plots? Anything interesting?
6. Algorithm 2 gives a recipe for making an off-axis scatter plot (or shadow plot) using a direction defined by the gradient of a least-squares-fit linear model. Run this algorithm on the data set you just generated; in other words, you already did steps 1 and 2 from Algorithm 2. Do you think average velocity u can be well approximated by a univariate function of $\mathbf{w}^T\mathbf{x}$ (where $\mathbf{x}$ are the normalized inputs and $\mathbf{w}$ is the vector from Algorithm 2)? Why or why not? What function would you choose for the univariate function?
7. Using the vector $\mathbf{w}$ from the previous step (Algorithm 2), fit a univariate curve to the input / output pairs $\{\mathbf{w}^T\mathbf{x}_j, f(\mathbf{x}_j)\}$. Plot the fitted function with the data.
8. Look at the numerical values of $\mathbf{w}$. According to this vector, what are the most important parameters for the average velocity u? What are the least important or unimportant parameters? Write a complete English sentence that interprets the precise notion of importance of these vector components. Do parameters that you've identified as *important* make sense, physically? Why or why not?
9. If you have time, multiply the lower bounds by 1/100 and the upper bounds by 100, and repeat the analysis (i.e., make the plots again). Does anything change?

Algorithm 1 Random sweeps.

Given $f : [-1,1]^m \to \mathbb{R}$, decide the number M of sweeps and the number N of points per sweep. For $i = 1, \ldots, M$,

1. Draw $\mathbf{x}_i$ uniformly from $[-1,1]^m$.
2. Draw $\mathbf{u}_i$ from the unit sphere. Let $\mathcal{I} \subseteq \{1, \ldots, m\}$ be the set of indices corresponding to positive elements of $\mathbf{u}_i$, and let $\sim \mathcal{I} \subseteq \{1, \ldots, m\}$ be the set of indices corresponding to negative elements of $\mathbf{u}_i$.
3. Compute

$$\begin{aligned} h_{\min} &= \max\left(\begin{bmatrix} \frac{\mathbf{1}-\mathbf{x}_i^{\sim\mathcal{I}}}{\mathbf{u}_i^{\sim\mathcal{I}}} \\ \frac{-\mathbf{1}-\mathbf{x}_i^{\mathcal{I}}}{\mathbf{u}_i^{\mathcal{I}}} \end{bmatrix}\right) \\ h_{\max} &= \min\left(\begin{bmatrix} \frac{\mathbf{1}-\mathbf{x}_i^{\mathcal{I}}}{\mathbf{u}_i^{\mathcal{I}}} \\ \frac{-\mathbf{1}-\mathbf{x}_i^{\sim\mathcal{I}}}{\mathbf{u}_i^{\sim\mathcal{I}}} \end{bmatrix}\right) \end{aligned} \tag{2}$$

4. Compute the parameter points

$$\mathbf{x}_{i,j} = \mathbf{x}_i + h_j \mathbf{u}_i, \quad h_j = h_{\min} + j\,\Delta h, \quad j = 0, \ldots, N, \tag{3}$$

   where $\Delta h = (h_{\max} - h_{\min})/N$.
5. Compute $y_{i,j} = f(\mathbf{x}_{i,j})$.
6. Plot $y_{i,j}$ versus $h_j + \mathbf{x}_i^T \mathbf{u}_i$.

Algorithm 2 "Ordinary Least Squares" (OLS)

1. For α between 2 and 10, draw $N = \alpha m$ independent samples $\{\mathbf{x}_j\}$ uniformly from $[-1,1]^m$.
2. For each sample $\mathbf{x}_j$, run the simulation and compute the quantity of interest, $y_j = f(\mathbf{x}_j)$.
3. Use least-squares to compute the coefficients c and $\mathbf{b}$ of the linear model,

$$y_j \;\approx\; c + \mathbf{b}^T \mathbf{x}_j, \quad j = 1, \ldots, N. \tag{4}$$

4. Compute the normalized gradient of the linear model

$$\mathbf{w} \;=\; \mathbf{b}/\|\mathbf{b}\|. \tag{5}$$

5. Plot y_j versus $\mathbf{w}^T \mathbf{x}_j$.