Midterm Exam: Spring, 2022

Due Date: 3/2/2022, 2:00 PM

Dr. Abd-AlRahman AlMomani *MA601: Data-Driven Science & ML*

Problem 1 (20 points):

a. Explain the following: The values of the parameters estimated with the least-squares method are sensitive to the singular values. Explain in detail.

b. Based on your analysis in part a, suggest a solution for this problem.

c. Demonstrate your suggested solution on one simple example. Use the discrete map

$$x_{n+1} = 1.43x_n - 0.143x_n^{10}$$

as your toy example. (When creating the data, add minimal noise, SNR = 15).

Problem 2 (20 points):

Basal Metabolic Rate (BMR) is the rate of energy expenditure per unit time by endothermic animals at rest. It is reported in energy units per unit time. There are several formulas (equations) to compute the BMR. However, a simple and efficient one can be given by:

$$y = aM^b \tag{1.1}$$

where y is the BMR, a allometric coefficient, M is body mass, and b is the allometric exponent.

Assume you are given the body mass M_i and the BMR y_i for a set of observations $i = 1, ..., n$. Suggest a regression framework to estimate the parameters a and b using the least-squares method. Note that it is white-box modeling. Meaning that you know the equation form, but you wish to estimate the parameters.

Problem 3 (30 points):
Given a set of observations $\{x_i\}_{i=1}^N$, we tried to investigate the least-squares method performance with a variable number of observations. So, we created some candidate functions matrix Φ, started with one observation, and increased the number gradually. At each number of observations, we computed the L_2-norm of the least-squares fitting by:

$$E = \|\Phi\Phi^\dagger f - f\|_2$$

where E is the fitting error, and $\Phi^\dagger f$ are the estimated parameters using the least squares. The results reported in Fig.1.1.

Looking at the curve, we see that the error keeps increasing as the number of observations increases, which contradicts the idea that we should get better accuracy when using more observations.

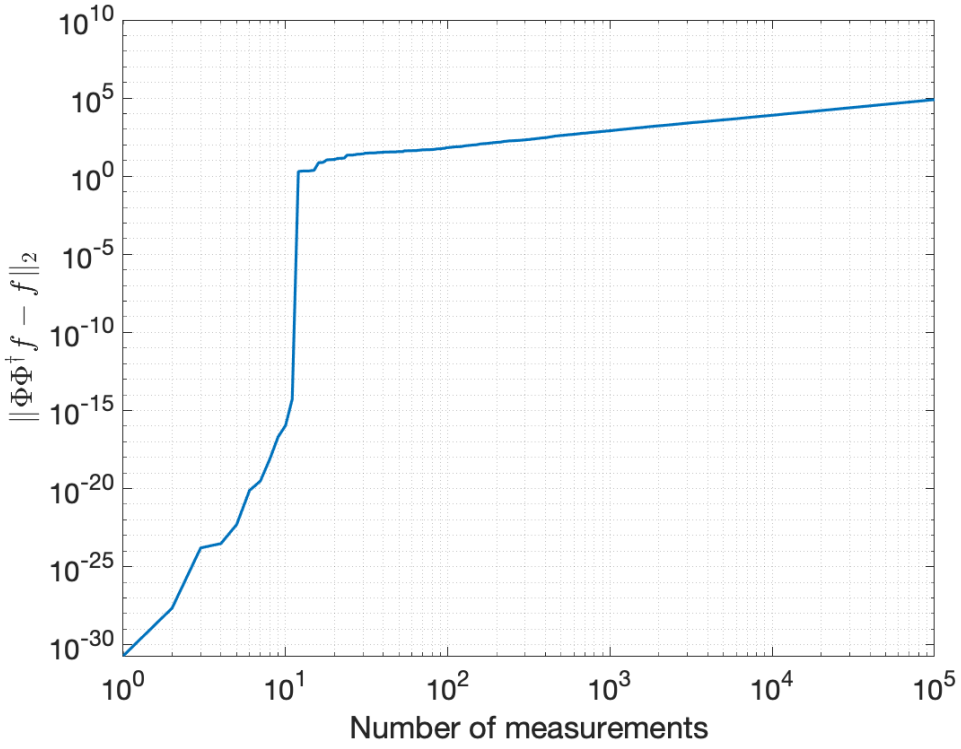

Figure 1.1: Problem 3

Answer the following, and provide all the details, theories, and analysis you used to build your opinion:

1. What is the number of candidate functions (number of columns) that is used in Φ.
2. Explain these unexpected results and discuss any misconceptions or mistakes in this experiment.

Problem 4 (30 points):
In the attached file (ExamData.mat) you can find a dataset that contains:

- X: Time series from a 6-dimensional continuous system. Each row represents one observation.
- t: The sampling time.

The measurements are very accurate, with $SNR \geq 20$. Answer the following.

1. To obtain the ODE model using power expansion and sparse regression, what regression method do you suggest to use, and why?
2. To create the candidate functions matrix Φ, what model order do you suggest? (quadratic, cubic ... etc.) and why?
3. If we assume the quadratic model, what is the number of candidate functions you will have? What is the total number of parameters? What are the dimensions of the parameters matrix β.

4. If we suggest using the cross-validation (CV) technique, what CV method do you recommend to use, and why?

5. Investigate the data closely. What are the challenges (if they exist) in modeling with this data? Explain how you can solve this/these challenges?.

6. What is your expectations about this system? Any weird behavior? Any phenomenon? Always explain why you think that.

7. With any method of your choice, obtain an accurate ODE model from this data. Write it as a 6-dimensional ODE system:

$$\begin{aligned}
\dot{x}_1 &= \dots \\
\dot{x}_2 &= \dots \\
\dot{x}_3 &= \dots \\
\dot{x}_4 &= \dots \\
\dot{x}_5 &= \dots \\
\dot{x}_6 &= \dots
\end{aligned}$$

8. Can we use the cross-validation (or other technique) to determine boundaries for the parameters? (min and max value for each parameter). Explain in detail and specify these boundaries.

Good Luck
Looking forward to seeing your excellent work.

If you have any questions, you can email me or drop by my office. Feel free to ask. The exam is a learning tool in the first place.