Applied Mathematics and Computational Sciences

MATH 302: Numerical Analysis

Instructor: Prof. Konstantinos Efstathiou

Approximation Theory and Related Applications

Author:
Qixuan Wang

We hear within us the perpetual call: There is the problem. Seek its solution. You can find it by pure reason, for in mathematics there is no *ignorabimus* – David Hilbert (1900).

Date: October 17, 2022

Table of Contents

List of Figures

1 Introduction

Approximation theory is the branch of mathematics which studies the process of approximating general functions by simple functions such as polynomials, finite elements or Fourier series. It therefore plays a central role in the analysis of numerical methods and is a subject that serves as an important bridge between pure and applied mathematics.

The origin of approximation theory is traced back to the fundamental work of Bernstein, Chebyshev, Haar, Hermite, Kolmogorov, Lagrange, Markov, and others. However, this branch of mathematics was not fully established until the founding of the *Journal of Approximation Theory* in 1968. Now with the rapid advance of digital computers, approximation theory has become a very important branch of mathematics.

In this project, we are going to discuss several techniques in approximation theory such as discrete least squares approximation, continuous least squares approximation, and Chebyshev polynomials. In addition, we will investigate the wide range of applications of discrete least squares approximation.

The main questions we would like to answer are as follows:

(1) How do we use least squares approximation to find the line of best fit in practical problems?
(2) Can we identify the trends in stocks and make prediction about the trends using least squares approximation?

The main method we use in the project is discrete least squares approximation. We will discuss the applications of polynomial least squares and of non-polynomials least squares.

There are three major results we would like to present. The first one is that we successfully apply polynomial least squares to investigating the relationship between the market price of gold and the stock price of a gold mining corporation. The second one is that we model the weather temperature in Suzhou from 2020 until October 2022 using non-polynomials least squares. The last one is that we use non-polynomials least squares approximation to identify the trends in real estate stocks and to make reasonable prediction about the future trends.

2 Background and literature review

To fully answer the research questions, we first need to have a good understanding of the techniques in approximation theory. Here I would like to discuss discrete least squares approximation, continuous least squares approximation, and Chebyshev polynomials. Then I will give a brief literature review.

2.1 Discrete least squares approximation

Generally speaking, there are two different types of discrete least squares approximation, which are polynomial least squares and non-polynomials least squares. However, the essence of these two methods is basically the same.

Polynomial least squares

Given data $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$, we would like to find the best polynomial approximation $P_n(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ where m will be usually much larger than n and $a_n \neq 0$.

In order to achieve this goal, we want to minimize

$$E = \sum_{i=1}^{m} (y_i - P_n(x_i))^2 = \sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{n} a_j x_i^j \right)^2$$

with respect to the parameters $a_n, a_{n-1}, \ldots, a_0$.

For a minimum to occur, the necessary conditions are

$$\frac{\partial E}{\partial a_k} = 0 \Rightarrow -\sum_{i=1}^{m} y_i x_i^k + \sum_{j=0}^{n} a_j \left(\sum_{i=1}^{m} x_i^{k+j} \right) = 0$$

for $k = 0, 1, \ldots, n$. Using algebra, we get a system of $(n+1)$ equations and $(n+1)$ unknowns

$$\sum_{j=0}^{n} a_j \left(\sum_{i=1}^{m} x_i^{k+j} \right) = \sum_{i=1}^{m} y_i x_i^k$$

for $k = 0, 1, \ldots, n$. We can write this system as a matrix equation

$$A\mathbf{a} = \mathbf{b}$$

where $\mathbf{a}$ is the unknown vector we are trying to find, and $\mathbf{b}$ is the constant vector

$$a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}, b = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} y_i x_i \\ \vdots \\ \sum_{i=1}^{m} y_i x_i^n \end{bmatrix}$$

and A is an $(n+1)$ by $(n+1)$ symmetric matrix with (kj)th entry A_{kj}, $k = 1, 2, \ldots, n+1$, $j = 1, 2, \ldots, n+1$ given by

$$A_{kj} = \sum_{i=1}^{m} x_i^{k+j-2}.$$

The matrix equation $A\mathbf{a} = \mathbf{b}$ has a unique solution if the x_i are distinct, and $n \leq m - 1$.

Solving this equation we can get the expression of $P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$.

Non-polynomials least squares

The method of least squares is not only for polynomials. For example, suppose we want to find the function

$$f(t) = a + bt + c\sin(\frac{2\pi t}{365}) + d\cos(\frac{2\pi t}{365})$$

that has the best fit to some data $(t_1, T_1), \ldots, (t_m, T_m)$ in the least-squares sense. This function is used in modeling weather temperature data, where t denotes time, and T denotes the temperature.

Similarly, we want to minimize the least squares error term

$$E = \sum_{i=1}^{m} (f(t_i) - T_i)^2 = \sum_{i=1}^{m} \left(a + bt_i + c\sin(\frac{2\pi t_i}{365}) + d\cos(\frac{2\pi t_i}{365}) - T_i \right)^2.$$

Therefore, we set its partial derivatives with respect to the unknowns a, b, c, d to zero to obtain the following equations:

$$\frac{\partial E}{\partial a} = 0 \Rightarrow \sum_{i=1}^{m} 2\left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0$$
$$\Rightarrow \sum_{i=1}^{m} \left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0,$$

$$\frac{\partial E}{\partial b} = 0 \Rightarrow \sum_{i=1}^{m} (2t_i)\left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0$$
$$\Rightarrow \sum_{i=1}^{m} t_i\left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0,$$

$$\frac{\partial E}{\partial c} = 0 \Rightarrow \sum_{i=1}^{m} \left(2\sin\left(\frac{2\pi t_i}{365}\right)\right)\left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0$$
$$\Rightarrow \sum_{i=1}^{m} \sin\left(\frac{2\pi t_i}{365}\right)\left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0,$$

$$\frac{\partial E}{\partial d} = 0 \Rightarrow \sum_{i=1}^{m} \left(2\cos\left(\frac{2\pi t_i}{365}\right)\right)\left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0$$
$$\Rightarrow \sum_{i=1}^{m} \cos\left(\frac{2\pi t_i}{365}\right)\left(a + bt_i + c\sin\left(\frac{2\pi t_i}{365}\right) + d\cos\left(\frac{2\pi t_i}{365}\right) - T_i\right) = 0.$$

Rearranging terms in the above equations, we get a system of four equations and four unknowns:

$$am + b\sum_{i=1}^{m} t_i + c\sum_{i=1}^{m}\sin\left(\frac{2\pi t_i}{365}\right) + d\sum_{i=1}^{m}\cos\left(\frac{2\pi t_i}{365}\right) = \sum_{i=1}^{m} T_i$$
$$a\sum_{i=1}^{m} t_i + b\sum_{i=1}^{m} t_i^2 + c\sum_{i=1}^{m} t_i\sin\left(\frac{2\pi t_i}{365}\right) + d\sum_{i=1}^{m} t_i\cos\left(\frac{2\pi t_i}{365}\right) = \sum_{i=1}^{m} T_i t_i$$
$$a\sum_{i=1}^{m}\sin\left(\frac{2\pi t_i}{365}\right) + b\sum_{i=1}^{m} t_i\sin\left(\frac{2\pi t_i}{365}\right) + c\sum_{i=1}^{m}\sin^2\left(\frac{2\pi t_i}{365}\right) + d\sum_{i=1}^{m}\sin\left(\frac{2\pi t_i}{365}\right)\cos\left(\frac{2\pi t_i}{365}\right)$$
$$= \sum_{i=1}^{m} T_i\sin\left(\frac{2\pi t_i}{365}\right)$$
$$a\sum_{i=1}^{m}\cos\left(\frac{2\pi t_i}{365}\right) + b\sum_{i=1}^{m} t_i\cos\left(\frac{2\pi t_i}{365}\right) + c\sum_{i=1}^{m}\sin\left(\frac{2\pi t_i}{365}\right)\cos\left(\frac{2\pi t_i}{365}\right) + d\sum_{i=1}^{m}\cos^2\left(\frac{2\pi t_i}{365}\right)$$
$$= \sum_{i=1}^{m} T_i\cos\left(\frac{2\pi t_i}{365}\right)$$

Using a short-hand notation where we suppress the argument $(\frac{2\pi t_i}{365})$ in the trigonometric functions, and the summation indices, we write the above equations as a matrix equation:

$$\underbrace{\begin{bmatrix} m & \sum t_i & \sum \sin(\cdot) & \sum \cos(\cdot) \\ \sum t_i & \sum t_i^2 & \sum t_i \sin(\cdot) & \sum t_i \cos(\cdot) \\ \sum \sin(\cdot) & \sum t_i \sin(\cdot) & \sum \sin^2(\cdot) & \sum \sin(\cdot)\cos(\cdot) \\ \sum \cos(\cdot) & \sum t_i \cos(\cdot) & \sum \sin(\cdot)\cos(\cdot) & \sum \cos^2(\cdot) \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \underbrace{\begin{bmatrix} \sum T_i \\ \sum T_i t_i \\ \sum T_i \sin(\cdot) \\ \sum T_i \cos(\cdot) \end{bmatrix}}_{\mathbf{r}}$$

Solve the equation $A\mathbf{x} = \mathbf{r}$ to get the unknowns a, b, c, d.

2.2 Continuous least squares approximation

In discrete least squares, our starting point was a set of data points. Here we will start with a continuous function f on $[a, b]$ and our goal is to find the 'best' polynomial $P_n(x) = \sum_{j=0}^{n} a_j x^j$ of degree at most n, that approximates f on $[a, b]$.

As before, 'best' polynomial will mean the polynomial that minimizes the least squares error:

$$E = \int_a^b \left(f(x) - \sum_{j=0}^{n} a_j x^j \right)^2 dx.$$

To minimize E we set $\frac{\partial E}{\partial a_k} = 0$, for $k = 0, 1, \ldots, n$, and observe

$$\begin{aligned} \frac{\partial E}{\partial a_k} &= \frac{\partial}{\partial a_k} \left(\int_a^b f^2(x) dx - 2 \int_a^b f(x) \left(\sum_{j=0}^{n} a_j x^j \right) dx + \int_a^b \left(\sum_{j=0}^{n} a_j x^j \right)^2 dx \right) \\ &= -2 \int_a^b f(x) x^k dx + 2 \sum_{j=0} a_j \int_a^b x^{j+k} dx = 0, \end{aligned}$$

which gives the $(n+1)$ equations for the continuous least squares problem:

$$\sum_{j=0}^{n} a_j \int_a^b x^{j+k} dx = \int_a^b f(x) x^k dx$$

for $k = 0, 1, \ldots, n$.

Note that the only unknowns in these equations are the a_j's, hence this is a linear system of equations. Solving the system of equations we can get the expression of $P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$.

2.3 Chebyshev polynomials

We will introduce another technique in approximation theory, which is Chebyshev polynomials, but we don't plan to go into details.

The Chebyshev polynomials of the first kind T_n are defined by

$$T_n(\cos\theta) = \cos(n\theta).$$

Hence, $T_n(x) = \cos(n \cos^{-1} x)$, $n \geq 0$.

The Chebyshev polynomials of the first kind are obtained from the recurrence relation $T_0(x) = 1$, $T_1(x) = x$, $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$.

Here is a plot of the first five Chebyshev polynomials:

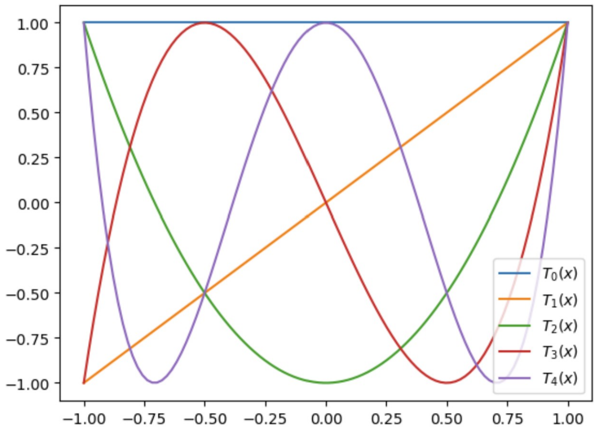

Figure 1: Plot of the first five Chebyshev polynomials.

Source: p. 203, First Semester in Numerical Analysis with Julia

Chebyshev polynomials can also be used to compute the least squares approximation to some functions. If we plot $y = e^x$ together with polynomial approximations of degree two and three using Chebyshev basis polynomials, we will get the following figure.

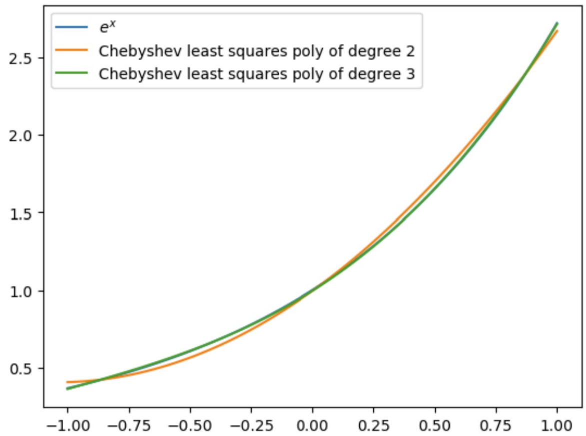

Figure 2: Approximating e^x using Chebyshev polynomials.

Source: p. 205, First Semester in Numerical Analysis with Julia

2.4 Literature review

As discussed before, (1) is used in modeling weather temperature data, where t denotes time, and T denotes the temperature.

$$f(t) = a + bt + c\sin(\frac{2\pi t}{365}) + d\cos(\frac{2\pi t}{365}) \tag{1}$$

In the book *First Semester in Numerical Analysis with Julia*, Ökten (2019) models the weather temperature in Australia using non-polynomials least squares. The following figure plots the daily maximum temperature during a period of 1056 days, from 2016 until November 21, 2018, as measured by a weather station at Melbourne airport, Australia.

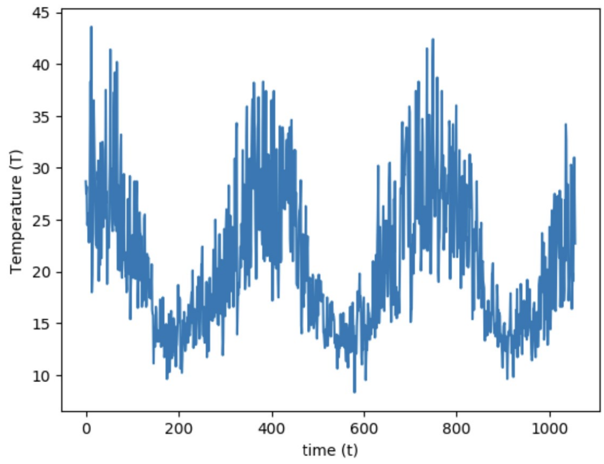

Figure 3: Line chart of temperature in Australia.

Source: p. 181, First Semester in Numerical Analysis with Julia

Using the same algorithm as discussed in Section 2.1, Ökten gets the explicit expression of

$$f(t) = a + bt + c\sin(\frac{2\pi t}{365}) + d\cos(\frac{2\pi t}{365})$$

where $a = 20.2898$, $b = 0.00116773$, $c = 2.72116$, $d = 6.88809$. The following figure shows that the approximating function fits the trend in original temperature data very well.

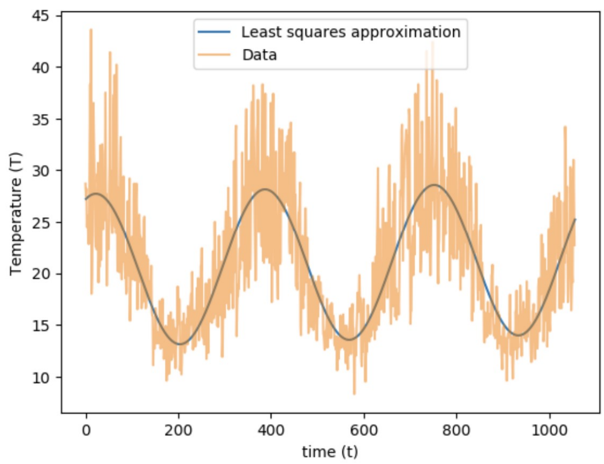

Figure 4: Line chart of temperature in Australia with approximating function.

Source: p. 186, First Semester in Numerical Analysis with Julia

In the following sections, we are going to apply non-polynomials least squares to approximating the trends in stocks as well.

3 Methods and preliminaries

The research questions we propose at the beginning are as follows:

(1) How do we use least squares approximation to find the line of best fit in practical problems?
(2) Can we identify the trends in stocks and make prediction about the trends using least squares approximation?

To answer the first question, we need to write a program that implements the algorithm of polynomial least squares approximation. In this project, we use Julia to solve the matrix equation $A\mathbf{a} = \mathbf{b}$ given in Section 2.1 and the code will be presented in Appendix A.

In terms of the implementation of non-polynomials least squares approximation, we can also write a Julia program. The key part of the program is to solve the matrix equation

$$\underbrace{\begin{bmatrix} m & \sum t_i & \sum \sin(\cdot) & \sum \cos(\cdot) \\ \sum t_i & \sum t_i^2 & \sum t_i \sin(\cdot) & \sum t_i \cos(\cdot) \\ \sum \sin(\cdot) & \sum t_i \sin(\cdot) & \sum \sin^2(\cdot) & \sum \sin(\cdot)\cos(\cdot) \\ \sum \cos(\cdot) & \sum t_i \cos(\cdot) & \sum \sin(\cdot)\cos(\cdot) & \sum \cos^2(\cdot) \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \underbrace{\begin{bmatrix} \sum T_i \\ \sum T_i t_i \\ \sum T_i \sin(\cdot) \\ \sum T_i \cos(\cdot) \end{bmatrix}}_{\mathbf{r}}$$

to get the explicit expression of

$$f(t) = a + bt + c\sin(\frac{2\pi t}{T}) + d\cos(\frac{2\pi t}{T})$$

where T is the period and a, b, c, d are the unknowns. Part of the code will be presented in Appendix B.

After knowing the methods and algorithms, we still need to cover some preliminaries.

Since we are going to investigate the trends in stocks, we should have an understanding of what kinds of stocks we are interested in. In this project, we will focus on the trends in cyclical stocks. Because in the above expression of $f(t)$ the period T is a crucial parameter and only the cyclical stocks rise and fall in line with the general economic cycle and are affected by macroeconomic changes in the overall economy. Cyclical stocks are generally concentrated in specific industries such as airlines, real estate, and construction.

Here comes another terminology – economic cycle. The term economic cycle refers to the fluctuations of the economy between periods of expansion (growth) and contraction (recession). In addition, the average length of recessions in the U.S. since World War II has been just around 11 months. U.S. expansions have typically lasted longer than U.S. recessions. However, for simplicity we assume that the average length of expansions is also 11 months.

In the project, we are going to investigate the trends in a real estate stock in the U.S., which belongs to cyclical stocks according to the above discussion. Therefore, the trends of this stock will be largely influenced by the U.S. economic cycle. Furthermore, we can assume that the period T of the stock is 22 months/660 days (the length of recessions is 11 months and the length of expansions is 11 months as well).

Hence, for the real estate stock, the approximating function becomes

$$f(t) = a + bt + c\sin(\frac{2\pi t}{660}) + d\cos(\frac{2\pi t}{660}) \tag{2}$$

where a, b, c, d are four unknowns.

In the next section, we will find the explicit expression of (2) and make prediction about the trends according to the approximating function.

4 Results

In this section, we will present the results in three parts. The first part is to investigate the relationship between the market price of gold and the stock price of a gold mining corporation.

The second part is to model the weather temperature in Suzhou using non-polynomials least squares. The last part is to identify the trends in a real estate stock and to make reasonable prediction about the future trends using non-polynomials least squares, which is also our main result.

4.1 Gold price versus stock price

Intuitively, the stock price of a gold mining corporation may be directly proportional to the market price of gold since as the gold price goes up, the revenue of the gold mining corporation will increase, which is very likely to boost its stock price.

Here we are going to trace the stock price of Newmont, which is one of the world's largest gold mining corporations, and the market price of gold from 2019/8/30 to 2022/10/6. The following figures plot the trends in gold price and stock price.

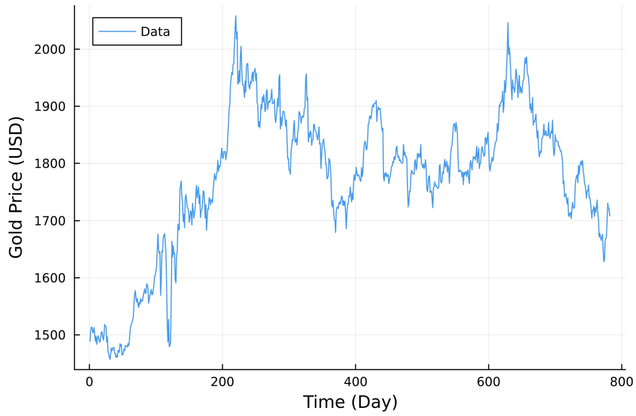

Figure 5: Trends in gold price.

Source: Plotted by Julia, data collected from https://www.investing.com/commodities/gold-historical-data

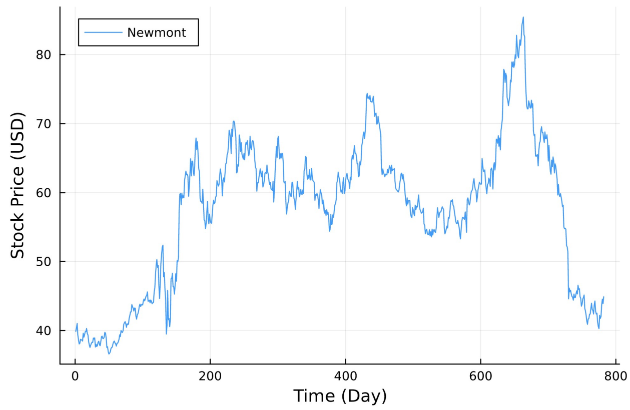

Figure 6: Trends in stock price of Newmont.

Source: Plotted by Julia, data collected from https://finance.yahoo.com/

Next we are going to investigate the relationship between the market price of gold and the stock price of Newmont. The following figure is the scatter plot of gold price and stock price. Note that gold price is the independent variable and stock price is the dependent variable.

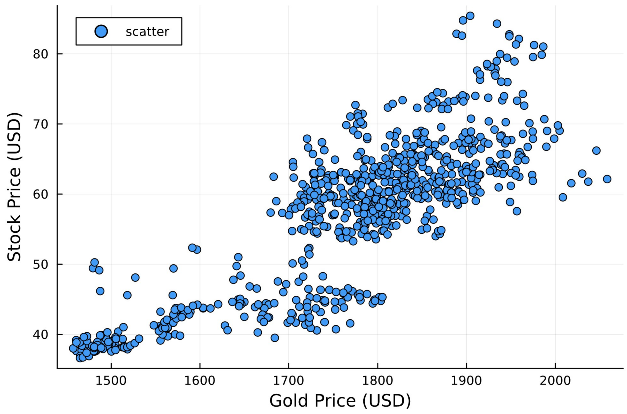

Figure 7: Scatter plot of gold price and stock price.

Source: Plotted by Julia

We implement the algorithm of polynomial least squares approximation and use Julia to solve the matrix equation $A\mathbf{a} = \mathbf{b}$ given in Section 2.1. Note that here we apply the polynomial of order 5 to approximating the trend. The code will be presented in Appendix C. We plot the fitting polynomial of order 5 together with the previous scatter plot.

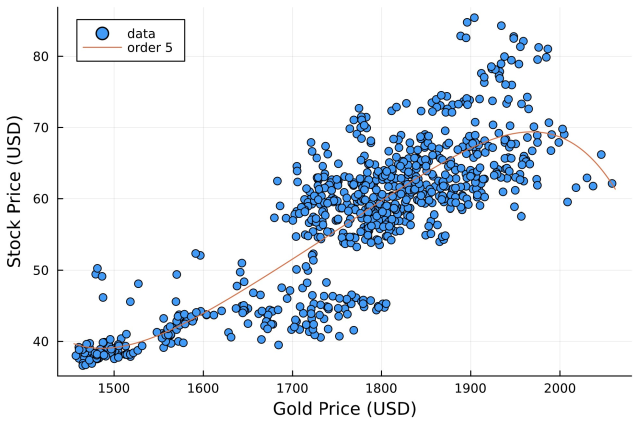

Figure 8: Scatter plot with the fitting polynomial.

Source: Plotted by Julia

Now we are able to make a few remarks.

(1) We can find the fitting polynomial in some practical problems applying the algorithm of polynomial least squares approximation.
(2) Generally speaking, the stock price of Newmont increases as the market price of gold goes up.
(3) From Figure 8, we can see that the relationship between the market price of gold and the stock price of Newmont is not just linear. It's a complicated relationship, but we can use polynomial least squares approximation to characterize it to some extent (polynomial least squares approximation helps us understand their relationship).
(4) In this problem, polynomial least squares approximation is influenced by some extreme values at both ends (when gold price is less than \$1500 or greater than \$1950).
(5) Investors should pay attention to the change of gold price when investing the stocks of gold mining corporations.

4.2 Model the weather temperature using non-polynomials least squares

We will model the weather temperature in Suzhou from February 2020 to October 2022 using non-polynomials least squares.

Following Ökten's algorithm, we can get the explicit expression of

$$f(t) = a + bt + c\sin(\frac{2\pi t}{365}) + d\cos(\frac{2\pi t}{365})$$

where $a = 16.2045$, $b = 0.000463146$, $c = 3.43072$, $d = -12.2178$ by solving the matrix equation $A\mathbf{x} = \mathbf{r}$ discussed in Section 2.1.

The following figure plots the daily average temperature in Suzhou during a period of 994 days (from February 2020 to October 2022) together with the approximating function $f(t)$.

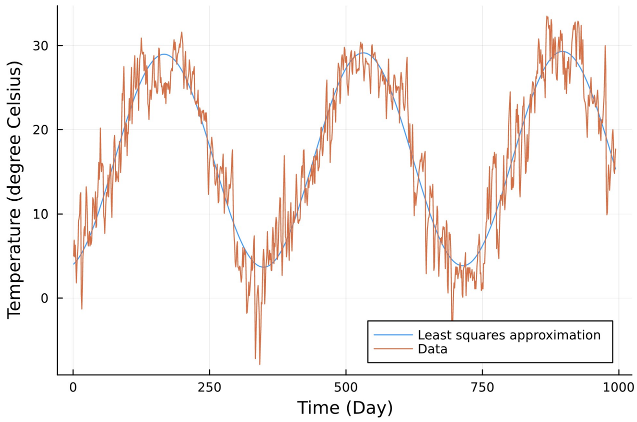

Figure 9: Line chart of temperature in Suzhou with approximating function.

Source: Plotted by Julia, data collected from https://www.visualcrossing.com/weather/weather-data-services

We now make some remarks about the above result.

(1) From Figure 9, we can see that the approximating function $f(t)$ fits the trend in original temperature data of Suzhou very well since they reach the peaks and troughs almost simultaneously.
(2) The approximating function $f(t)$ can help us understand the trend in the temperature data, but we may not be able to make an accurate weather forecast using $f(t)$ since it cannot tell us the exact temperature on a given day (it can only give us the approximate values).
(3) It is interesting to point out that the parameter $b = 0.000463146 > 0$ implies the phenomenon of 'global warming' since the term bt in the approximating function $f(t)$ refers to the general trend in $f(t)$ – either increasing or decreasing when the time t increases by one period (365 days).

4.3 Identify the trends in stocks using non-polynomials least squares

According to the discussion in Section 3, we are going to investigate the trends in a real estate stock in the U.S., which is Lennar. Lennar is the leading homebuilder in the U.S. and thus belongs to cyclical stocks. We will trace its stock price from 2012/7/2 to 2018/9/27. The following figure plots the trends in stock price of Lennar in the given period.

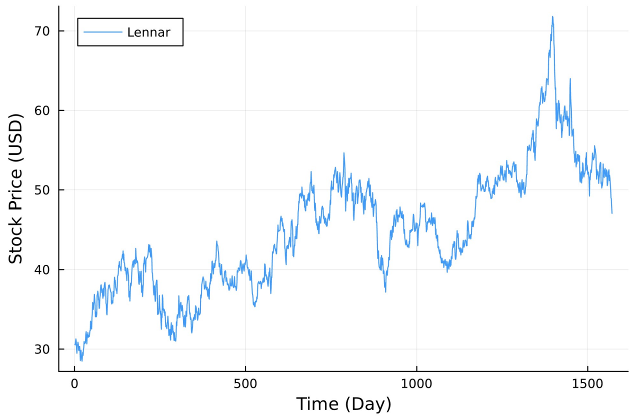

Figure 10: Trends in stock price of Lennar.

Source: Plotted by Julia, data collected from https://finance.yahoo.com/

We adopt the same idea as that of Section 4.2 and we would like to find the explicit expression of

$$f(t) = a + bt + c\sin(\frac{2\pi t}{T}) + d\cos(\frac{2\pi t}{T})$$

where a, b, c, d are the unknowns and $T = 660$ days according to the assumption in Section 3.

We use Julia to solve the matrix equation $A\mathbf{x} = \mathbf{r}$ given in Section 2.1. Then we get

$$f(t) = a + bt + c\sin(\frac{2\pi t}{660}) + d\cos(\frac{2\pi t}{660})$$

where $a = 33.3416$, $b = 0.0141036$, $c = 2.60417$, $d = 3.08978$. Therefore, we can plot the trends in stock price together with the above non-polynomials least squares approximation.

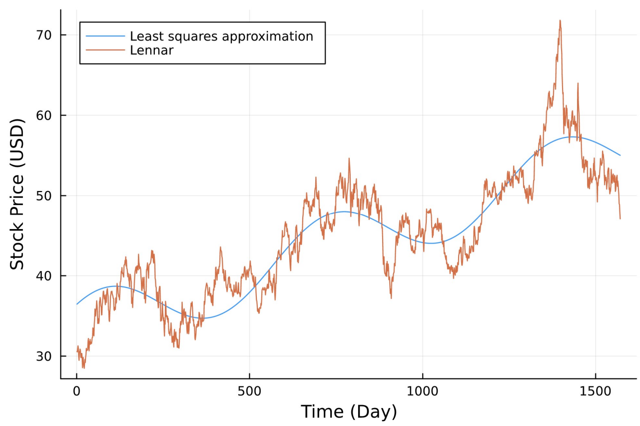

Figure 11: Trends in stock price of Lennar with approximating function.

Source: Plotted by Julia

Furthermore, one way to test the non-polynomials least squares approximation is to introduce new data. It can also determine if one can make prediction about the future trends using the original approximating function $f(t)$. We plot the trends in stock price of Lennar from 2012/7/2 to 2020/3/30 together with the original approximating function.

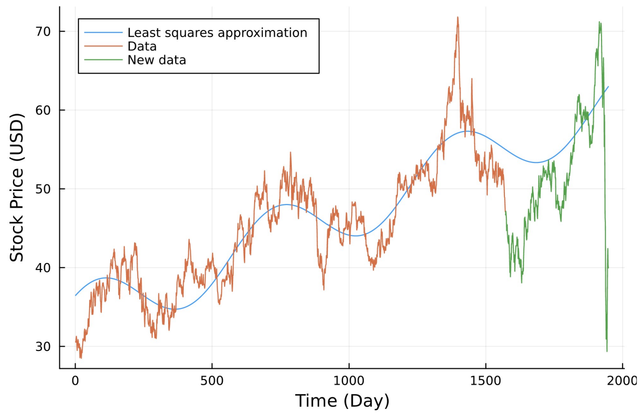

Figure 12: Trends in stock price of Lennar before 2020/3/30 with approximating function.

Source: Plotted by Julia

From Figure 12, we can see that there was a sharp decline in the stock price of Lennar in March 2020. This was because the Covid-19 pandemic began to influence U.S. economy and stock market to a large extent.

Now we are able to make a few remarks about the above result.

(1) From Figure 11 and 12, we may conclude that it is reasonable to apply non-polynomials least squares approximation to the trends in cyclical stocks since the approximating function $f(t)$ fits the overall trends quite well (they reach the peaks and troughs almost simultaneously).
(2) The term bt with $b > 0$ in the approximating function $f(t)$ implies that the real estate industry develops steadily as U.S. economy grows in the long term although there are some recessions.
(3) One limitation of this approach is that we can only investigate the trends in cyclical stocks using non-polynomials least squares approximation since the period T is a fundamental parameter in the expression of $f(t)$ and typically non-cyclical stocks don't have periods regarding their stock prices.
(4) The other limitation is that black swan events such as the Covid-19 pandemic are very likely to reduce the effectiveness of non-polynomials least squares approximation significantly. For example, from Figure 12 we can see that the stock price of Lennar decreased a lot in March 2020, but the approximating function $f(t)$ was still increasing. After all, no one can predict when the black swan events will happen.
(5) One strategy for investing in stocks is that one may invest heavily when black swan events start to cause a significant decline in stock market since the intrinsic value of some stocks may not be largely affected. As Warren Buffett (1986) once said that it is wise for investors to be 'fearful when others are greedy, and **greedy when others are fearful**.' If we have a look at the stock price of Lennar in April 2020, we will see that its stock price bounced back quite rapidly.

5 Conclusion and discussion

Approximation theory, which is often classified as approximations and expansions, is a subject that serves as an important bridge between pure and applied mathematics. In this project, we discuss some common techniques in approximation theory such as discrete least squares approximation, continuous least squares approximation, and Chebyshev polynomials. We also investigate the wide range of applications of polynomial least squares and of non-polynomials least squares.

The major results we would like to summarize are as follows:

(1) We successfully apply polynomial least squares to investigating the relationship between the market price of gold and the stock price of a gold mining corporation.
(2) We model the weather temperature in Suzhou from 2020 until October 2022 using non-

polynomials least squares.
(3) We use non-polynomials least squares approximation to identify the trends in real estate stocks and to make reasonable prediction about the future trends.

These results also fully answer the two research questions we raise at the beginning. Besides, we make two useful suggestions for investment in stocks. One is that when investing the stocks of gold mining corporations investors should pay attention to the change of gold price. The other is that one may invest heavily when black swan events start to cause a significant decline in stock market (same suggestion as that of Warren Buffett).

In terms of the extensions and improvements, we would like to propose the following ideas:

(1) One can introduce some new terms such as $m \cdot \ln(x)$ (where m is a constant) to the non-polynomials approximating function

$$f(t) = a + bt + c\sin(\frac{2\pi t}{T}) + d\cos(\frac{2\pi t}{T}). \tag{3}$$

Because when we try to model the weather temperature or some financial problems, the trends may be quite complicated to characterize and it is reasonable to involve some other terms like $m \cdot \ln(x)$ in the approximating function.
(2) We can set the period T in (3) an unknown as well. Then we need to solve a different matrix equation than $A\mathbf{x} = \mathbf{r}$ given in Section 2.1 in order to get the unknowns a, b, c, d, and T. This may give us a more accurate approximating function since in some cases we get the value of T from the assumption.

Bibliography

Achuthan, Lakshman (2022). *Business Cycle: What It Is, How to Measure It, the 4 Phases.* URL: https://www.investopedia.com/terms/b/businesscycle.asp#citation-17 (visited on 10th Oct. 2022).

Arula, Kadi (2022). *What are Cyclical Stocks? — Definition Examples.* URL: https://finbold.com/guide/cyclical-stocks-definition/ (visited on 10th Oct. 2022).

Burden, Richard L., J. Douglas Faires and Annette M. Burden (2014). *Numerical Analysis.* Cengage Learning.

Chui, Charles K. (2003). 'Approximations and Expansions'. In: *Encyclopedia of Physical Science and Technology* 3, pp. 581–607.

Ökten, Giray (2019). *First Semester Numerical Analysis with Julia.* Florida State University Libraries.

Appendix

A Code demo 1

Please see the complete code in Least squares approximation.ipynb.

```
In [2]: function leastsqfit(x::Array,y::Array,n)
            m=length(x) # number of data points
            d=n+1 # number of coefficients to determine
            A=zeros(d,d)
            b=zeros(d,1)
            # the linear system we want to solve is Ax=b
            p=Array{Float64}(undef,2*n+1)
            for k in 1:d
                sum=0
                for i in 1:m
                    sum=sum+y[i]*x[i]^(k-1)
                end
                b[k]=sum
            end
            # p[i] below is the sum of the (i-1)th power of the x coordinates
            p[1]=m
            for i in 2:2*n+1
                sum=0
                for j in 1:m
                    sum=sum+x[j]^(i-1)
                end
                p[i]=sum
            end
            # We next compute the upper triangular part of the coefficient
            # matrix A, and its diagonal
            for k in 1:d
                for j in k:d
                    A[k,j]=p[k+j-1]
                end
            end
            # The lower triangular part of the matrix is defined using the
            # fact the matrix is symmetric
            for i in 2:d
                for j in 1:i-1
                    A[i,j]=A[j,i]
                end
            end
            a=A\b
        end
```

B Code demo 2

In [29]:

```
A=zeros(4,4);
A[1,1]=994
A[1,2]=sum(time)
A[1,3]=sum(t->sin(2*pi*t/365),time)
A[1,4]=sum(t->cos(2*pi*t/365),time)
A[2,2]=sum(t->t^2,time)
A[2,3]=sum(t->t*sin(2*pi*t/365),time)
A[2,4]=sum(t->t*cos(2*pi*t/365),time)
A[3,3]=sum(t->(sin(2*pi*t/365))^2,time)
A[3,4]=sum(t->(sin(2*pi*t/365)*cos(2*pi*t/365)),time)
A[4,4]=sum(t->(cos(2*pi*t/365))^2,time)
for i=2:4
    for j=1:i
        A[i,j]=A[j,i]
    end
end
```

C Code demo 3

In [88]:

```
b1=leastsqfit(gold_vals,NEM_val,5)
xaxis= 1455:1/100:2062
yvals=map(x->poly(x,b1),xaxis)
scatter(gold_vals,NEM_val, label="data")
plot!(xaxis,yvals, label="order 5", xlabel="Gold Price (USD)" , ylabel="Stock Price (USD)", legend=:topleft)
```