Life expectancy around the world

In this problem you will work with the United Nations data on the life expectancy in different coutries in the world. Life expectancy is "a statistical measure of the average time an organism is expected to live" (Wikipedia). As an optional reading you can take a look at an article discussing past and future trends in this quality of life metric here (quite interesting read).

You will use linear least-squares techniques to create a model for a selected country and try to predict the life expectancy future trends. A big aspect of this assignment is **critical thinking** and justification of your choices of viable models.

Task 1 - data preparation

Read the file "`global life expectancy dataset.csv`". Extract the data for the US and a selected country along with the dates for which this data was obtained. Plot the data for both countries as a function of time. What are your conclusions from examining the data? How does life expectancy in the US compares to that of your selected country?9(73,3:end);

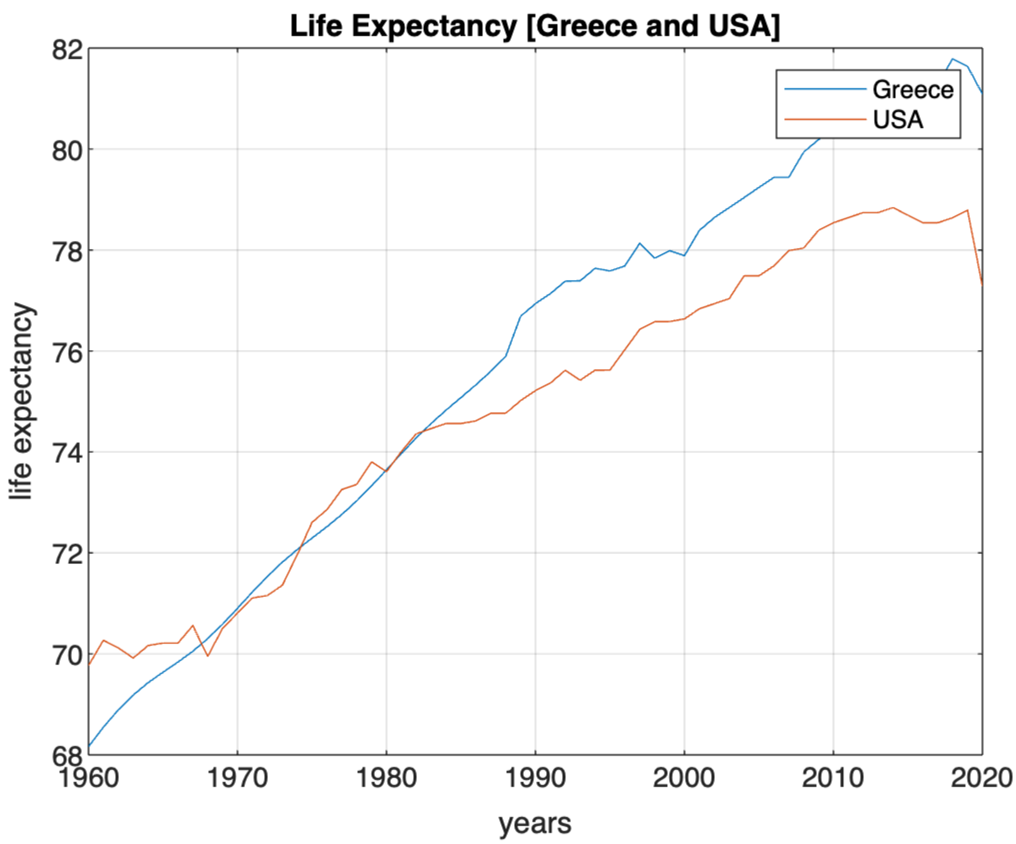
```
%T = readtable('global_life_expectancy_dataset.csv')
```

```
table_matrix = readmatrix('global_life_expectancy_dataset.csv')
```

```
table_matrix = 206×63
10^3 ×
       NaN       NaN    1.9600    1.9610    1.9620    1.9630    1.9640    1.9650 ···
       NaN       NaN    0.0657    0.0661    0.0664    0.0668    0.0671    0.0674
       NaN       NaN    0.0324    0.0330    0.0335    0.0340    0.0345    0.0349
       NaN       NaN    0.0375    0.0378    0.0381    0.0384    0.0388    0.0391
       NaN       NaN    0.0623    0.0633    0.0642    0.0649    0.0655    0.0658
       NaN       NaN    0.0515    0.0526    0.0536    0.0546    0.0556    0.0565
       NaN       NaN    0.0651    0.0652    0.0653    0.0653    0.0654    0.0655
       NaN       NaN    0.0660    0.0664    0.0668    0.0673    0.0677    0.0682
       NaN       NaN    0.0620    0.0625    0.0630    0.0635    0.0640    0.0644
       NaN       NaN    0.0708    0.0710    0.0709    0.0709    0.0709    0.0709
     ⋮
```

```
T1_dates = table_matrix(1,3:end);

Greece = table_matrix(73,3:end);
USA = table_matrix(194,3:end);
new_table = [T1_dates; Greece; USA];
Greece_ = plot(T1_dates, Greece);
hold on;
USA_ = plot(T1_dates, USA);

grid on;
xlabel('years')
ylabel('life expectancy')
title('Life Expectancy [Greece and USA]')
```

```
legend('Greece','USA')
```



Answer: The graph shows how life expectancy has decresead in the US relative to Greece. This trend begins to appear around the early 1980's. The US appears to have a higher life expectancy prior to the 1970's. This could possibly be a product of the post WW2 era.

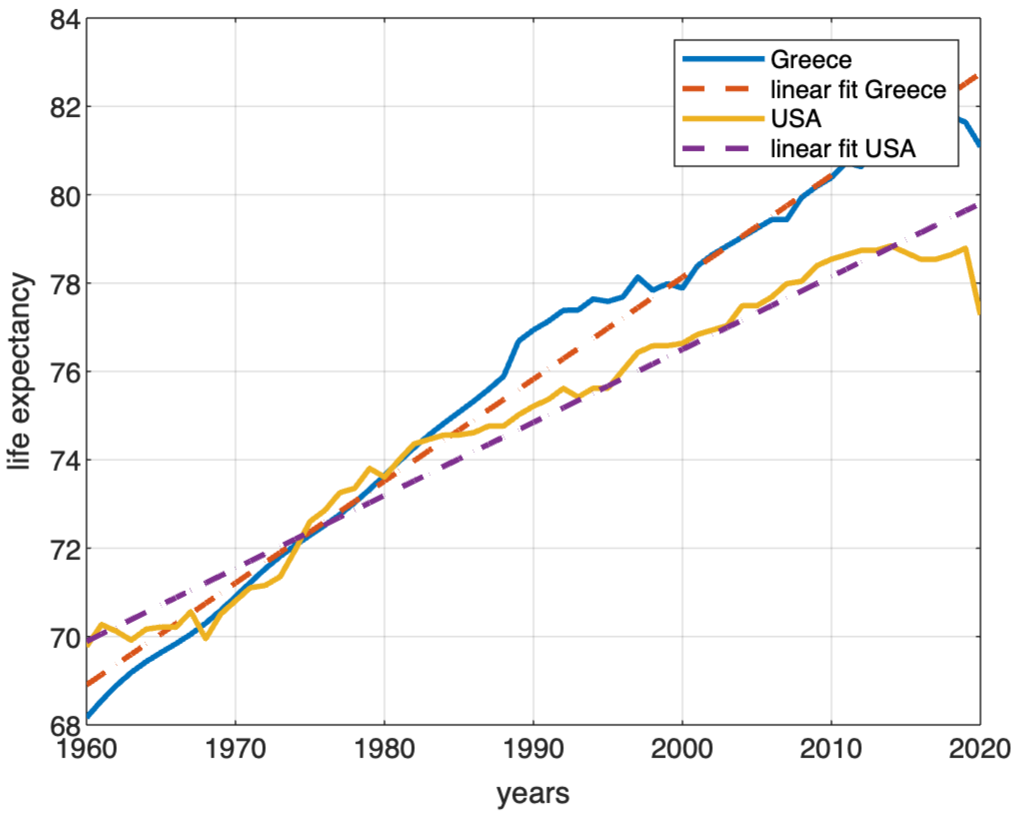
Task 2 - propose a model

Based on the plot in Task 1, propose at least three different functions (models) which you will try to fit to the data. Use the **normal equation approach** to fit all three models to the data and compute the residual of each fit. Plot the resulting models and the data. Which model fits the data best?

Line of best fit

```
close all
T1_dates = table_matrix(1,3:end)';
Greece = table_matrix(73,3:end)';
USA = table_matrix(194,3:end)';


n = length(T1_dates);
A = [T1_dates ones(n,1)];
c = (A'*A)\A'*Greece;
```

```
c1 = (A'*A)\A'*USA;
figure(2)
h = @(T1_dates) c(1)*T1_dates + c(2);
h1 = @(T1_dates) c1(1)*T1_dates + c1(2);
plot(T1_dates,Greece,T1_dates,h(T1_dates),'--','LineWidth',2,'MarkerSize',10)
;
hold on
plot(T1_dates,USA,T1_dates,h1(T1_dates),'--','LineWidth',2,'MarkerSize',10);
xlabel('years'); ylabel('life expectancy')
legend('Greece','linear fit Greece','USA','linear fit USA')
grid on
```



Quadratic Polynomial

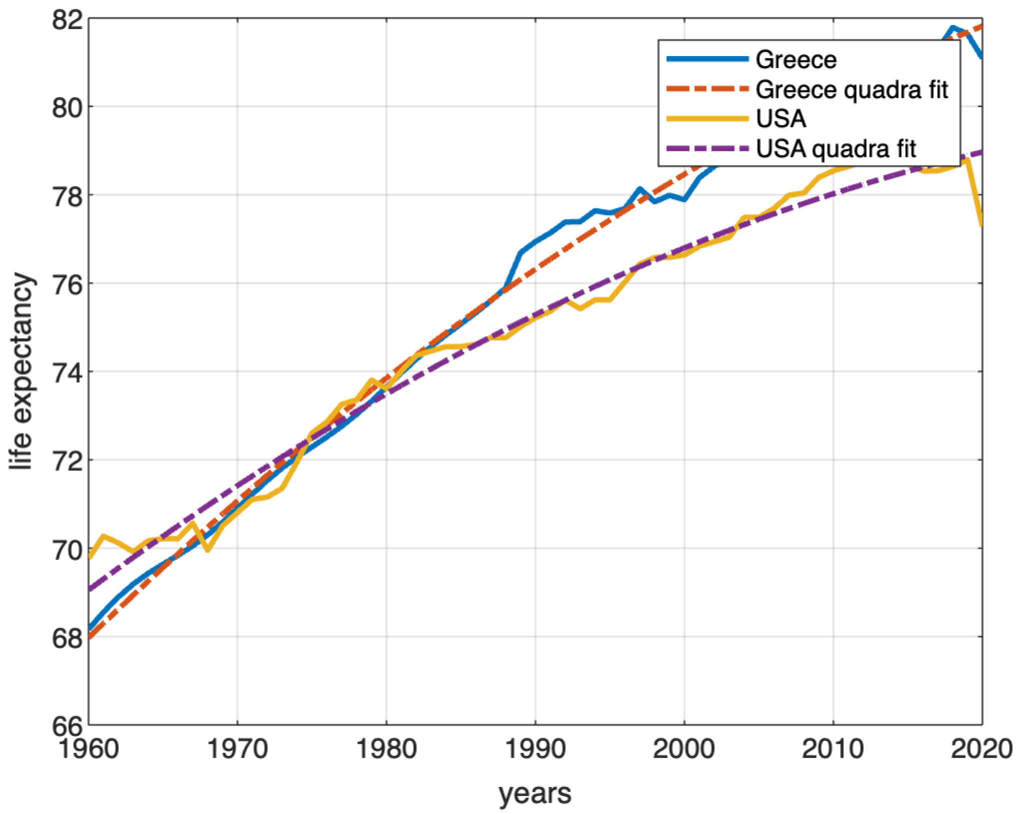
```
close all
T1_dates = table_matrix(1,3:end)';
Greece = table_matrix(73,3:end)';
USA = table_matrix(194,3:end)';


n = length(T1_dates);
A = [T1_dates.^2 T1_dates ones(n,1)];
c = (A'*A)\A'*Greece;
```

```
Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND =
3.117602e-22.
```

```
c1 = (A'*A)\A'*USA;
```

```
Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND =
3.117602e-22.
```

```
figure(1)
g = @(T1_dates) c(1)*T1_dates.^2 + c(2)*T1_dates + c(3);
g1 = @(T1_dates) c1(1)*T1_dates.^2 + c1(2)*T1_dates + c1(3);
plot(T1_dates,Greece,T1_dates,g(T1_dates),'-.','LineWidth',2,'MarkerSize',10)
hold on;
plot(T1_dates,USA,T1_dates,g1(T1_dates),'-.','LineWidth',2,'MarkerSize',10)
xlabel('years'); ylabel('life expectancy')
legend('Greece','Greece quadra fit','USA','USA quadra fit')
grid on
```



Cubic Polynomial

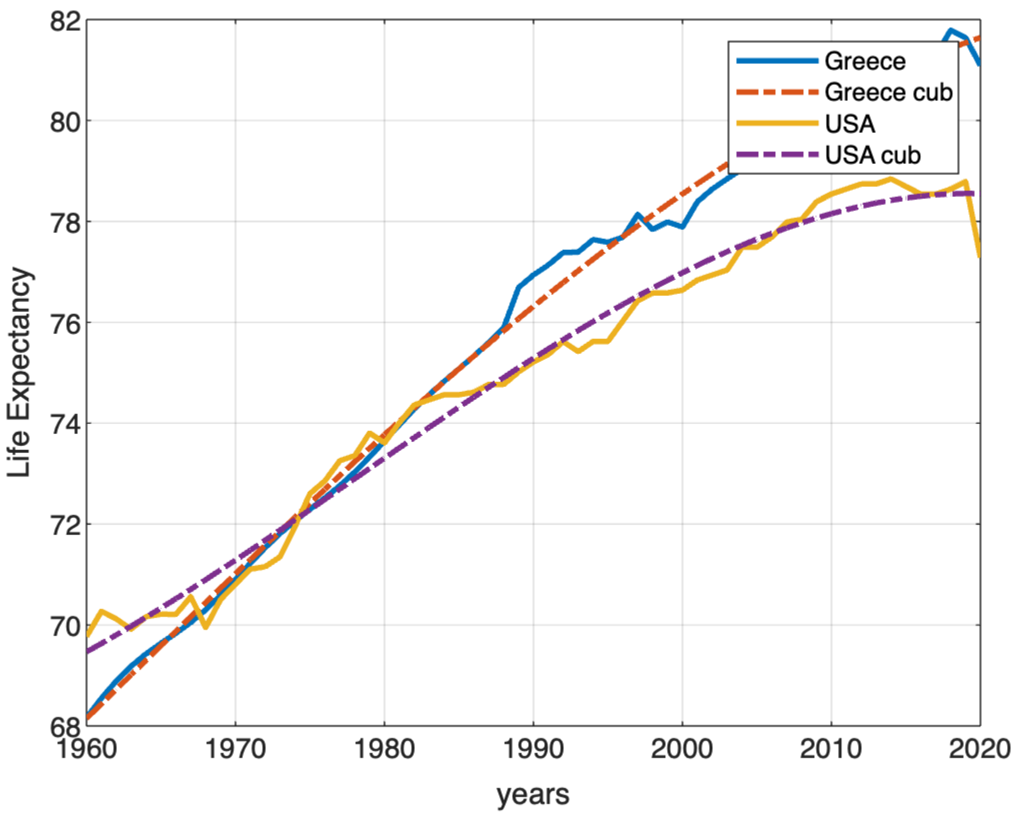
```
close all
T1_dates = table_matrix(1,3:end)';
Greece = table_matrix(73,3:end)';
USA = table_matrix(194,3:end)';


n = length(T1_dates);
A = [T1_dates.^3 T1_dates.^2 T1_dates ones(n,1)];
c = (A'*A)\A'*Greece
```

```
Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND =
4.751405e-33.
c = 4×1
```
10^5 ×
```
   -0.0000
    0.0000
   -0.0020
    1.2764
```

```
c1 = (A'*A)\A'*USA
```

```
Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND =
4.751405e-33.
c1 = 4×1
```
10^5 ×
```
   -0.0000
    0.0000
   -0.0048
    3.1308
```

```
figure(3)
g = @(T1_dates) c(1)*T1_dates.^3 + c(2)*T1_dates.^2 + c(3)*T1_dates + c(4);
g1 = @(T1_dates) c1(1)*T1_dates.^3 + c1(2)*T1_dates.^2 + c1(3)*T1_dates +
c1(4);
plot(T1_dates,Greece,T1_dates,g(T1_dates),'-.','LineWidth',2,'MarkerSize',10)
hold on;
plot(T1_dates,USA,T1_dates,g1(T1_dates),'-.','LineWidth',2,'MarkerSize',10)

grid on


xlabel('years'); ylabel('Life Expectancy')
legend('Greece','Greece cub','USA','USA cub')
grid on
```



Answer:

The cubic polymomial fits the life expectancy data the best, it closely represents how life expectancy has changed relative to the United States. The cubic polynomial in comparison to the quadratic polynomial provides more detail, particularly at the tail end of the data where we can begin to see a gradual negative slope only on the cubic polynomial graph.

Task 3 - predict the future trends

Using the best model you chose in Task 2 (or more than one model if you are not sure which one is best) predict the population of your selected country in year 2050 and 2100. Compare your result against the United Nations projections available here (make sure you add your selected country to the graph visible in the link, you can also access tabulated data and download it). How close is your prediction to the UN? If you are testing more than one model, did the model which fit the historical data best still performed the best in predicting the future?

```
% Greece
k = polyfit(T1_dates,Greece,2) % fitting using quadratic
```

```
k = 1×3
10^3 ×
   -0.0000    0.0065   -6.6299
```

```
prediction_Greece = polyval(k,2050);
```

```
fprintf('prediction in 2050 is %7.4f \n',prediction_Greece)
```

```
prediction in 2050 is 84.4795
```

```
prediction1 = polyval(k,2100);
fprintf('prediction in 2100 is %7.4f \n',prediction1)
```

```
prediction in 2100 is 82.6066
```

```
%USA (aditional optional task)
m = polyfit(T1_dates,USA,2); % fitting using quadratic
%
% prediction_USA = polyval(k,2050);
% fprintf('prediction in 2050 is %7.4f \n',prediction_USA)
%
% prediction2 = polyval(k,2100);
% fprintf('prediction in 2100 is %7.4f \n',prediction2)
```

Answer:

At first I noticed how the cubic polynomial closely represented the data. However, when looking at the results over a longer period of time, and then compare our results to the United Nations projections for 2050 and 2100, we can see that the quadratic polynomial is a much better fit in the long term over the cubic polynomial. The quadratic polynimial is a better fit for both projections 2100 and 2050, this could be the result of external factorsd that are not represented in my cubic polynomial, such as future pandemics, wars, and resource management problems for example. Therefore, the cubic polynomial is overfitting the data because I am using limmited information to make accurate predictions beyond a quadratic polynimial.

Optional: while examining the plot in Task 1, you may notice some outlying data point (i.e. those which are significantly off the general trend) due to some events which temporarily decrease life expectancy (i.e. major pandemic, war, etc). To get a model fit better you may need to remove those outliers from the data.

Task 4 - When are we going to live up to 100 years?

Having (hopefully) some confidence in your model by now, plot the historical data and a prediction of your model far into the future. We want to check whether (and when) people in your selected country can expect to live to 100 years. Create a function which takes in a year and returns the life expectancy (it can be a function handle with only one input, and can be built using polyfit and polyval or by solving the normal equation, or by using "\" operator directly). Use known to you zero-finding method (bisection, Newton, secant, fzero) to find a date when life expectancy reaches 100 years (i.e. where the model function crosses y = 100) line. Alternatively, if your model predicts total population decline (life expectancy of 0), find for which year this happens.

```
polyval(k,T1_dates)
```

```
ans = 61×1
   67.9741
   68.2979
   68.6185
   68.9360
   69.2504
   69.5615
   69.8695
   70.1744
   70.4761
   70.7747
      ⋮
```

```
R = @(T1_dates) polyval(k,T1_dates);
P = @(T1_dates) R(T1_dates)-84; % function
x0 = 1900; %initial point
x = fzero(P,x0)
```

```
x = 2.0413e+03
```

```
fprintf('Life expectancy in Greece of 82 years will be in %7.0f \n',x)
```

```
Life expectancy in Greece of 82 years will be in    2041
```

Answer:

After trying diffrent methods my function was not able to give me an exact time period for when people in Greece would live to be 100 years old. I did however manage to discover that in 2041 life expectancy in Greece will increase to 84 years. For the United States I was also unable to find a solution for a 100 year life expectancy. However the overall trend shows that it is bellow the Life expectancy of Greece and there for it makes sence that we unable to go beyond the results that we got for Greece. There seems to be an overall trend where the US seems consistently bellow Greece since the 1960's.

If you are interested in how this will look like in the US, you can do similar analysis as an optional task.

```
R1 = @(T1_dates) polyval(m,T1_dates);
P1 = @(T1_dates) R1(T1_dates)-80; % function
x0(1) = 1900; %initial point
x1 = fzero(P1,x0(1))
```

```
x1 = 2.0400e+03
```

```
fprintf('Life expectancy in the USA is 80 years will be in %7.0f \n',x1)
```

```
Life expectancy in the USA is 80 years will be in    2040
```