**Math 330 Fall 2020 Homework 3**

1. Consider the equations, $2x_1 + 3x_2 - x_3 = 5, 4x_1 + 4x_2 - 3x_3 = 3, -2x_1 + 3x_2 - 7x_3 = 1$.
(a) Write a Matlab script where you form the appropriate matrix $\mathcal{A}$ and vector $b$, and solve this system using Gaussian elimination (backslash command) to find $x = (x_1, x_2, x_3)^T$
(b) In this same script, compute the absolute residual $res = ||\mathcal{A}x - b||$ and relative residual $relres = res/||b||$.
(c) Compute $cond(A)$. Comment on how the magnitude of the relative residual and this condition are related.

2. *Condition number and relative error for linear systems of equations:* Download the matlab function genSPDlinearsystem.m from the Files tab. Given inputs $N$ and $cnd$, this function generates a random $N \times N$ matrix A and right-hand-side $b$ of size $N \times 1$. We want to use this script to test how the condition number of a matrix affects (1) the relative residual solving $Ax = b$ using Gauss Elimination, (2) the number of iterations Conjugate Gradient takes to solve $Ax = b$ and (3) How both these algorithms perform (how much time they take).

Consider the values $N = [250\ 500\ 1000\ 2000\ 4000]$ and $cnd = [10\ 100\ 1000\ 10000]$. Write a Matlab script that, for each value of N and cnd, does the following: (Hint: use two nested for loops)

- Generates SPD matrix $A$ and right-hand-side $b$ given size $N(i)$ and condition number value $cnd(j)$.
- Compute the solution of $Ax = b$ using backslash.
- Find the relative residual $res(i, j) = \text{norm}(Ax - b)$.
- Using tic and toc, measures the time $T(i, j)$ that it takes to compute the solution and the residual.

Note both res and T are matrices of size $5 \times 4$.

(a) What is the relationship between cond($A$) and the residuals you observe? What can you expect for the solution relative error $||x - x^{true}||/||x^{true}||$?

(b) Based on the output of this script, plot $log10(N)$ against $log10(T)$ of the time it took to compute the solution and the residual. Use option '-o', make LineWidth thicker, add legends and axis labels.

(c) Use either the linear fitting tools (in the plot menu), polyfit or backslash to find the best linear fit for each of the 4 curves in the plot (each corresponds to a different condition number). Consider that, if

$$\log_{10} T \simeq k \log_{10} N + c$$

Then, this is telling us that

$$T \simeq (10^c) N^k$$

Based on this, conclude: does the time your computer takes to solve using backslash (Gaussian elimination) depend on N? Does it depend on the condition number? If so, how?

(d) Write a different script based on a modification of the previous one. Make it so that, for each linear system $A * x = b$, it solves the linear system using Conjugate Gradient instead of backslash. Do this by calling the function $pcg$ as follows:

$$[x, fl, resCG(i,j), iterCG(i,j)] = pcg(A, b, 1e-6, 10000);$$

Based on the output of this script, once again plot $log10(N)$ against $log10(T)$ of the time it took to compute the solution and the residual. Use either the linear fitting tools (in the plot menu), polyfit or backslash to find the best linear fit for each of the 4 curves in the plot (each corresponds to a different condition number). Based on this, answer the following: does the time your computer takes to solve using Conjugate Gradient depend on N? Does it depend on the condition number? If so, how?

3. *Applications of least squares:* Load the *LS_data2.mat* file included in the Files tab. It should include an array $X$ of size $9 \times 100$, as well as three vectors $Y, Y_2, Y_3$ of size $100 \times 1$.

(a) Present all columns of $X$ in the same plot, adding a legend to distinguish them. Corroborate that:

$$X = \begin{bmatrix} 1 & sin(x_i) & cos(x_i) & sin(2x_i) & cos(2x_i) & sin(3x_i) & cos(3x_i) & sin(x_i)^2 & cos(x_i)^2 \end{bmatrix}$$

for 100 equispaced points $x_i$ on $[0, 2\pi]$.

(b) $Y$ is a true linear combination of the columns of $X$. Using backslash, find a vector of coefficients. Explain: is this vector the unique solution for $X\alpha = Y$? If not, what is it?

(c) $Y_2$ and $Y_3$ are noisy versions of $Y$ (of the form $Y + \varepsilon$). Using backslash, find for each one a vector of coefficients $\alpha_2$ and $\alpha_3$. For each one, find the residual $||Y_i - X\alpha_i||^2$. Present $Y_i$ (using the 'o' format) and $X\alpha_i$ (using continuous red lines) in one plot for $Y_2$ and $Y_3$.

4. *Importing data and modeling:* In class, we imported population data from an excel spreadsheet to fit exponential and logistic growth models. Start by downloading the excel file *PredatorPreydata.xlsx* from the Files tab, and import it into your workspace. This file contains data with estimates for the population of two animal species: baboons (prey) and cheetahs (predators) in the hundreds of individuals at time t, for t from 0 to 86.13 months.

(a) Write a script that imports this file, and creates three variables: t, P_B and P_C representing time in months, population of Baboons and population of Cheetahs. Draw a plot representing both populations (make sure to change line width, add legends and axis labels). What do can you observe about the data for both populations?

(b) From observation and more research, you conclude the populations are both periodic, with a period of $T \simeq 14.357$ months. You recall from several mathematics courses that the best model for periodic data is a combination of sines and cosines, that is,

$$P(t) = a_0 + \sum_{k=1}^{m} a_k \cos\left(\frac{2\pi k}{T}t\right) + b_k \sin\left(\frac{2\pi k}{T}t\right)$$

write a Matlab function that, given a vector of times $t$ of size $N \times 1$ and positive integer $m$, generates the matrix $X$ of size $N \times 2m + 1$ where the first column of $X$ is all ones, and the following columns of X are the functions $\cos\left(\frac{2\pi k}{T}t\right)$ and $\sin\left(\frac{2\pi k}{T}t\right)$ for $k$ from 1 to m.

(c) Set $m = 10$. Call this Matlab function in your script to define $X$. Using backslash (least squares), find the vector of 21 coefficients $\alpha_B$ such that $P_B \simeq X\alpha_B$. Do the same thing for $P_C$, that is, find $\alpha_C$ such that $P_C \simeq X\alpha_C$ Plot both the population data and the models. How good are they?

(d) Someone suggests that a better model is to approximate the natural log of the populations with a combination of sines and cosines, that is, applying what you did before but to $\log(P_B)$ and $\log(P_C)$. Find $\beta_B$ such that $\log(P_B) \simeq X\beta_B$, and do the same for $P_C$.

(e) Plot $P_B, P_C$ and $exp(X\beta_B), exp(X\beta_C)$. Explain which model you consider best (between this and the previous one) and why.