Computer project 1

Hand-in report: In paper form at the beginning of the solution session on 16 April 2019 at 15:00.

Groups: The project should be done in groups of 1–3 students, and you organize the groups among yourselves. If you are looking for a group to join, then send a mail to `eskil.hansen@math.lth.se`

Report: Hand in your results, even if you did not fully succeed on some of the problems. In this case describe the problems you had. You should add code if it helps to clarify your results. Never hand in only code. The important parts are your solutions, results and conclusions. Include plots with your results whenever possible. It is preferable that you use Matlab to obtain solutions. If you make use of another program/language make sure to comment the code properly as your reviewers might be unfamiliar with it. Finally, leave some room in the report for the reviewer's comments.

Review: During the solution session on 16 April 2019, one report will be handed-out to each group for reviewing. Make sure to write your names on the report your reviewing. The review consists of trying to understand another groups report. The reviewing group should comment if they do not understand, correct if errors are found and if the report contains questions try to answer them. Be polite and try to give constructive feedback!

Hand-in review: During the lecture on 6 May 2019 at 8:00.

Interpolating population data

The aim of this project is to study interpolation and extrapolation in the context of population data and to investigate the role that the basis functions play for the accuracy (conditioning) of the interpolation procedure.

To this end, we consider the Swedish population between 1940 and 2000 (data from Statistiska centralbyrån):

Year (t_i)	Population (y_i)
1940	6,371,432
1950	7,041,829
1960	7,497,967
1970	8,081,229
1980	8,317,937
1990	8,590,630
2000	8,882,792

The influence of the basis

The seven data points (t_0, y_0) to (t_6, y_6) can be interpolated by a sixth degree polynomial $p_6(t)$, which may be represented in many different ways.
If we choose a particular basis $\{\phi_0(t), \phi_1(t), \ldots, \phi_6(t)\}$ then our interpolation polynomial is represented as

$$p_6(t) \;=\; c_0\phi_0(t) + c_1\phi_1(t) + \ldots + c_6\phi_6(t)$$

and the task of interpolating the data points is then to find the coefficients c_i such that

$$p_6(t_i) = y_i, \quad i = 0, \ldots, 6.$$

As we derived in class, the coefficients can be computed by solving a linear system $Ac = y$, where the matrix A depends on the basis functions.

In order to investigate the role of the basis, we will consider the following basis functions:

$$\begin{array}{ll} a) & \phi_i(t) = t^i \\ b) & \phi_i(t) = (t - 1940)^i \\ c) & \phi_i(t) = (t - 1970)^i \\ d) & \phi_i(t) = ((t - 1970)/30)^i \end{array}$$

Problem 1. Present the linear system $Ac = y$ which you need to solve in order to find the coefficients c_i and compute the condition number of the matrix A for each basis. Compare the condition numbers and explain your results. Do your results depend on the choice of the matrix norm used to compute the condition number?

Interpolation and extrapolation

After having determined the coefficients of our interpolation polynomial $p_6(t)$, it remains to evaluate $p_6(t)$ for different times t. The most straightforward approach is to compute each term $c_i\phi_i(t)$ individually and thereafter to sum them all up in order to obtain $p_6(t)$. However, we can save quite a lot computations if we apply *Horner's method.* To illustrate the method we consider the first basis (a), i.e.,

$$p_6(t) = c_0 + c_1 t + \ldots + c_6 t^6.$$

We rewrite the polynomial in nested form as

$$p_6(t) = c_0 + t(c_1 + t(c_2 + t(c_3 + t(c_4 + t(c_5 + tc_6)))))$$

and the polynomial can then be evaluated by first computing $p_1(t) = c_5 + tc_6$, then $p_2(t) = c_4 + tp_1(t)$ and after six iterations we obtain $p_6(t)$.

Problem 2. Represent the polynomial $p_6(t)$ in the best-conditioned basis found in the previous part. Rewrite $p_6(t)$ in nested form and evaluate it

at one-year intervals by using Horner's method. Finally plot $p_6(t)$ together with the interpolation points.

Problem 3. Use the polynomial $p_6(t)$ to extrapolate the population to 2010. How close is the extrapolated value to the true value of 9,345,135? Explain your observations.

Newton interpolation and cubic splines

A different approach to interpolation, which does not result in solving a *full* 7×7 system, is to use Newton interpolation. The basis functions are then given as

$$\phi_i(t) = \prod_{k=0}^{i-1} (t - t_k),$$

where $\phi_0(t) = 1$ by definition. The coefficients of the interpolation polynomial is then obtained by either solving a lower triangular system (of size 7×7) or by using the so called divided differences.

Problem 4. Determine the Newton form of $p_6(t)$, i.e., represent the polynomial interpolating the seven data points with the Newton basis and use divided differences to find the coefficients of the polynomial. Next, determine the Newton form of the seventh degree polynomial $p_7(t)$ that also interpolates the data given for 2010, without starting from scratch (use the Newton form of $p_6(t)$, which is already computed, to determine $p_7(t)$). Plot both polynomials over the interval from 1940 to 2010.

Finally, we will try out a completely different approach to interpolation. Instead of fitting one single polynomial to all data points, we can fit a low order polynomial between each pair of data points. The result, which we refer to as a *spline*, will be a chain of polynomials fitted together at the data points (t_i, y_i). The simplest example is when one uses linear polynomials and the spline is then obtained by simply "connecting the dots (t_i, y_i) " with straight line segments.

Problem 5. Use the built in Matlab function `spline` to interpolate the seven data points by a spline consisting of cubic polynomials. Read through the documentation of the `spline` function and write a short description of what it does and how you have used it in your own code. Compare the cubic spline interpolation with the previous interpolation you obtained in problem 2 (and 4) by plotting them in the same graph with one-year intervals between 1940 and 2000. Is there any difference? Which type of interpolation would you recommend and why?