UCSC AM 147: Computational Methods and Applications

Additional Exercises with Solutions

Abhishek Halder

Problem 1

Numerical errors

The MATLAB command `chop(x,n)` rounds x up to n significant digits. For example,

```
>> format long
>> pi
ans =
   3.141592653589793
>> chop(pi,5)
ans =
   3.141600000000000
```

Consider two functions $f(x) = \dfrac{1-\cos^2(x)}{x^2}$ and $g(x) = \dfrac{\sin^2 x}{x^2}$.

(a) Are the two functions f and g equal? Why/why not?

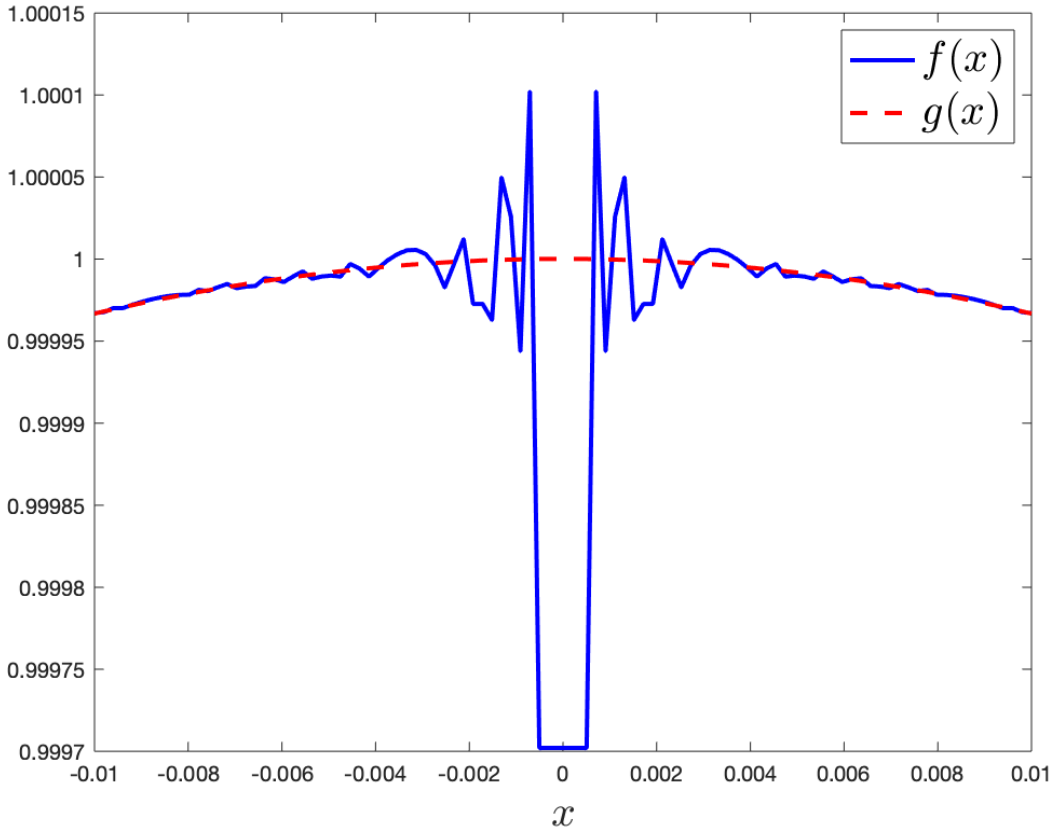
Solution: Yes, $f = g$. This is because $1 - \cos^2 x = \sin^2 x$ for all x. At $x = 0$, both functions have the same limiting value:

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} g(x) = 1.$$

(b) Now consider the following commands typed in a MATLAB terminal:

```
>> x = linspace(-0.01,0.01,100);
>> f = (1 - (chop(cos(x),10)).^2)./(x.^2);
>> g = ((chop(sin(x),10)).^2)./(x.^2);
>> plot(x, f, '-b', x, g, '--r', 'LineWidth', 2)
```

which produces the figure below. Explain the discrepancy you notice.



Solution: In this case, large round-off error occurred due to subtraction of two numbers which are almost (but not exactly) equal. (We saw before another instance of this phenomenon in Lecture 4, pages 9–10.) When x is small, $1 \approx \cos^2 x$. So the round-off error in $\cos x$ becomes significant in the evaluation of $f(x)$.

To get a concrete example of what's going on, fix $x = 5 \times 10^{-5}$ (a small number). Then, $\cos(x) = 0.99999999875000...$, which rounded to 10 significant digits gives 0.9999999988. Using this rounded value of $\cos(x)$, we then get

$$f(x) = \frac{1 - \cos^2 x}{x^2} = \frac{1 - (0.9999999988)^2}{(5 \times 10^{-5})^2} = 0.9599...$$

which has only one significant digit (the true value is 0.9999...). However, since the evaluation of $g(x)$ does not involve a subtraction, the round off error does not grow very large. For example, at $x = 5 \times 10^{-5}$, the true value of $\sin(x) = 0.499999999791667... \times 10^{-5}$ rounded to 10 significant digits is $0.4999999998 \times 10^{-5}$. Using this rounded value of $\sin(x)$, we then get

$$g(x) = \frac{\sin^2 x}{x^2} = \frac{(0.4999999998 \times 10^{-5})^2}{(5 \times 10^{-5})^2} = 0.9999...$$

which has ten correct significant digits. In summary, f and g are mathematically equivalent, but numerically not so because of the round-off error.

Problem 2

Newton's method

(a) Use Newton's method to derive a recursive algorithm to compute the square root of N.

Solution: To compute $x = \sqrt{N}$ is same as solving for x from the nonlinear equation $f(x) = x^2 - N = 0$. Since $f'(x) = 2x$, the Newton's method, in this case, becomes

$$x_{k+1} = x_k - \frac{x_k^2 - N}{2x_k} = \frac{1}{2}\left(x_k + \frac{N}{x_k}\right).$$

(b) Suppose we use the above recursion from part (a) to compute $\sqrt{-1}$. Since the real number sequence $\{x_k\}$ cannot approach the true imaginary solutions $\pm i$, the Newton's method will fail. To investigate exactly how does it fail, let us start from the initial guess $x_0 = \frac{1}{\sqrt{3}}$ and compute (by hand) first three iterates x_1, x_2, x_3. What do you think is happening?

Solution: We get $x_1 = -\frac{1}{\sqrt{3}}$, $x_2 = +\frac{1}{\sqrt{3}}$, and $x_3 = -\frac{1}{\sqrt{3}}$. The recursion from part (b) results in an oscillation $\{\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, ...\}$.

Problem 3

Fixed point equation

We would like to solve the fixed point equation $x = g(x)$ where $g(x) = -\frac{5}{2}x^3 + \frac{15}{2}x^2 - 5x + 1$. This fixed point equation has 3 roots. Prove that $x = 1$ is a root. Find the other two roots.

Solution: By direct substitution, we verify that $1 = g(1)$, and hence $x = 1$ is a root.
To find the other two roots, we rewrite $x = g(x)$ as $f(x) = -5x^3 + 15x^2 - 12x + 2 = 0$. Since $f(x)$ is a cubic polynomial with $x = 1$ as one root, we must have $f(x) = (x-1)(ax^2 + bx + c)$ for some to-be-determined constants a, b, c. Equating the coefficients, we get $a = -5, b = 10, c = -2$. Now, the quadratic $-5x^2 + 10x - 2$ has two roots $1 \mp \sqrt{3/5}$. Therefore, the three roots of $f(x) = 0$, or equivalently of $x = g(x)$ are $x = 1, 1 \mp \sqrt{3/5}$.

Problem 4

Number of solutions for a square linear system

Show that the following system of equations:

$$\begin{aligned} x_1 + 4x_2 + \alpha x_3 &= 6, \\ 2x_1 - x_2 + 2\alpha x_3 &= 3, \\ \alpha x_1 + 3x_2 + x_3 &= 5, \end{aligned}$$

has (a) unique solution when $\alpha = 0$, (b) no solution when $\alpha = -1$, and (c) infinitely many solutions when $\alpha = 1$.

Solution: (a) A system of linear equations of the form $\boldsymbol{Ax} = \boldsymbol{b}$ has unique solution if and only if $\det(\boldsymbol{A}) \neq 0$. In our case,

$$\boldsymbol{A} = \begin{pmatrix} 1 & 4 & \alpha \\ 2 & -1 & 2\alpha \\ \alpha & 3 & 1 \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} 6 \\ 3 \\ 5 \end{pmatrix}.$$

We have $\det(\boldsymbol{A}) = \alpha^2 - 1$, in which substituting $\alpha = 0$ gives $-1 \neq 0$. Hence there is a unique solution.

(b). For $\alpha = \pm 1$, $\det(\boldsymbol{A}) = 0$, and it is not clear, just by looking at the determinant, when we have no solution, and when we have infinitely many solutions. To examine what happens for $\alpha = -1$, we start solving the system $\boldsymbol{Ax} = \boldsymbol{b}$ as follows. For $\alpha = -1$, we perform: (first equation) $-\frac{1}{2}\times$ (second equation). This gives $x_2 = 1$. Substituting $x_2 = 1$ in the third equation results $-x_1 + x_3 = 2$. Similarly, substituting $x_2 = 1$ in the second equation yields $x_1 - x_3 = 2$. Adding them together, we get:

$$(-x_1 + x_3) + (x_1 - x_3) = 2 + 2, \quad \Leftrightarrow \quad 0 = 4,$$

which is impossible. Thus, for $\alpha = -1$, the system has no solution.

(c) To investigate the case $\alpha = 1$, we eliminate x_2 from the first and second equation to obtain $x_1 + x_3 = 2$, which in turn gives $x_2 = 1$. We get the exactly same result if we eliminate x_2 from the second and third equations. Thus, choosing x_3 as the "free variable", the solution set is given by:

$$x_1 = 2 - x_3, \quad x_2 = 1, \quad x_3 = x_3.$$

In other words, there are infinitely many solutions depending on what value we choose for x_3.

Problem 5

Square linear system and ill-conditioned matrix

Consider the matrix $\boldsymbol{A} = \begin{pmatrix} 1+\varepsilon & 1 & 2 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$.

(a) Argue why $\boldsymbol{A}$ is singular for $\varepsilon = 0$.

Solution: Substituting $\varepsilon = 0$ we directly compute the 3×3 determinant, to find that determinant equals $1 \times (-1 - 0) - 1 \times (1 - 0) + 2 \times (0 - (-1)) = 0$. Hence singular.

(b) Verify that for $\varepsilon \neq 0$, the inverse is given by

$$\boldsymbol{A}^{-1} = \frac{1}{\varepsilon} \begin{pmatrix} 1 & 1 & -2 \\ 1 & 1-\varepsilon & -2 \\ -1 & -1 & 2+\varepsilon \end{pmatrix}.$$

Solution: We directly verify if the above expression satisfies the definition of the matrix inverse: $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$. For this purpose, we compute the matrix-matrix product:

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \frac{1}{\varepsilon} \begin{pmatrix} \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{pmatrix} = \boldsymbol{I}.$$

This verifies the claim.

(c) Show that for $\varepsilon \neq 0$, we have $\kappa_2(\boldsymbol{A}) \geq 1/|\varepsilon|$. What is the implication of this result?

Solution: Consider a specific 3×1 unit vector: $\boldsymbol{z} = (1, 0, 0)^\top$. By definition,

$$\|\boldsymbol{A}\|_2 \geq \|\boldsymbol{A}\boldsymbol{z}\|_2 = \sqrt{(1+\varepsilon)^2 + 1^2 + 1^2} \geq \sqrt{2} > 1.$$

Likewise,

$$\|\boldsymbol{A}^{-1}\|_2 \geq \|\boldsymbol{A}^{-1}\boldsymbol{z}\|_2 = \frac{1}{|\varepsilon|}\sqrt{1^2 + 1^2 + (-1)^2} = \frac{\sqrt{3}}{|\varepsilon|} > \frac{1}{|\varepsilon|}.$$

Therefore, $\kappa_2(\boldsymbol{A}) := \|\boldsymbol{A}\|_2 \|\boldsymbol{A}^{-1}\|_2 \geq 1/|\varepsilon|$. The implication is that when $|\varepsilon|$ is small, then the condition number is large, that is, the matrix $\boldsymbol{A}$ is then ill-conditioned.

(d) Consider the following sequence of commands typed in MATLAB terminal.

```
>> format long;
>> eps = 1e-4;
>> A = [(1 + eps), 1, 2; 1, -1, 0; 1, 0, 1];
>> b = ones(3,1);
>> x = A\b
```

The output is

```
x =
  -0.000000000000375
  -1.000000000000375
   1.000000000000375
```

If we now type

```
>> b_perturbed = b + [0.001; 0; 0];
>> x_perturbed = A\b_perturbed
```

then the output is

```
x_perturbed =
  10.000000000000613
   9.000000000000611
  -9.000000000000611
```

What qualitative observation can you make from this? Give reasons to explain your observation.

Solution: We observe that small change in vector $\boldsymbol{b}$ produces a large change in the solution vector $\boldsymbol{x}$. The reason is that the matrix $\boldsymbol{A}$ is ill-conditioned as proved in part (c), that is, $\kappa_2(\boldsymbol{A})$ is much larger than 1.

Problem 6

Weighted regression

We have experimentally collected N data samples:

$$\left(x^{(i)}, y^{(i)}\right), \qquad i = 1, 2, \ldots, N,$$

and want to perform a regression $y \approx \widehat{f}\,(x) := \theta_1 f_1(x) + \theta_2 f_2(x) + \ldots + \theta_p f_p(x)$, where a dictionary of functions $f_1, f_2, ..., f_p$ is given.

Suppose we also know that some experimental data are more accurate than others. This is expressed by a known collection of normalized weights $w_1, w_2, \ldots, w_N$ satisfying $w_i \geq 0$, $\sum_{i=1}^{N} w_i = 1$. For example, if w_i is close to 1 (say, 0.9) then the ith sample is a highly accurate experimental data. On the other hand, if w_i is close to 0 (say, 0.1) then the ith sample is a rather inaccurate experimental data.

Then it is natural to perform weighted regression that takes into account this variable accuracy among data samples by minimizing the weighted mean square error, i.e., by solving

$$\arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \sum_{i=1}^{N} w_i \left(y^{(i)} - \widehat{f}\left(x^{(i)}\right)\right)^2.$$

Re-write the above weighted regression problem in standard/ordinary least squares form

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^p}{\text{minimize}} \, \|\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{y}\|_2^2.$$

In other words, find the matrix $\boldsymbol{A}$, and the vector $\boldsymbol{y}$, as well as their dimensions.

Remark: Notice that the problem in Lec. 19, p. 5, is a special case: $w_1 = w_2 = \ldots = w_N$.

Solution: Let us define the diagonal matrix $\boldsymbol{W} := \operatorname{diag}(w_1, w_2, \ldots, w_N)$. Then, $\sqrt{\boldsymbol{W}} = \operatorname{diag}\left(\sqrt{w_1}, \sqrt{w_2}, \ldots, \sqrt{w_N}\right)$. Also, let

$$\widetilde{\boldsymbol{A}} := \begin{pmatrix} f_1(x^{(1)}) & \ldots & f_p(x^{(1)}) \\ f_1(x^{(2)}) & \ldots & f_p(x^{(2)}) \\ \vdots & \vdots & \vdots \\ f_1(x^{(N)}) & \ldots & f_p(x^{(N)}) \end{pmatrix}, \qquad \widetilde{\boldsymbol{y}} := \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix}, \qquad \boldsymbol{\theta} := \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}.$$

Now notice that the given objective

$$\sum_{i=1}^{N} w_i \left(y^{(i)} - \widehat{f}\left(x^{(i)}\right)\right)^2 = \left(\widetilde{\boldsymbol{A}}\boldsymbol{\theta} - \widetilde{\boldsymbol{y}}\right)^\top \boldsymbol{W} \left(\widetilde{\boldsymbol{A}}\boldsymbol{\theta} - \widetilde{\boldsymbol{y}}\right) = \left\| \sqrt{\boldsymbol{W}} \left(\widetilde{\boldsymbol{A}}\boldsymbol{\theta} - \widetilde{\boldsymbol{y}}\right) \right\|_2^2 = \|\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{y}\|_2^2,$$

where $\boldsymbol{A} := \sqrt{\boldsymbol{W}}\,\widetilde{\boldsymbol{A}}$ and $\boldsymbol{y} := \sqrt{\boldsymbol{W}}\,\widetilde{\boldsymbol{y}}$. We have $\boldsymbol{A} \in \mathbb{R}^{N\times p}$, $\boldsymbol{\theta} \in \mathbb{R}^p$, and $\boldsymbol{y} \in \mathbb{R}^N$.

Problem 7

Least norm solution

Given nonsingular $\boldsymbol{P} \in \mathbb{R}^{n\times n}$, and full row rank matrix $\boldsymbol{C} \in \mathbb{R}^{m\times n}$ where $m < n$, express the

problem

$$\begin{aligned} &\underset{\boldsymbol{y}\in\mathbb{R}^n}{\text{minimize}} \ \|\boldsymbol{P}\boldsymbol{y}+\boldsymbol{q}\|_2 \\ &\text{subject to } \boldsymbol{C}\boldsymbol{y}=\boldsymbol{d} \end{aligned}$$

in standard least norm form:

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} \ \|\boldsymbol{x}\|_2 \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \end{aligned}$$

for appropriately defined matrix $\boldsymbol{A}$, and vectors $\boldsymbol{x}, \boldsymbol{b}$.

Explain if and how it is possible to compute the solution of the original problem.

Solution: Introducing the change-of-variable $\boldsymbol{x} := \boldsymbol{P}\boldsymbol{y}+\boldsymbol{q}$, we get

$$\boldsymbol{C}\boldsymbol{y} = \boldsymbol{C}\boldsymbol{P}^{-1}\boldsymbol{x} - \boldsymbol{C}\boldsymbol{P}^{-1}\boldsymbol{q}.$$

We thus arrive at the standard least norm form:

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} \ \|\boldsymbol{x}\|_2 \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \end{aligned}$$

with

$$\boldsymbol{A} := \boldsymbol{C}\boldsymbol{P}^{-1}, \quad \boldsymbol{b} := \boldsymbol{C}\boldsymbol{P}^{-1}\boldsymbol{q}+\boldsymbol{d}.$$

Since $\text{rank}(\boldsymbol{A}) = \text{rank}\,(\boldsymbol{C}\boldsymbol{P}^{-1}) \leq \min\{m,n\} = m$, and Sylvester's rank inequality gives $\text{rank}\,(\boldsymbol{C})+\text{rank}\,(\boldsymbol{P}^{-1})-n = m \leq \text{rank}\,(\boldsymbol{C}\boldsymbol{P}^{-1})$, therefore, $\text{rank}(\boldsymbol{A}) = \text{rank}\,(\boldsymbol{C}\boldsymbol{P}^{-1}) = m$, i.e., a full row rank matrix. So we can indeed compute the unique solution of the re-written problem using Lec. 20, p. 2.

Problem 8

Numerical differentiation

The purpose of this exercise is to help you appreciate that the tools we learnt in this course are also useful in solving partial differential equations (PDEs). To work on this exercise, you don't need any prior knowledge of PDEs.

Consider the one dimensional heat equation

$$\frac{\partial u}{\partial t} = \alpha^2 \frac{\partial^2 u}{\partial x^2}, \quad 0 \leq x \leq L, \quad 0 \leq t \leq T,$$

that describes the evolution of temperature $u(x,t)$ along a bar of length L, as a function of the position x and time t. Here α denotes the thermal diffusivity of the material. Let us assume the following boundary conditions:

$$u(0,t) = u(L,t) = 0 \quad \text{for} \quad 0 < t < T, \quad u(x,0) = \phi(x) \quad \text{for} \quad 0 \le x \le L.$$

Partition the space interval $[0, L]$ as

$$x_i = i\Delta x, \quad i = 0, 1, \ldots, M, \quad \Delta x = \frac{L}{M}.$$

Partition the time interval $[0, T]$ as

$$t_k = k\Delta t, \quad k = 0, 1, \ldots, N, \quad \Delta t = \frac{T}{N}.$$

Let us introduce the notation $u_{i,k} = u(x_i, t_k)$.

(a) Write the two point forward difference approximation for the left hand side time derivative of the heat equation. Also write the three point central difference approximation for the right hand side spatial derivative of the heat equation.

Solution: The two point forward difference approximation for the first derivative with respect to time gives

$$\frac{\partial u}{\partial t} \approx \frac{u_{i,k+1} - u_{i,k}}{\Delta t}.$$

The three point central difference approximation for second derivative with respect to space, gives

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u_{i+1,k} - 2u_{i,k} + u_{i-1,k}}{(\Delta x)^2}.$$

(b) Use your answer from part (a) to approximate the heat equation as

$$u_{i,k+1} = (1 - 2\lambda)\, u_{i,k} + \lambda \left(u_{i+1,k} + u_{i-1,k}\right),$$

where you need to determine the parameter λ as function of $\alpha, \Delta t, \Delta x$.

Solution: Substituting the finite difference approximations from part (a) in the heat equation, we get

$$\frac{u_{i,k+1} - u_{i,k}}{\Delta t} = \alpha^2 \frac{u_{i+1,k} - 2u_{i,k} + u_{i-1,k}}{(\Delta x)^2}$$

$$\Rightarrow u_{i,k+1} - u_{i,k} = \lambda \left(u_{i+1,k} - 2u_{i,k} + u_{i-1,k}\right)$$

$$\Rightarrow u_{i,k+1} = (1 - 2\lambda)\, u_{i,k} + \lambda \left(u_{i+1,k} + u_{i-1,k}\right),$$

where $\lambda := \dfrac{\alpha^2 \Delta t}{(\Delta x)^2}$. The above formula is often called the FTCS (forward in time, central in space) approximation.

(c) Next, notice that the boundary conditions can be written in discrete form: $u_{i,0} = \phi(x_i)$ for $i = 0, 1, \ldots, M$, and $u_{0,k} = u_{M,k} = 0$ for $k = 1, \ldots, N$. Using this information, and your answer in part (b), rewrite the approximated heat equation in matrix-vector form

$$\begin{pmatrix} u_{1,k+1} \\ u_{2,k+1} \\ \vdots \\ u_{M-1,k+1} \end{pmatrix} = \begin{pmatrix} a & b & 0 & 0 & 0 & \ldots & 0 \\ b & a & b & 0 & 0 & \ldots & 0 \\ 0 & b & a & b & 0 & \ldots & 0 \\ 0 & 0 & b & a & b & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & b & a \end{pmatrix} \begin{pmatrix} u_{1,k} \\ u_{2,k} \\ \vdots \\ u_{M-1,k} \end{pmatrix},$$

where you need to determine the constants a, b as function of the parameter λ.

Solution: We substitute $i = 1, 2, ..., M-1$, in the recursive FTCS formula derived in part (b), and utilize the boundary conditions given in the question, to get the matrix-vector equation

$$\begin{pmatrix} u_{1,k+1} \\ u_{2,k+1} \\ \vdots \\ u_{M-1,k+1} \end{pmatrix} = \begin{pmatrix} 1-2\lambda & \lambda & 0 & 0 & 0 & \ldots & 0 \\ \lambda & 1-2\lambda & \lambda & 0 & 0 & \ldots & 0 \\ 0 & \lambda & 1-2\lambda & \lambda & 0 & \ldots & 0 \\ 0 & 0 & \lambda & 1-2\lambda & \lambda & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \lambda & 1-2\lambda \end{pmatrix} \begin{pmatrix} u_{1,k} \\ u_{2,k} \\ \vdots \\ u_{M-1,k} \end{pmatrix}.$$

Comparing the above matrix-vector equation with the one given in the question, we deduce that $a = 1 - 2\lambda$, and $b = \lambda$.

(d) Assuming the heat distribution at time $t = 0$ as $u(x, 0) = \phi(x) = \sin\left(\dfrac{\pi x}{L}\right)$, the analytical solution is

$$u(x, t) = \sin\left(\frac{\pi x}{L}\right) \exp\left(-\frac{\pi^2 \alpha t}{L^2}\right).$$

To numerically solve the heat equation, fix $\alpha = 1$, $L = 1$, $\Delta x = 0.01$, $\Delta t = 0.001$, $T = 1$. Write a MATLAB code `PracticeProblem7d.m` to iteratively solve the approximate heat equation derived in part (c) with the same initial heat distribution as above. Your code should plot the initial heat distribution, numerically computed heat distributions for the first 5 time steps, as well as the analytically computed heat distributions for the same time steps, all in the same

figure window. Please use different color/linestyles/legends, as you feel appropriate, to clarify which curve is which.

Solution: See the posted code in CANVAS File Section folder: "HW Problems and Solutions", filename `PracticeProblem7d.m`. In the figure generated by this code, the green line plot is the initial distribution. The blue solid lines are the FTCS approximate solutions, and the red dashed lines are the analytical solutions for the first 5 time steps.

(e) Looking at your plot from part (d), which location of the rod is most hot at any given time t? Explain what you observe.

Solution: From the plots generated in via the MATLAB script in part (d), it is observed that the midpoint of the rod (that is, the location $x = L/2$) is most hot at all times t. The intuitive reasoning lies in the boundary conditions, which say that both ends of the rod (location $x = 0$ and $x = L$) are kept at zero temperature for all times, that is, $u_{0,k} = u_{M,k} = 0$ for all $k = 0, 1, 2, \ldots, N$. As a result, the temperature increases as we go away from each end. Farther the location is from one end, higher the temperature. The midpoint is the farthest location from both ends, hence the answer.
We can also explain this from the analytical solution as follows. At any fixed time t, maximum temperature occurs at that x which solves

$$\frac{\partial u(x,t)}{\partial x} = 0 \;\Rightarrow\; \exp\left(-\frac{\alpha\pi^2 t}{L^2}\right)\left(\frac{\partial}{\partial x}\sin\frac{\pi x}{L}\right) = 0 \;\Rightarrow\; \cos\frac{\pi x}{L} = 0 = \cos\frac{\pi}{2} \;\Rightarrow\; x = L/2.$$

You may check that the solution $x = L/2$ corresponds to a maximum, not a minimum or saddle, by verifying that the second derivative is negative at $x = L/2$.

Problem 9

Numerical integration

Consider the definite integral $I := \displaystyle\int_0^1 \frac{x^4(1-x)^4}{1+x^2}\,\mathrm{d}x$.

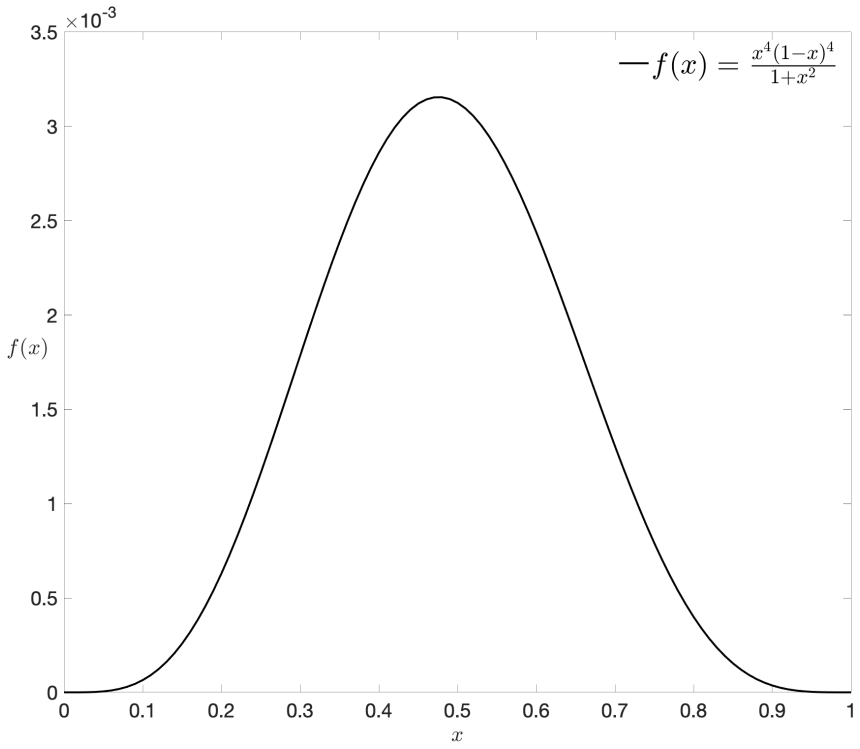
(a) Just by looking at the integrand, what can you tell about the sign of I (without any explicit calculation)? Give reasons to support your answer.

(b) We can analytically compute I as

$$\begin{aligned} I &= \int_0^1 \frac{x^4 - 4x^5 + 6x^6 - 4x^7 + x^8}{1 + x^2} \\ &= \int_0^1 \left(x^6 - 4x^5 + 5x^4 - 4x^2 + 4 - \frac{4}{1 + x^2} \right) \, \mathrm{d}x \\ &= \left(\frac{x^7}{7} - \frac{2x^6}{3} + x^5 - \frac{4x^3}{3} + 4x - 4 \arctan x \right) \Big|_{x=0}^{x=1} \\ &= \frac{1}{7} - \frac{2}{3} + 1 - \frac{4}{3} + 4 - \pi \\ &= \frac{22}{7} - \pi. \end{aligned}$$

Combine your answer in part (a) with the above exact value for I, to derive an inequality involving π.

(c) The integrand plotted over the domain $[0, 1]$ is shown below. Use this figure to approximate I using the trapezoid method with x-axis partition $[0, 1/2]$, $[1/2, 1]$. Keep your answer in fractions (rational number format).



Solution: (a) Since the integrand is zero at the two endpoints ($x = 0$ and $x = 1$) and positive everywhere in the open interval $(0, 1)$, hence the area enclosed by it must be positive, that is, $I > 0$.

(b) Combining part (a) with the exact value, we deduce: $\pi < \dfrac{22}{7}$.

(c) From the figure, we notice that the integrand curve $f(x)$ is symmetric about $x = 0.5$. Hence numerical approximation of I using the trapezoid method with the given partition, equals the area of the triangle with vertices $(0,0)$, $(1/2, f(1/2))$, $(1,0)$. Hence

$$I_{\text{trapezoid}} = \frac{1}{2} \times \underbrace{1}_{\text{base}} \times \underbrace{f(1/2)}_{\text{height}} = \frac{1}{2} \times \frac{2^{-8}}{1+2^{-2}} = \frac{1}{640}.$$

You can check in decimals that this gives a reasonably good approximation of the true answer.

Problem 10

ODE IVP

Consider the ODE IVP:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = x + y - xy, \qquad y(0) = 1.$$

(a) Is this a linear or nonlinear ODE? Give reasons to support your answer.

(b) Use the forward Euler method to approximate $y(0.3)$ with stepsize $h = 0.1$.

Solution: (a) This is a linear ODE because it can be written as $y' + (x-1)y = x$, where the left hand side is linear in y and its derivatives.

(b) In the forward Euler method, we use the approximation $y(x+h) \approx y(x) + hy'(x)$. For our ODE initial value problem with stepsize $h = 0.1$, this gives

$y(0.1) = y(0+0.1) = y(0) + 0.1 \times y'(0) = 1 + 0.1 \times 1 = 1.1,$

$y(0.2) = y(0.1+0.1) = y(0.1) + 0.1 \times y'(0.1) = 1.1 + 0.1 \times (0.1 + 1.1 - 0.1 \times 1.1) = 1.209,$

$y(0.3) = y(0.2+0.1) = y(0.2) + 0.1 \times y'(0.2) = 1.209 + 0.1 \times (0.2 + 1.209 - 0.2 \times 1.209) = 1.32572.$

Problem 11

PageRank Algorithm

Recall that the PageRank vector $\boldsymbol{p}(\alpha) \in \mathbb{R}^n$ is the eigenvector associated with the dominant (in the maximum modulus sense) eigenvalue unity of the $n \times n$ Google matrix $\boldsymbol{G} = \alpha \boldsymbol{S} + (1-\alpha)\boldsymbol{q}\boldsymbol{1}^\top$, where the vector $\boldsymbol{q} \in \mathbb{R}^n$ is elementwise nonnegative and $\boldsymbol{1}^\top \boldsymbol{q} = 1$. As always, $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ is

column-stochastic, and $\mathbf{1}$ denotes the $n \times 1$ vector of ones. In this exercise, we will use (but not prove) the fact that $\boldsymbol{p}(\alpha)$ is a differentiable function of the damping factor α, where $0 < \alpha < 1$. The purpose of this exercise is to estimate how sensitive is $\boldsymbol{p}(\alpha)$ with respect to change in α, that is, to estimate the norm of the $n \times 1$ vector $\dfrac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha}$. We will proceed in the following steps.

(a) A matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ is called strictly diagonally dominant if the magnitude of the diagonal entry in each row is greater than the sum of the magnitudes of the non-diagonal entries in that row: $|M_{ii}| > \sum_{j \neq i} |M_{ij}|$ for all $i = 1, \ldots, n$. Show that a diagonally dominant matrix must be nonsingular.

(b) Use part (a) to prove that the matrix $\boldsymbol{I} - \alpha \boldsymbol{S}$ is nonsingular.

(c) Use part (b) to prove the identity: $\dfrac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha} = (\boldsymbol{I} - \alpha \boldsymbol{S})^{-1} \left(\boldsymbol{S} - \boldsymbol{q}\mathbf{1}^\top\right) \boldsymbol{p}$.

(d) Use part (c) to deduce that $\left\| \dfrac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha} \right\|_1 \leq \dfrac{2}{1-\alpha}$. What is the implication of this result?

Solution: (a) We prove by contradiction. Suppose if possible, $\boldsymbol{M}$ is diagonally dominant and singular, that is, $\det(\boldsymbol{M}) = 0$. Then, there exists a non-zero vector $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{M}\boldsymbol{x} = \mathbf{0}$. Choose a component of $\boldsymbol{x}$ that has maximum magnitude, that is, choose $r = \arg\max_i |x_i|$. Now consider the r-th row of the matrix-vector product $\boldsymbol{M}\boldsymbol{x}$, given by

$$M_{rr} x_r + \sum_{j \neq r} M_{rj} x_j = 0.$$

Therefore, $|M_{rr}||x_r| = \left| \sum_{j \neq r} M_{rj} x_j \right| \leq \sum_{j \neq r} |M_{rj}||x_j| \leq \left(\sum_{j \neq r} |M_{rj}| \right) |x_r|$, where the last inequality used $|x_j| \leq |x_r|$ for all $j = 1, \ldots, n$. Since $|x_r| \neq 0$, we thus get $|M_{rr}| \leq \left(\sum_{j \neq r} |M_{rj}| \right)$, which is a contradiction to $\boldsymbol{M}$ being diagonally dominant.

(b) Since $\lambda_i(\boldsymbol{I} - \alpha \boldsymbol{S}) = \lambda_i\left((\boldsymbol{I} - \alpha \boldsymbol{S})^\top\right) = \lambda_i\left(\boldsymbol{I} - \alpha \boldsymbol{S}^\top\right)$, hence letting $\boldsymbol{M} := \boldsymbol{I} - \alpha \boldsymbol{S}^\top$, we note that $M_{ii} = 1 - \alpha S_{ii}^\top = \underbrace{\alpha\left(1 - S_{ii}^\top\right)}_{>0} + \underbrace{(1-\alpha)}_{>0} = |M_{ii}|$. On the other hand, $M_{ij} = -\alpha S_{ij}^\top \Rightarrow |M_{ij}| = \alpha S_{ij}^\top \Rightarrow \sum_{j \neq i} |M_{ij}| = \alpha \sum_{j \neq i} S_{ij}^\top = \alpha\left(1 - S_{ii}^\top\right)$. Consequently, $|M_{ii}| > \sum_{j \neq i} |M_{ij}|$ for all $i = 1, \ldots, n$. Therefore, $I - \alpha \boldsymbol{S}^\top$ is strictly diagonally dominant, and by part (a), must be nonsingular. So its transpose, $\boldsymbol{I} - \alpha \boldsymbol{S}$ must also be nonsingular.

(c) Differentiating both sides of $\mathbf{1}^\top \boldsymbol{p} = 1$ with respect to α, we get $\mathbf{1}^\top \dfrac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha} = 0$. On the other hand, differentiating both sides of the dominant eigenvalue equation $\boldsymbol{G}\boldsymbol{p} = \boldsymbol{p}$ with respect to α gives

$$\frac{\mathrm{d}\boldsymbol{G}}{\mathrm{d}\alpha}\boldsymbol{p} + \boldsymbol{G}\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha} = \frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha} \;\Rightarrow\; \left(\boldsymbol{S} - \boldsymbol{q}\mathbf{1}^\top\right)\boldsymbol{p} = (\boldsymbol{I} - \boldsymbol{G})\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha} = (\boldsymbol{I} - \alpha\boldsymbol{S})\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha},$$

wherein the last equality used $\mathbf{1}^\top \dfrac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha} = 0$. Since $\boldsymbol{I} - \alpha\boldsymbol{S}$ is invertible from part (b), we pre-multiply both sides of the above equality by $(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}$ to get the desired identity.

(d) Taking 1-norm to both sides of the identity derived in part (c), we obtain

$$\begin{aligned}\left\|\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha}\right\|_1 \le \|\,(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\,\|_1\,\|\left(\boldsymbol{S} - \boldsymbol{q}\mathbf{1}^\top\right)\boldsymbol{p}\|_1 &= \|\,(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\,\|_1\,\|\boldsymbol{S}\boldsymbol{p} + \left(-\boldsymbol{q}\mathbf{1}^\top\right)\boldsymbol{p}\|_1 \\ &\le \|\,(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\,\|_1\,\left(\|\boldsymbol{S}\boldsymbol{p}\|_1 + \|\boldsymbol{q}\mathbf{1}^\top\boldsymbol{p}\|_1\right).\end{aligned}$$

Since $\boldsymbol{S}$ is column-stochastic, hence $\|\boldsymbol{S}\boldsymbol{p}\|_1 = 1$. Furthermore, since $\boldsymbol{q}\mathbf{1}^\top\boldsymbol{p} = \boldsymbol{q}$, we also have $\|\boldsymbol{q}\mathbf{1}^\top\boldsymbol{p}\|_1 = 1$. Therefore,

$$\left\|\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha}\right\|_1 \le 2\,\|\,(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\,\|_1.$$

Now notice that $(\boldsymbol{I} - \alpha\boldsymbol{S})^\top\mathbf{1} = \boldsymbol{I}\mathbf{1} - \alpha\boldsymbol{S}^\top\mathbf{1} = \mathbf{1} - \alpha\mathbf{1} = (1-\alpha)\mathbf{1}$, and thus $(1-\alpha)^{-1}\mathbf{1} = (\boldsymbol{I} - \alpha\boldsymbol{S})^{-\top}\mathbf{1} \;\Rightarrow\; (1-\alpha)^{-1}\mathbf{1}^\top = \mathbf{1}^\top(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1} = \sum_i \left((\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\right)_{ij}$. Since $(1-\alpha)^{-1} > 0$ for $0 < \alpha < 1$, we have

$$\|\,(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\,\|_1 = \max_j \sum_i \left|\,\left((\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\right)_{ij}\,\right| = (1-\alpha)^{-1}.$$

Consequently,

$$\left\|\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\alpha}\right\|_1 \le \frac{2}{1-\alpha},$$

as desired.

The implication of this result is that the PageRank vector $\boldsymbol{p}(\alpha)$ is sensitive to variations in α. If α is very close to (meaning slightly less than) unity, then although the accuracy of ranking increases, it becomes more sensitive to the particular numerical value of α, that is, small changes in α would then produce large changes in $\boldsymbol{p}(\alpha)$.