AM205: Take-home midterm exam

This exam was posted at 5 PM on November 12th. Answers are due at 5 PM on November 14th. Solutions should be uploaded to Canvas.

For queries, contact the teaching staff using a private message on Piazza. Any clarifications will be posted on Piazza.

The exam is open book—any class notes, books, or online resources can be used. The exam must be completed by yourself and no collaboration with classmates or others is allowed. The exam will be graded out of forty points. Point values for each question are given in square brackets.

1. **Solving an ODE with a termination condition [15].** Suppose we wish to solve an ODE for a scalar function $y(t)$ that satisfies $y' = f(t, y)$. We have an initial condition $y(0) = 1$, and we wish to solve for y up until the time t_s such that $y(t_s) = 0$. Write a program to do this using the following steps, writing the routines yourself without employing any library functions:

   - Implement the classical fourth-order Runge–Kutta method. Using a step size of h, solve for $y_k = y(t_k)$, where $t_k = hk$ for $k = 0, 1, 2, \ldots$.
   - Terminate the Runge–Kutta method immediately when the condition $y_k < 0$ is first satisfied.
   - Construct an approximating cubic $y_c(t)$ on the interval $[t_{k-1}, t_k]$. The cubic is uniquely determined by the four conditions $y_c(t_{k-1}) = y_{k-1}$, $y_c(t_k) = y_k$, $y_c'(t_{k-1}) = f(t_{k-1}, y_{k-1})$, $y_c'(t_k) = f(t_k, y_k)$.
   - Use a bisection search to estimate t_s as the root of $y_c(t)$ in the interval $[t_{k-1}, t_k)$. Terminate the bisection when the interval that the root is contained within is smaller than 10^{-15}.

   To test your program, consider the differential equation

   $$y' = -\sqrt{y+1}. \tag{1}$$

   First, solve the equation by hand and calculate t_s analytically. Next, try your numerical procedure on this equation, using at least ten values of h from 0.3 to 0.003. Make a log–log plot of the absolute error of t_s as a function of h. Calculate the order of convergence and carefully explain why this is the case.

2. **A numerical scheme for the transport equation [15].**

   (a) Consider the numerical first derivative

   $$f_{\text{diff}}(x; \Delta x, \alpha) = \frac{(\alpha - 1) f(x - \Delta x) + (1 - 2\alpha) f(x) + \alpha f(x + \Delta x)}{\Delta x}$$

   of a smooth function $f(x)$, where Δx is a step size and α is a parameter. Use Taylor series to show that f_{diff} is first-order accurate for any value of α. Determine the specific value of α for which f_{diff} is second-order accurate.

   (b) Consider solving the transport equation $u_t + c u_x = 0$ for a function $u(t, x)$, using the numerical scheme

   $$\frac{u_j^{n+1} - u_j^n}{\Delta t} + c \frac{(\alpha - 1) u_{j-1}^n + (1 - 2\alpha) u_j^n + \alpha u_{j+1}^n}{\Delta x} = 0 \tag{2}$$

where $u_j^n = u(n\Delta t, j\Delta x)$. Define the CFL number as $\nu = c\Delta t/\Delta x$, and consider the case when $\nu = 0.1$. Perform a stability analysis by searching for solutions of the form $u_j^n = (\lambda(\omega, \alpha))^n e^{i\omega j}$ where $\lambda(\omega, \alpha)$ is an amplification factor. Define $Q = \cos\omega - 1$, and show that $|\lambda(\omega, \alpha)|^2$ can be written as a quadratic in Q. By considering the quadratic over the range $Q \in [-2, 0]$, show that the method is stable if and only if $\alpha_{\mathrm{lo}} \le \alpha \le \alpha_{\mathrm{hi}}$. Calculate α_{lo} and α_{hi}.

(c) Write a program to implement the scheme in Eq. 2 on the periodic interval with $x \in [0, 1)$. Use the initial condition $u(0, x) = (2 + \cos(2\pi x))^{-1}$. Use the parameters $h = \Delta t = 0.02$ and $c = 0.1$, and define the grid positions at $x_j = hj$ for $j = 0, \ldots 49$. Integrate the system for 500 timesteps from $t = 0$ to $t = 10$. Due to the periodicity, we have $u(0, x) = u(10, x)$. For the five values of α in $\{\alpha_{\mathrm{lo}}, \alpha_{\mathrm{hi}}, -\frac{1}{2}, 0, \frac{1}{2}\}$, calculate the measure of absolute error,

$$E = \sqrt{\sum_{j=0}^{49} (u_j^{500} - u_j^0)^2}. \tag{3}$$

Do stable values of α always give a lower E than unstable values?

3. **Fitting a matrix transformation [10].** Consider an experiment involving a soft square section of a gel, shown in Fig. 1. The gel is stretched and displaced during the experiment, so that its final position $\mathbf{r} = (r, s)$ is given in terms of the initial position $\mathbf{x} = (x, y)$ by

$$\mathbf{r} = F\mathbf{x} + \mathbf{c} \tag{4}$$

where F is a 2×2 matrix and $\mathbf{c} = (c, d)$ is a vector. We aim to determine F and $\mathbf{c}$. There are a number of fluorescent beads embedded in the gel that move with it. The initial positions $\mathbf{x}_i$ and final positions $\mathbf{r}_i$ are measured for $i = 1, \ldots, n$, and are provided in a text file `beads.txt`. Both $\mathbf{x}_i$ and $\mathbf{r}_i$ may have measurement error. From tracking how the beads move, F and $\mathbf{c}$ can be estimated.

(a) By formulating it as a linear least squares problem, or otherwise, find $F_{\mathrm{a}} \in \mathbb{R}^{2\times 2}$ and $\mathbf{c}_{\mathrm{a}} \in \mathbb{R}^2$ that minimize the residual

$$r(F_{\mathrm{a}}, \mathbf{c}_{\mathrm{a}}) = \sum_{i=1}^{n} ||F_{\mathrm{a}}\mathbf{x}_i + \mathbf{c}_{\mathrm{a}} - \mathbf{r}_i||_2^2. \tag{5}$$

(b) Use a similar approach to find $F_{\mathrm{b}} \in \mathbb{R}^{2\times 2}$ and $\mathbf{c}_{\mathrm{b}} \in \mathbb{R}^2$ that minimize the residual

$$r(F_{\mathrm{b}}, \mathbf{c}_{\mathrm{b}}) = \sum_{i=1}^{n} ||F_{\mathrm{b}}\mathbf{r}_i + \mathbf{c}_{\mathrm{b}} - \mathbf{x}_i||_2^2. \tag{6}$$

Calculate $||F_{\mathrm{a}} - F_{\mathrm{b}}^{-1}||_F$ and show that it is non-zero.

(c) The solutions from parts (a) and (b) are somewhat unsatisfactory. Part (a) assumes that the initial measurements are perfect, and that the errors in the final measurements should be minimized. Part (b) does the opposite. We now consider an approach that finds F

taking into account the errors in both. To begin, calculate shifted points $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ and $\tilde{\mathbf{r}}_i = \mathbf{r}_i - \bar{\mathbf{r}}$ where $\bar{\mathbf{x}}$ and $\bar{\mathbf{r}}$ are the mean bead positions. Then construct the matrix

$$A = \begin{pmatrix} \tilde{x}_1 & \tilde{y}_1 & \tilde{r}_1 & \tilde{s}_1 \\ \tilde{x}_2 & \tilde{y}_2 & \tilde{r}_2 & \tilde{s}_2 \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_n & \tilde{y}_n & \tilde{r}_n & \tilde{s}_n \end{pmatrix}. \tag{7}$$

Calculate the singular value decomposition (SVD) of A, and use it to construct a rank two matrix A' such that $||A - A'||_F$ is minimized.

i. Calculate a matrix F_c that exactly solves the equation

$$\tilde{\mathbf{r}}'_i = F_\mathrm{c} \tilde{\mathbf{x}}'_i \tag{8}$$

for all i, where $\tilde{\mathbf{r}}'_i$ and $\tilde{\mathbf{x}}'_i$ are given by the corresponding entries of A'.

ii. Calculate

$$E(\tilde{\mathbf{x}}'_i, \tilde{\mathbf{r}}'_i) = \sum_{i=1}^{n} ||\tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}'_i||_2^2 + \sum_{i=1}^{n} ||\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}'_i||_2^2. \tag{9}$$

(Here, $E(\tilde{\mathbf{x}}'_i, \tilde{\mathbf{r}}'_i)$ is a function defined on the collection of all the $\tilde{\mathbf{x}}'_i$ and $\tilde{\mathbf{r}}'_i$.) Determine a second expression for E in terms of the singular values of A, and show that the two values are the same.

iii. Suppose that there is another matrix F_d that exactly satisfies the equation

$$\tilde{\mathbf{r}}''_i = F_\mathrm{d} \tilde{\mathbf{x}}''_i \tag{10}$$

for all i, for some $\tilde{\mathbf{r}}''_i$ and $\tilde{\mathbf{x}}''_i$. Show that $E(\tilde{\mathbf{x}}''_i, \tilde{\mathbf{r}}''_i) \geq E(\tilde{\mathbf{x}}'_i, \tilde{\mathbf{r}}'_i)$.

(d) In reality, the true F used to generate the data is

$$F = \begin{pmatrix} 1.5 & 1 \\ -1 & 2 \end{pmatrix}. \tag{11}$$

Calculate $||F - F_\mathrm{a}||_F$, $||F - F_\mathrm{b}^{-1}||_F$, and $||F - F_\mathrm{c}||_F$. Show that F_c is the closest to F with respect to this norm.

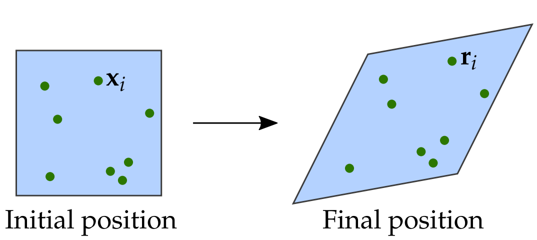

Figure 1: Diagram of the gel experiment considered in question 3. A gel is stretched and translated. Beads shown in green move with the gel, from initial points $\mathbf{x}_i$ to final points $\mathbf{r}_i$, and can be used to determine how the gel transforms.