CS412 Spring Semester 2011

Midterm #2 - Solutions

1. [30% = 6 questions × 5% each] MULTIPLE CHOICE SECTION. Circle or underline the correct answer (or answers). You do not need to provide a justification for your answer(s).

   (1) If the $n \times n$ matrix $\mathbf{A}$ is poorly conditioned (i.e. it has a very large condition number), then ...
   (Circle or underline the ONE most correct answer)

   (a) Solving $\mathbf{Ax} = \mathbf{b}$ accurately would be difficult with LU decomposition or Gauss elimination, but iterative methods (Jacobi, Gauss-Seidel) would not have a problem.

   (b) Solving $\mathbf{Ax} = \mathbf{b}$ accurately with iterative methods (Jacobi, Gauss-Seidel) would be difficult, but LU decomposition with pivoting would not have a problem.

   (c) Solving $\mathbf{Ax} = \mathbf{b}$ accurately will be challenging regardless of the method we use.

   (2) Consider the rectangular $m \times n$ matrix $\mathbf{A}$ (with $m > n$) and the vector $\mathbf{b} \in \mathbf{R}^m$. If $\mathbf{x}$ is the *least squares solution* to $\mathbf{Ax} \approx \mathbf{b}$, can we say that $\mathbf{x}$ is an actual solution to $\mathbf{Ax} = \mathbf{b}$?
   (Circle or underline the ONE most correct answer)

   (a) Yes, in fact $\mathbf{Ax} = \mathbf{b}$ has many solutions and the "least squares solution" is the one with the smallest 2-norm of the residual vector $\|\mathbf{r}\|_2$.

   (b) No, the system $\mathbf{Ax} = \mathbf{b}$ will generally not have a solution. What we call the "least squares solution" is the vector $\mathbf{x}$ with the smallest 2-norm of the error vector $\|\mathbf{x} - \mathbf{x}_{\text{exact}}\|_2$.

   (c) No, the system $\mathbf{Ax} = \mathbf{b}$ will generally not have a solution. What we call the "least squares solution" is the vector $\mathbf{x}$ with the smallest 2-norm of the residual vector $\|\mathbf{b} - \mathbf{Ax}\|_2$.

(3) Which of the following are good reasons for using an iterative method (e.g. Jacobi or Gauss-Seidel) instead of a direct method (e.g. Gauss Elimination or LU factorization) to solve the $n \times n$ system $\mathbf{Ax} = \mathbf{b}$?
(Circle or underline ALL correct answers)

(a) When an iterative method is convergent and the matrix $\mathbf{A}$ is relatively sparse, the computational cost of finding a good approximation of the simultion using an iterative approac could be significantly lower than using a direct method.

(b) Iterative methods work very well with poorly conditioned matrices, while direct methods face problems in this case.

(c) Iterative methods do not require pivoting when $\mathbf{A}$ is diagonally dominant or symmetric positive definite, while a direct method would require pivoting in this case.

(4) Imagine that we perform an evaluation of a certain composite integration rule, partitioning the integration interval $[a, b]$ into equal size subintervals, with length h. We observe that by doubling the number of data points, the error in the approximation of the integral is reduced by a factor of eight. Which of the following are true?
(Circle or underline ALL correct answers)

(a) The integration rule is third order accurate.

(b) The global integration error scales proportionately to h^4.

(c) The local integration error scales proportionately to h^4.

(5) When we try to solve an Initial Value Problem $y' = f(t, y),\ y(t_0) = y_0$, why is it desirable for the differential equation to have stable solutions?
(Circle or underline ALL correct answers)

(a) Because in this case numerical methods for approximating the solution will be stable as well.

(b) Because in this case it is possible for an propely designed numerical method to match the asymptotic behavior of the exact solution.

(c) Because if the solutions were unstable, any errors or inaccuracies incurred at any part of the solution process could be amplified without bound as $t \to \infty$.

(6) Which of the following statements about norms are true?
(Circle or underline ALL correct answers)

(a) $\|\mathbf{x}\|_\infty \geq \|\mathbf{x}\|_1$ for all $\mathbf{x} \in \mathbf{R}^n$ $(n \geq 2)$.

(b) $\|\mathbf{Ax}\| = \|\mathbf{A}\|\|\mathbf{x}\|$ for any matrix $\mathbf{A} \in \mathbf{R}^{n \times n}$ and vector $\mathbf{x} \in \mathbf{R}^n$.

(c) $\|\mathbf{A}^2\| \leq \|\mathbf{A}\|^2$ for any matrix $\mathbf{A} \in \mathbf{R}^{n \times n}$.

2. [20% = 4 questions × 5% each] SHORT ANSWER SECTION. Answer each of the following questions in no more than 1-2 sentences.

   (a) Why do we often prefer to use the composite Simpson's rule, instead of the composite Trapezoidal rule to approximate a definite integral?

   Answer: For a slight (almost negligible) increase in complexity, Simpson's rule offers significantly increased accuracy; namely it is 4th order accurate, while Trapezoidal rule is 2nd order accurate.

   (b) Write down three ordinary differential equations, one with asymptotically stable solutions, one with stable (but not asymptotically so) solutions, and one with unstable solutions.

   Answer: The canonical examples would be:

   - $y' = \lambda y,\ \lambda < 0$ (e.g. $y' = -5y$) is an equation with asymptotically stable solutions.
   - $y' = 0$ (or any equation of the form $y'(t) = f(t)$) is an equation with stable, but not asymptotically stable solutions.
   - $y' = \lambda y,\ \lambda > 0$ (e.g. $y' = 2y$) has unstable solutions.

   (c) When solving Initial Value Problems, why does an iteration of an implicit method often require more computational effort, than an iteration of an explicit method?

   Answer: While an explicit method isolates y_{n+1} on the left-hand-side by construction, an implicit method will generally require solving a (possibly nonlinear) equation to find this result, incurring additional cost per iteration.

   (d) List one of the conditions that would guarantee convergence of the Jacobi method for solving a linear system $\mathbf{Ax} = \mathbf{b}$.

   Answer:

   - $\mathbf{A}$ is diagonally dominant by rows, or
   - $\mathbf{A}$ is symmetric and positive (or negative) definite

3. [16%] Determine the order of accuracy for the following numerical integration rule:

$$\int_a^b f(x)dx \approx \frac{b-a}{2}\left[f\left(\frac{2a+b}{3}\right)+f\left(\frac{a+2b}{3}\right)\right]$$

Solution

We test this integration rule by checking if it integrates exactly monomials of the form $f(x) = x^d$.

- For $f(x) = 1$ we get

$$I_{\text{rule}} = \frac{b-a}{2}[1+1] = b-a \equiv \int_a^b 1dx$$

- For $f(x) = x$ we get

$$I_{\text{rule}} = \frac{b-a}{2}\left[\left(\frac{2a+b}{3}\right)+\left(\frac{a+2b}{3}\right)\right] = \frac{(b-a)(b+a)}{2} = \frac{b^2}{2}-\frac{a^2}{2} \equiv \int_a^b xdx$$

- For $f(x) = x^2$ we get

$$I_{\text{rule}} = \frac{b-a}{2}\left[\left(\frac{2a+b}{3}\right)^2+\left(\frac{a+2b}{3}\right)^2\right] = \frac{(b-a)(5b^2+8ba+5a^2)}{2}$$

$$\neq \frac{b^3}{3}-\frac{a^3}{3} \equiv \int_a^b x^2dx$$

We know that if the integration rule computes monomials up to order $d-1$, the rule is d-order accurate. In our case, the rule integrates exactly up to first order monomials ($f(x) = x$), thus the rule is second order accurate.

4. [14%] Consider the 5 points:

$$\begin{aligned}(x_1, y_1) &= (-3, -1)\\(x_2, y_2) &= (-2, 1)\\(x_3, y_3) &= (0, 2)\\(x_4, y_4) &= (1, 3)\\(x_5, y_5) &= (3, 2)\end{aligned}$$

(a) We want to determine a straight line $y = c_1 x + c_0$ that approximates these points as closely as possible, in the least squares sense. Write a least squares system $\mathbf{Ax} \approx \mathbf{b}$ which can be used to determine the coefficients c_1 and c_0.

(b) Solve this least squares system, using the method of normal equations.

Solution

We want the constants c_1 and c_0 to be such that the following equations are are closely approximated as possible, in the least squares sense:

$$\begin{aligned}c_1 x_1 + c_0 &\approx y_1\\c_1 x_2 + c_0 &\approx y_2\\c_1 x_3 + c_0 &\approx y_3\\c_1 x_4 + c_0 &\approx y_4\\c_1 x_5 + c_0 &\approx y_5\end{aligned}$$

These equations are written in matrix form as the least-squares system:

$$\begin{pmatrix} x_1 & 1\\ x_2 & 1\\ x_3 & 1\\ x_4 & 1\\ x_5 & 1\end{pmatrix}\begin{pmatrix} c_1\\ c_0\end{pmatrix} \approx \begin{pmatrix} y_1\\ y_2\\ y_3\\ y_4\\ y_5\end{pmatrix} \quad \text{or} \quad \underbrace{\begin{pmatrix} -3 & 1\\ -2 & 1\\ 0 & 1\\ 1 & 1\\ 3 & 1\end{pmatrix}}_{\mathbf{A}}\underbrace{\begin{pmatrix} c_1\\ c_0\end{pmatrix}}_{\mathbf{x}} \approx \underbrace{\begin{pmatrix} -1\\ 1\\ 2\\ 3\\ 2\end{pmatrix}}_{\mathbf{b}}$$

The least squares solution to this system $\mathbf{Ax} \approx \mathbf{b}$ is given by the *normal equations* system $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$, or:

$$\begin{pmatrix} -3 & -2 & 0 & 1 & 3\\ 1 & 1 & 1 & 1 & 1\end{pmatrix}\begin{pmatrix} -3 & 1\\ -2 & 1\\ 0 & 1\\ 1 & 1\\ 3 & 1\end{pmatrix}\begin{pmatrix} c_1\\ c_0\end{pmatrix} = \begin{pmatrix} -3 & -2 & 0 & 1 & 3\\ 1 & 1 & 1 & 1 & 1\end{pmatrix}\begin{pmatrix} -1\\ 1\\ 2\\ 3\\ 2\end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 23 & -1\\ -1 & 5\end{pmatrix}\begin{pmatrix} c_1\\ c_0\end{pmatrix} = \begin{pmatrix} 10\\ 7\end{pmatrix}$$

which yeilds the solution $c_1 = 0.5$, $c_0 = 1.5$.

5. [20%] Consider the following family of methods for solving Initial Value Problems of the form $y' = f(t, y),\ y(t_0) = y_0$:

$$y_{k+1} = y_k + \Delta t\left[(1-w)f(t_k, y_k) + wf(t_{k+1}, y_{k+1})\right] \tag{1}$$

In equation (1) the constant w can take any value in the interval $[0, 1]$; different values produce different methods. We can see, for example, that $w = 0$ corresponds to Forward Euler, $w = 0.5$ is Trapezoidal Rule and $w = 1$ produces Backward Euler.

(a) Show that for $0.5 \leq w \leq 1$, the method of equation (1) is unconditionally stable on the model equation $y' = \lambda y,\ \lambda < 0$.

(b) For $0 \leq w < 0.5$, determine the stability condition for the method of equation (1) when applied to the model equation $y' = \lambda y,\ \lambda < 0$.

Hint: Remember that stability of a method on this model equation is equivalent to showing that $y_k \to 0$ as $k \to \infty$.

Solution

For the model equation, we have $f(t, y) = \lambda y$. We substitute this into (1) to obtain:

$$y_{k+1} = y_k + \Delta t\left[(1-w)\lambda y_k + w\lambda y_{k+1}\right] \Rightarrow$$
$$\Rightarrow \left[1 - w\lambda\Delta t\right] y_{k+1} = \left[1 + (1-w)\lambda\Delta t\right] y_k \Rightarrow$$
$$y_{k+1} = \frac{1 + (1-w)\lambda\Delta t}{1 - w\lambda\Delta t} y_k \Rightarrow$$

From this equation, in order to guarantee that $y_k \to 0$ as $k \to \infty$, we need to require that

$$\left|\frac{1 + (1-w)\lambda\Delta t}{1 - w\lambda\Delta t}\right| < 1 \Rightarrow$$
$$\Rightarrow |1 + (1-w)\lambda\Delta t| < |1 - w\lambda\Delta t|$$

Since $\lambda < 0$ we have $1 - w\lambda\Delta t > 0$, thus the last inequality is equivalently written as

$$|1 + (1-w)\lambda\Delta t| < 1 - w\lambda\Delta t \Rightarrow$$
$$\Rightarrow -1 + w\lambda\Delta t < 1 + (1-w)\lambda\Delta t < 1 - w\lambda\Delta t \Rightarrow$$
$$\Rightarrow -2 + w\lambda\Delta t < (1-w)\lambda\Delta t < -w\lambda\Delta t \Rightarrow$$
$$\Rightarrow -2 - w|\lambda|\Delta t < -(1-w)|\lambda|\Delta t < w|\lambda|\Delta t$$

The second part of this inequality is always true when $w \in [0, 1]$. The first part of the inequality is equivalently written as

$$(1 - 2w)|\lambda|\Delta t < 2$$

When $w \geq 0.5$ this inequality always holds, since the left-hand side is a negative (or zero) quantity; thus for $w \geq 0.5$ the method is unconditionally stable.

When $w \in [0, 0.5)$, we divide both sides with the positive quantity $1 - 2w$, to obtain the final stability condition:

$$\Delta t < \frac{2}{(1 - 2w)|\lambda|}$$