Rutgers University
The State University Of New Jersey
College of Engineering
Department of Electrical and Computer Engineering

332:501   Systems Analysis   Fall 2002
                                    MIDTERM EXAMINATION

You have until 8pm to answer the following three (3) questions with point values as shown.

You are allowed two sides of handwritten notes on an 8.5×11 sheet of paper. Think through each

problem BEFORE you begin to write and don't get stuck on one problem. Move on if you are

stumped. YOU MUST SHOW ALL WORK. ANSWERS GIVEN WITHOUT WORK RECEIVE NO

CREDIT.

GOOD LUCK!

1. *(30 points)* **Solving Differential Equations:**

    (a) *(10 points)* You are given the following differential equation:

    $$\frac{d^4 x}{dt^4} + 4\frac{d^3 x}{dt^3} + 6\frac{d^2 x}{dt^2} + 4\frac{dx}{dt} + x = 0$$

    with initial conditions at $t = 0$ of $\frac{d^3 x}{dt^3} = 6$, $\frac{d^2 x}{dt^2} = 2$, $\frac{dx}{dt} = 1$, and $x(0) = 1$. Please provide the first four terms of the series solution for $x(t)$, $t \geq 0$.
    HINT: THINK FIRST
    SOLUTION: *At $t = 0$, the coefficient of t^n determines the value of the nth derivative. So the first four terms of the series must be $1 + t + t^2 + t^3$.*

    (b) *(10 points)* Please solve the previous differential equation in closed form for $x(t)$, $t \geq 0$.
    SOLUTION: *This is a linear constant coefficient differential equation so that we can use the laplace transform – which yields $(s + 1)^4 = 0$. So the homogenous solution is*

    $$x(t) = ae^{-t} + bte^{-t} + ct^2 e^{-t} + dt^3 e^{-t} = (a + bt + ct^2 + dt^3)e^{-t}$$

    The $x(0) = 1$ condition requires that $a = 1$. The first derivative at $t = 0$ is

    $$\left[(b - a) + (2c - b)t + (3d - c)t^2 - dt^3\right] e^{-t}$$

    which at $t = 0$ becomes $b - a = b - 1 = 1$ so that $b = 2$.
    The second derivative at $t = 0$ is

    $$\left[2c - 2b + a + (6d - 4c + b)t + (c - 6d)t^2 + dt^3\right] e^{-t}$$

    which at $t = 0$ is $2c - 2b + a = 2$ so that so that $c = 2.5$. And finally, the third derivative is

    $$\left[(6d - 6c + 3b - a) + (-18d + 6c - b)t + (9d - c)t^2 - dt^3\right] e^{-t}$$

    which at $t = 0$ is $6d - 6c + 3b - a = 6$ or $d = 5/3$.

(c) *(10 points)* Let

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

where θ is some constant. Find the state transition matrix for the system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ for $t \geq 0$.

SOLUTION: *Perhaps the easiest thing to do is to use residues on $(s\mathbf{I} - \mathbf{A})^{-1}$ though the matrix could also be easily diagonalized.*

$$(s\mathbf{I}-\mathbf{A})^{-1} = \begin{bmatrix} s - \cos\theta & \sin\theta \\ -\sin\theta & s - \cos\theta \end{bmatrix}^{-1} = \frac{1}{s^2 - 2\cos\theta + 1}\begin{bmatrix} s - \cos\theta & -\sin\theta \\ \sin\theta & s - \cos\theta \end{bmatrix}$$

In preparation for residues we can rewrite as

$$(s\mathbf{I} - \mathbf{A})^{-1} = \frac{1}{(s - \cos\theta + j\sin\theta)(s - \cos\theta - j\sin\theta)}\begin{bmatrix} s - \cos\theta & -\sin\theta \\ \sin\theta & s - \cos\theta \end{bmatrix}$$

We then do a partial fractions expansion on

$$\frac{s - \cos\theta}{(s - \cos\theta - j\sin\theta)(s - \cos\theta + j\sin\theta)} = \frac{0.5}{s - \cos\theta - j\sin\theta} + \frac{0.5}{s - \cos\theta + j\sin\theta}$$

to which applying the residue theorem produces

$$\frac{1}{2}e^{t\cos\theta}e^{jt\sin\theta} + \frac{1}{2}e^{t\cos\theta}e^{-jt\sin\theta} = e^{t\cos\theta}\cos(t\sin\theta)$$

and also on

$$\frac{\sin\theta}{(s - \cos\theta - j\sin\theta)(s - \cos\theta + j\sin\theta)} = \frac{0.5j}{s - \cos\theta + j\sin\theta} - \frac{0.5j}{s - \cos\theta - j\sin\theta}$$

to obtain

$$\frac{j}{2}e^{t\cos\theta}e^{-jt\sin\theta} - \frac{j}{2}e^{t\cos\theta}e^{+jt\sin\theta} = e^{t\cos\theta}\sin(t\sin\theta)$$

so that

$$e^{\mathbf{A}t} = e^{t\cos\theta}\begin{bmatrix} \cos(t\sin\theta) & -\sin(t\sin\theta) \\ \sin(t\sin\theta) & \cos(t\sin\theta) \end{bmatrix}$$

2. *(20 points)* **Rutgera and Computational Complexity:** Rutgera Univera, the world famous Rutgers University graduate student has been asked to numerically solve the following initial value problem

$$\mathbf{x}(n + 1) = \mathbf{A}\mathbf{x}(n)$$

with $\mathbf{x}(0) = \mathbf{x}_0$. She immediately sees that $\mathbf{x}(k) = \mathbf{A}^k\mathbf{x}_0$ and equally immediately sees that calcuating $\mathbf{A}^k$ will require on the order of kN^3 multiply operations (where N is the dimension of the matrix $\mathbf{A}$). If possible, help Rutgera find a way to reduce the computational complexity of finding $\mathbf{A}^k$ and explicitly state the number of operations necessary for your method.

SOLUTION: *We first state the characteristic equation as*

$$\psi(s) = \sum_{\ell=0}^{N} a_\ell s^\ell$$

2

where we usually assume $a_N = 1$.

$$\mathbf{A}^N = -\sum_{\ell=0}^{N-1} a_\ell \mathbf{A}^\ell$$

which implies that any power of $\mathbf{A}$ can be written in terms of lower powers of $\mathbf{A}$. So we precalculate these matrices and store them. Then for any power required, we compute the requisite polynomial

$$\mathbf{A}^k = \sum_{\ell=0}^{N-1} b_\ell \mathbf{A}^\ell$$

perform the requisite $O(N^3)$ multiplies (a constant times a matrix is N^2 multiplies and there are N matrices in the sum). So instead of kN^3 we have a constant N^3 complexity (excluding the complexity of calculating the coefficients b_ℓ.

Now, if $\mathbf{A}$ is diagonalizable as well, then the complexity is even more reduced. That is

$$\mathbf{A}^k = \mathbf{T}\Lambda^k\mathbf{T}^{-1}$$

which implies (using dumb algorithms to compute the powers of a constant) kN multiplies (for the diagonal terms of Λ) and then another N^2 multiplies to compute the product $\mathbf{T}\Lambda^k$ and finally N^3 operations to compute the product $\mathbf{T}\Lambda^k\mathbf{T}^{-1}$. And only three matrices need be stored after the initial computation of the diagonal form (on the order of N^3 operations as well).

If one uses smarter algorithms to compute powers, then for very large k we have on the order of $\log k$ multiplies to compute each λ_i^k since we can compute λ^{128}, for instance by computing in sequence λ^2 and λ^4, λ^8 etc.

3. *(50 points)* **Convexity and Mappings:** Convex functions on convex sets have unique minima. A marble rolled down a convex surface (with frictional drag) will eventually settle at this minimum. We will find later that in describing systems we will often try to figure out the properties of the surface traced out by system variable trajectories. And we will sometimes use contraction mapping theory to examine extensive properties like stability.

Here we explore the relationship (or lack thereof) between convexity and contraction mappings.

   (a) *(5 points)* Carefully write down the definition of a convex set.
   SOLUTION: *A convex set $\mathcal{C}$ has elements $x \in \mathcal{C}$ which obey $\lambda x_1 + (1 - \lambda)x_2 \in \mathcal{C}$ for $\lambda \in [0, 1]$.*

   (b) *(5 points)* Carefully write down the definition of a convex function.
   SOLUTION: *A convex function obeys*

$$f(\lambda x_1 + (1 - \lambda)x_2) \le \lambda f(x_1) + (1 - \lambda)f(x_2)$$

   where $\lambda \in [0, 1]$. and the x_i form a convex set in the domain of the function. Strict convexity implies that the inquality holds with equality iff $\lambda = 0, 1$

(c) *(10 points)* Suppose we define a set $\mathcal{S}$ and then an iteration

$$x_{n+1} = H(x_n) = x_n - \Delta \frac{df(x)}{dx}\Big|_{x=x_n}$$

where we assume $H(x) \in \mathcal{S}$ $\forall x \in \mathcal{S}$. Let $\frac{df(x)}{dx}$ be a bounded function (in the formal sense) with bounding constant K and then assume $\Delta < 1/K$. Finally, suppose $f()$ is strictly convex over the set $\mathcal{S}$. Is the mapping $H()$ a contraction mapping in general? Why/why not?

HINT: Use $\rho(x,y) = |x - y|$ as your measure.

SOLUTION: *Back to basics*

$$|H(x_1) - H(x_2)| = \left| x_1 - x_2 - \Delta\left(\frac{df}{dx}\Big|_{x=x_1} - \frac{df}{dx}\Big|_{x=x_2}\right)\right|$$

Now, since $f()$ is strictly convex, $\frac{df}{dx}$ must be a strictly increasing function which implies that if $x_1 > x_2$ then $\frac{df}{dx}\big|_{x=x_1} > \frac{df}{dx}\big|_{x=x_2}$. So since $\frac{df}{dx}$ is bounded and strictly increasing, we have assuming (with no loss of generality) that $x_1 > x_2$

$$0 < \epsilon \le \Delta\left(\frac{df}{dx}\Big|_{x=x_1} - \frac{df}{dx}\Big|_{x=x_2}\right) < x_1 - x_2$$

which implies

$$\left| x_1 - x_2 - \Delta\left(\frac{df}{dx}\Big|_{x=x_1} - \frac{df}{dx}\Big|_{x=x_2}\right)\right| \le |x_1 - x_2 - \epsilon| < |x_1 - x_2|$$

so the mapping is a contraction.

(d) *(10 points)* Now turn the problem around. Suppose the mapping $H(x)$ **IS** a contraction. Does that guarantee $f(x)$ convex? Why/why not? You may assume everything from the previous part, except the assumption of convexity of $f()$.

SOLUTION: *If the mapping is a contraction we must have for $x_1 - x_2 > 0$*

$$0 < \Delta\left(\frac{df}{dx}\Big|_{x=x_1} - \frac{df}{dx}\Big|_{x=x_2}\right) < x_1 - x_2$$

which implies that $\frac{df}{dx}$ is a strictly increasing function which automatically implies convexity.

(e) *(10 points)* Now try to extend (and contrast) the results for continuous vector mappings. Let

$$\dot{\mathbf{x}} = -\nabla f(\mathbf{x})$$

where $f(\mathbf{x})$ is a continuous, strictly convex scalar function with a vector argument and $\nabla f(\mathbf{x})$, the gradient of $f()$ is (formally) bounded in some metric space (you may assume $\rho(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$ if you like). Prove that for all possible initial conditions $\mathbf{x}(0)$ we must have

$$\lim_{t \to \infty} x(t) = \mathbf{x}^*$$

where $\mathbf{x}^*$ is the unique point such that

$$\nabla f(\mathbf{x}^*) = 0$$

4

HINT: I've not worked out the solution to this one – and almost invariably when I've not worked the solution out ahead of time it's either demonically hard or trivial.

SOLUTION: *First, this is not an iterated mapping problem (i.e., it's a continuous time problem whereas the prior parts were a discrete time mapping). Therefore, our contraction mapping theory cannot be directly applied without modification.*

So, the first thing to do is assume that the fixed point (that point where $\nabla f() = 0$) is in the domain set. Otherwise, the problem is ill-formed and we have to define boundaries on the domain. Then we define $\mathbf{G}(\mathbf{x}) = \nabla f(\mathbf{x})$ for notational simplicity.

Now, for global convergence, what we'd REALLY like is for $|\dot{\mathbf{x}}|$ to be monotonically DEcreasing no matter what the current value of $\mathbf{x}$ is. So we form

$$|\dot{\mathbf{x}}|^2 = \mathbf{G}(\mathbf{x})^\top \mathbf{G}(\mathbf{x})$$

and now we look at the first time derivative.

$$\frac{d\mathbf{G}^\top(\mathbf{x})\mathbf{G}(\mathbf{x})}{dt} = 2\dot{\mathbf{G}}^\top(\mathbf{x})\mathbf{G}(\mathbf{x})$$

We have via the chain rule that

$$\dot{\mathbf{G}}\mathbf{x} = \begin{bmatrix} \sum_{i=1}^N \frac{\partial^2 f}{\partial x_1 x_i}\dot{x}_i \\ \vdots \\ \sum_{i=1}^N \frac{\partial^2 f}{x_N x_i}\dot{x}_i \end{bmatrix} = \mathbf{H}_f\dot{\mathbf{x}}$$

where $\mathbf{H}$ is the Hessian of $f()$. This allows us to rewrite

$$2\dot{\mathbf{G}}^\top(\mathbf{x})\mathbf{G}(\mathbf{x}) = 2\dot{\mathbf{x}}^\top \mathbf{H}_f \mathbf{G}(\mathbf{x}) = -2\dot{\mathbf{x}}^\top \mathbf{H}_f \dot{\mathbf{x}}$$

since $\dot{\mathbf{x}} = -\mathbf{G}(\mathbf{x})$.

Since $\mathbf{H}_f$ is the Hessian of a convex function, it is positive definite which means that

$$-2\dot{\mathbf{x}}^\top \mathbf{H}_f \dot{\mathbf{x}} \le 0$$

with equality iff $\mathbf{x} = \mathbf{x}^*$.

Which completes the proof since we've shown that the magnitude of $\dot{\mathbf{x}}$ is decreasing and its magnitude is bounded from below by zero.

NOTE: It took me 15 minutes to type the solution in, solving and formatting on the fly (which means about 5 minutes to write it down by hand – which means under optimal conditions it might have taken you about 20 minutes to write down). So I guess this one was neither fiendishly hard nor trivial! This is your first peek at Lyapunov functions, by the way which will be important for stability.

(f) *(10 points)* Now suppose you are told that for all $\mathbf{x}(0)$

$$\lim_{t \to \infty} x(t) = \mathbf{x}^*$$

where $\mathbf{x}^*$ is unique. Does that guarantee $f()$ convex?

SOLUTION: *Consider the function $1 - e^{-x^2}$. A marble started from anywhere will eventually roll down toward $x = 0$ where the gradient is zero, but the function is clearly not convex. The problem is that the analogy is not exact. Here we require only convergence and not a contraction mapping as in the iterated sequence. If we'd required only convergence for the iterated sequence as well, then nonconvex functions would have been possible there too.*