Homework 1

January 18, 2025

Please submit your HW on Canvas; include a PDF printout of any code and results, clearly labeled, e.g. from a Jupyter notebook. For coding problems, we recommend using Julia, but you can use other languages if you wish. It is due Friday January 24th by 11:59pm EST.

Problem 1 (10 points)

Start reading the draft course notes (linked from https://github.com/mitmath/matrixcalc/). Find a place that you found confusing, and write a paragraph explaining the source of your confusion and (ideally) suggesting a possible improvement.

(Any other corrections/comments are welcome, too.)

Problem 2 (10 points)

Later in the class we will look analytically at derivatives of eigenproblems. Here you will begin to look at them empirically, with numerical experiments, using finite differences $\delta f = f(x + \delta x) - f(x) \approx f'(x)[\delta x]$.

Let $f(A) = \lambda$ be a function that maps a real-symmetric matrix A to one of its eigenvalues λ (e.g. the smallest one), satisfying $Av = \lambda v$ for some eigenvector v normalized to $\|v\| = 1$, and suppose that the eigenvalue is *not* a multiple eigenvalue. Demonstrate through numerical evidence on random 4×4 matrices that $\nabla\lambda = vv^T$.

(This result is known as the "Hellmann–Feynman theorem" in physics.)

Problem 3 (4+4+4+4 points)

Find the derivatives f' of the following functions. If f maps column vectors or matrices to scalars, give ∇f (so that $f'(x)[dx] = \langle \nabla f, dx \rangle$ in the usual inner product). If f maps column vectors to column vectors, give the Jacobian matrix. Otherwise, simply write down f' as a linear operation.

1. $f(x) = g.(x)$, denoting (in Julia notation) element-wise application of some scalar function g (in terms of its scalar derivative g'), for $x \in \mathbb{R}^m$.
2. $f(A) = (A^T A)^{-1}$ where A is an $m \times n$ matrix (with $m \geq n$ so that $A^T A$ is invertible for most A).
3. $f(x) = (I + xx^T)^{-1} x$ for $x \in \mathbb{R}^n$.
4. $f(A) = \operatorname{trace}(A^3)$ where A is an $m \times m$ matrix.

Problem 4 (5+5 points)

Use the "linear operator" definition of Kronecker products, where $A \otimes B$ is interpreted as a linear operator on matrices given by $(A \otimes B)[C] = BCA^T$, to show

1. Associativity:
$$A \otimes (B \otimes C) = (A \otimes B) \otimes C.$$
2. Mixed Products:
$$(A \otimes B)(C \otimes D) = (AC \otimes BD).$$

Problem 5 (4+4+4 points)

For this problem, recall that Newton's method finds a root $f(x) = 0$ of a function $f(x)$ by iteratively improving a guess x to $x \to x - f'(x)^{-1} f(x)$, assuming the derivative (or Jacobian) f' is square and invertible. It converges extremely rapidly if you start with a good enough initial guess.

If A is an $n \times n$ matrix, an ordinary eigenvalue λ solves $\det(A - \lambda I) = 0$. More generally, you can find the roots $\det M(\lambda) = 0$ where $M(\lambda)$ is some arbitrary matrix-valued function of λ: this is called a *nonlinear eigenvalue problem* and arises in lots of applications. Newton's method (and variants thereof) can be a very good way to solve nonlinear eigenproblems!

1. If M maps scalars $\lambda \in \mathbb{R}$ to $n \times n$ matrices $M(\lambda)$, explain why $M'(\lambda)$ (starting from the general definition of derivatives in class) is simply a matrix whose entries are the derivatives of each entry of M with respect to λ.
2. If $f(\lambda) = \det M(\lambda)$, find the Newton step $f'(\lambda)^{-1} f(\lambda)$ in terms of $M'(\lambda)$. (Simplify your answer: one term cancels nicely.)
3. Implement Newton's method to solve $\det M(\lambda) = 0$ for $M(\lambda) = A - \lambda I + \alpha\lambda \sin(\lambda) B$ with example 3×3 matrices `A = [-2 -1 -7; -1 6 5; -7 5 6]`, `B = [7 -1 8; -1 7 -1; 8 -1 3]`, and $\alpha = 0.01$. As your starting guess, use the largest eigenvalue of A (i.e. a solution for $\alpha = 0$), and find the resulting "nonlinear eigenvalue" of $M(\lambda)$ to at least 6 significant digits. (It should require very few Newton steps!)

Problem 6 (4+4+4+4 points)

Let $f(A)$ be a function that maps $m \times m$ matrices to $m \times m$ matrices. Recall that its derivative $f'(A)$ is a linear operator that maps any change δA in A to the corresponding change $\delta f = f(A + \delta A) - f(A) \approx f'(A)[\delta A]$, to first order in δA.

In this problem, you will study and prove a remarkable identity (Mathias, 1996): if $f(A)$ is sufficiently smooth,[1] then for *any* δA (not necessarily small!) the following formula holds *exactly*:

$$f\left(\underbrace{\begin{bmatrix} A & \delta A \\ & A \end{bmatrix}}_{M}\right) = \begin{bmatrix} f(A) & f'(A)[\delta A] \\ & f(A) \end{bmatrix}.$$

That is, one applies f to a $2m \times 2m$ "block upper-trianguar" matrix M (blank lower-left = zeros), and the desired derivative is in the upper-right $m \times m$ corner of the result $f(M)$. (Note: please do your *own* derivation here, don't just look it up.)

1. Check this identity numerically against a finite-difference approximation $f(A+\delta A) - f(A)$, which should match the exact $f'(A)[\delta A]$ to a few digits, for $f(A) = \exp(A)$ (the matrix exponential e^A, computed by `exp(A)` in Julia, or `expm` in Scipy or Matlab), for a random 3×3 matrix `A = randn(3,3)` and a random small perturbation `dA = randn(3,3) * 1e-8`. Note that you can make the block matrix above in Julia by `using LinearAlgebra` followed by `M = [A dA; 0I A]`, and you can extract an upper-right corner by (*e.g.*) `M[1:3,4:6]`.

   It is also worth verifying (by a finite-difference check) that the derivative of the matrix exponential is *not* simply multiplication by e^A (on either the left or right or both): $(e^A)'[dA] \neq e^A\, dA$ or $dA\, e^A$ or $e^{A/2}\, dA\, e^{A/2}$, unlike the scalar case.
2. Prove the identity by explicit computation for the cases: $f(A) = I$, $f(A) = A$, $f(A) = A^2$, and $f(A) = A^3$. (Two of these are trivial! This is "bargain-basement induction": do a few small examples and see the pattern.)
3. Prove the identity for $f(A) = A^n$ for any $n \geq 0$ by induction: assume it works for A^{n-1} and show using the product rule that it therefore must work for A^n. (You already proved the trivial $n = 0$ base case in the previous part.)

   Remark: Once it works for any A^n, it immediately follows that it works for any $f(A)$ described by a Taylor series, such as $\exp(A) = I + A + A^2/2 + A^3/6 + \cdots + A^n/n! + \cdots$, since such a function is just a linear combination of A^n terms.
4. Prove the identity for $f(A) = A^{-1}$: since we know (from class) that $f'(A)[\delta A] = -A^{-1}\, \delta A\, A^{-1}$, plug this into the right-hand side of the formula above and show that it is the inverse of M (multiply by M and show you get I).[2]

[1] The result is easiest to show when $f(A)$ has a Taylor series (is "analytic"), and in fact you will do this below, but Higham (2008) shows that it remains true whenever f is $2m - 1$ times differentiable, or even just differentiable if A is "normal" ($AA^T = A^T A$).

[2] This is a special case of the famous "Schur complement" formula for the inverse of a 2×2 block matrix.