Homework 1

January 20, 2024

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 26th by 11:59pm EST.

Problem 1

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

A directional derivative of $f(x)$ in a direction v is sometimes described as the derivative $\frac{d}{d\alpha} f(x+\alpha v)\big|_{\alpha=0}$, where $\alpha \in \mathbb{R}$ is a scalar; that is, it is $g'(0)$ for $g(\alpha) = f(x+\alpha v)$. If $f(x)$ is a function from some input vector space $x \in X$ to some output vector space $f(x) \in Y$ with a derivative $f'(x)$ as defined in class, apply the chain rule to obtain this $g'(0)$ (for some $v \in X$) in terms of f'.

Problem 3

Find the derivatives f' of the following functions. If f maps column vectors to scalars, give ∇f (so that $f'(x)[dx] = (\nabla f)^T dx$ as in our definition of the gradient), and if f maps column vectors to column vectors gives the Jacobian matrix. Otherwise, simply write down f' as a linear operation.

1. $f(x) = \|x\| = \sqrt{x^T x}$ for $x \in \mathbb{R}^m$.
2. $f(x) = \frac{x^T(A+\|x\|^2 I)x}{x^T x}$ for $x \in \mathbb{R}^m$, A being a constant $m \times m$ matrix, and I being the $m \times m$ identity matrix.
3. $f(A) = A^{-2}$ where A is an $m \times m$ matrix.
4. $f(A) = (\operatorname{trace} A)^9$ where A is an $m \times m$ matrix.
5. $f(x) = A(x \,.\!*\, x)$ where A is an $m \times n$ matrix, $x \in \mathbb{R}^n$, and $.*$ denotes *elementwise* multiplication (also called a Hadamard product) in Julia/Matlab notation.

Problem 4

Suppose that $f(t) = A(t)$ is a function that maps scalars $t \in \mathbb{R}$ to $m \times n$ matrices $A(t)$. For example, $A(t) = \begin{pmatrix} \sin(t) & 0 & \cos(t) \\ t & t^2 & t^3 \end{pmatrix}$. Explain why $f'(t)[dt]$, following our general definition, must simply correspond to taking the ordinary single-variable calculus derivative of each element of $A(t)$ (the "elementwise" derivative) and multiplying it by the scalar dt. That is, $f'(t) = A'(t)$ is the elementwise derivative.

Problem 5

If you are not familiar with 2d convolution operations (or even if you are), watch (a little of the start of) the YouTube video https://www.youtube.com/watch?v=yb2tPt0QVPY that explains them.

1. The (linear) convolution of an $m \times n$ array ("matrix") with the 3×3 Sobel kernel $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$ results in a ______ by ______ array?

2. Explain why this map from arrays to arrays is a linear operation.
3. Let X be a 2024×2024 array, and Y be the result of convolving X with Sobel. Describe the matrix M that satisfies:

$$\mathrm{vec}(Y) = M\,\mathrm{vec}(X)\,.$$

   What is the size of M? Express M in terms of Kronecker products of much smaller matrices (hint: 2d convolution with this Sobel kernel is "separable" into 1d convolutions acting on the rows and columns of X, and you can express these as matrices multiplying X on the ______ and ______, respectively).
4. Convolutions like this are very common linear operations, and yet they are not normally implemented by constructing an explicit matrix then multiplying it by a vector (even for 1d convolutions, much less 2d), no matter what you may have learned in linear algebra classes. Why is that?

Problem 6

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:

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

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 $f(M)$.

1. Check this identity numerically in Julia against a finite-difference approximation 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 `A = randn(3,3)` and a random small perturbation `dA = randn(3,3) * 1e-8`; note that you can make the block matrix above 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]`.
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 is 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 $f(A) = A^{-1}$ by explicit computation: 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.

[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 diagonalizable.