# Homework 1 Solutions

January 26, 2025

**See also the accompanying Julia notebook for computational solutions.**

## 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.)

**Solution:**

Student-dependent, but full marks if clearly written and explained.

## 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 $\lambda$ (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 $4 \times 4$ matrices that $\nabla \lambda = vv^T$.

**Solution:**

See attached Julia notebook.

## 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 $\nabla f$ (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. $f(A) = \text{trace}(A^3)$ where $A$ is an $m \times m$ matrix.

**Solution:**

1. The Jacobian is just a diagonal matrix of the element-wise derivatives. $f'(x) = \begin{bmatrix} g'(x_1) & & & \\ & g'(x_2) & & \\ & & \ddots & \\ & & & g'(x_n) \end{bmatrix}$. (In Julia

notation , $f'(x)[dx] = g'.(x) .* dx$, or $f'(x) = \text{Diagonal}(g'.(x))$.)

2. Applying the matrix-inverse chain rule and the product rule, we have

$$df = -(A^T A)^{-1} d(A^T A)(A^T A)^{-1}$$
$$= \boxed{-f(A)(dA^T A + A^T dA)f(A) = f'(A)[dA]},$$

which clearly a linear operation on $dA$.

3. Again by the chain and product rules, we get:

$$df = d((I + xx^T)^{-1})x + (I + xx^T)^{-1}dx$$
$$= -(I + xx^T)^{-1}d(I + xx^T)(I + xx^T)^{-1}x + (I + xx^T)^{-1}dx$$
$$= (I + xx^T)^{-1}\left(-(dx\, x^T + x\, dx^T)f(x) + dx\right).$$

To get this in the form of a Jacobian matrix, i.e. $f'(x)dx$, we simply employ the identity $dx^T y = y^T dx$ to move $dx^T$ to the right, along with the fact that $x^T f(x)$ is a scalar so it commutes with the other operations, giving:

$$\boxed{f'(x) = (I + xx^T)^{-1}\left(-(x^T f(x))I - xf(x)^T + I\right) = (1 - x^T f(x))(I + xx^T)^{-1} - f(x)f(x)^T}.$$

4. Since the output is a scalar, we want $\nabla f$ such that $df = \operatorname{trace}(\nabla f^T dA)$. Since trace is linear, we can pass the differential through the trace, apply the product rule for $A^3$, and then apply the cyclic rule to push all of the $dA$ terms to the right:

$$df = \operatorname{trace}(d(A^3)) = \operatorname{trace}(dA\, A^2 + A\, dA\, A + dA\, A^2)$$
$$= \operatorname{trace}(3A^2\, dA),$$

which by inspection gives $\boxed{\nabla f = (3A^2)^T}$.

## 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:
$$(A \otimes B)(C \otimes D) = (AC \otimes BD).$$

**Solution:**

1. **Associativity**: In order to derive this from the "'operator" viewpoint, as discussed in class, we have to first extend the "linear operator" definition $Y = (A \otimes B \otimes C)[X]$ for the triple Kronecker product to make sense as a linear operator on "3d arrays" $X$ and $Y$. Alternatively, you can prove this problem using the "matrix" viewpoint.

   *Matrix viewpoint*:

   Applying the matrix definition $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$, we obtain

$$A \otimes (B \otimes C) = \begin{bmatrix} a_{11}\begin{bmatrix} b_{11}C & \cdots & b_{1n'}C \\ \vdots & \ddots & \vdots \\ b_{m'1}C & \cdots & b_{m'n'}C \end{bmatrix} & \cdots & a_{1n}\begin{bmatrix} b_{11}C & \cdots & b_{1n'}C \\ \vdots & \ddots & \vdots \\ b_{m'1}C & \cdots & b_{m'n'}C \end{bmatrix} \\ \vdots & \ddots & \vdots \\ a_{m1}\begin{bmatrix} b_{11}C & \cdots & b_{1n'}C \\ \vdots & \ddots & \vdots \\ b_{m'1}C & \cdots & b_{m'n'}C \end{bmatrix} & \cdots & a_{mn}\begin{bmatrix} b_{11}C & \cdots & b_{1n'}C \\ \vdots & \ddots & \vdots \\ b_{m'1}C & \cdots & b_{m'n'}C \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11}C & \cdots & a_{11}b_{1n'}C & & a_{1n}b_{11}C & \cdots & a_{1n}b_{1n'}C \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{11}b_{m'1}C & \cdots & a_{11}b_{m'n'}C & & a_{1n}b_{m'1}C & \cdots & a_{1n}b_{m'n'}C \\ & \vdots & & \ddots & & \vdots & \\ a_{m1}b_{11}C & \cdots & a_{m1}b_{1n'}C & & a_{mn}b_{11}C & \cdots & a_{mn}b_{1n'}C \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m'1}C & \cdots & a_{m1}b_{m'n'}C & & a_{mn}b_{m'1}C & \cdots & a_{mn}b_{m'n'}C \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} & \cdots & a_{11}b_{1n'} & & a_{1n}b_{11} & \cdots & a_{1n}b_{1n'} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{11}b_{m'1} & \cdots & a_{11}b_{m'n'} & & a_{1n}b_{m'1} & \cdots & a_{1n}b_{m'n'} \\ & \vdots & & \ddots & & \vdots & \\ a_{m1}b_{11} & \cdots & a_{m1}b_{1n'} & & a_{mn}b_{11} & \cdots & a_{mn}b_{1n'} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m'1} & \cdots & a_{m1}b_{m'n'} & & a_{mn}b_{m'1} & \cdots & a_{mn}b_{m'n'} \end{bmatrix} \otimes C$$

$$= (A \otimes B) \otimes C.$$

   Q.E.D.

*Operator viewpoint*:

As an operator, $A \otimes (B \otimes C)$ denotes contraction in dimensions 1 and 2 followed by contraction in dimension 3, while $(A \otimes B) \otimes C$ denotes contraction in dimension 1 followed by the contractions in dimensions 2 and 3.

Written out, for 3d arrays $X$ of the right size, associativity is the obvious identity that

$$\sum_{j_3} A_{i_3 j_3} \sum_{j_1, j_2} B_{i_2 j_2} C_{i_1 j_1} X_{j_1 j_2 j_3} = \sum_{j_2, j_3} A_{i_3 j_3} B_{i_2 j_2} \sum_{j_1} C_{i_1 j_1} X_{j_1 j_2 j_3},$$

in that both are equal to

$$Y_{i_1 i_2 i_3} = \sum_{j_1, j_2, j_3} A_{i_3 j_3} B_{i_2 j_2} C_{i_1 j_1} X_{j_1 j_2 j_3}.$$

Geometrically, in terms of a data brick, the left hand side may be viewed as the matrix Kronecker product in the planes in dimensions 1 and 2 followed by a contraction in dimension 3. The right hand side may be viewed as a contraction in dimension 1 followed by matrix kronecker products in the planes in dimensions 2 and 3.

For those who might enjoy how expressive Julia can be, try this out:

```julia
m₁=2; m₂=6; m₃=3
n₁=3; n₂=4; n₃=5
X = rand(n₁,n₂,n₃)
C = rand(m₁,n₁); B=rand(m₂,n₂); A=rand(m₃,n₃);

using LinearAlgebra
const ⊗ = kron # the matrix definition

A ⊗ (B ⊗ C) ≈ (A ⊗ B) ⊗ C # returns true

# the operator definition, acting on slices:
matvec(M,d)    = X -> mapslices(v->M*v,   X, dims=d) # matvec on 1d lines
kronprod(A,B,d) = X -> mapslices(Y->B*Y*A',X, dims=d) # kron on 2d planes

lhs =      ( matvec(A,3) ∘ kronprod(B,C,[1,2]))(X)  # A ⊗ (B ⊗ C)
rhs =      ( kronprod(A,B,[2,3]) ∘ matvec(C,1))(X)   # (A ⊗ B) ⊗ C

A ⊗ (B ⊗ C) * vec(X) ≈ vec(lhs) # returns true
(A ⊗ B) ⊗ C * vec(X) ≈ vec(rhs) # returns true

# the operator definition, this time with Einstein summation
using Einsum
Y = zeros(m₁,m₂,m₃)
@einsum Y[i₁,i₂,i₃] =  A[i₃,j₃] * B[i₂,j₂] * C[i₁,j₁] * X[j₁,j₂,j₃]

lhs ≈ Y ≈ rhs # returns true
```

2. **Mixed Products**:

Here, the linear operator definition on matrices, $(C \otimes D)[X] = DYC^T$, is all that is needed. The key fact is that the product of two matrices corresponds to the composition of the linear operators (i.e., associativity). What you are asked to prove then follows by elementary algebra: for any $X$ of the right size,

$$\begin{aligned}
((A \otimes B)(C \otimes D))[X] &= (A \otimes B)[(C \otimes D)[X]] \\
&= (A \otimes B)[DXC^T] \\
&= B(DXC^T)A^T \\
&= (BD)X(AC)^T = ((AC) \otimes (BD))[X].
\end{aligned}$$

## 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 $\lambda$ 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 $\lambda$: 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 $\lambda$.

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 $3 \times 3$ 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!)

**Solution:**

1. $dM = M(\lambda + d\lambda) - M(\lambda) = M'(\lambda)d\lambda$ is element-wise subtraction, so each element $dM_{i,j} = M_{i,j}(\lambda + d\lambda) - M_{i,j}(\lambda) = M'_{i,j}d\lambda$ is simply given in terms of the derivative of that element. Hence the entries of $M'$ are the derivatives of each element of $M$ with respect to $\lambda$.

2. Applying the rule for the derivative of the determinant from class, $df = \text{trace}[\det(M)M^{-1}dM] = \det(M)\,\text{trace}[M^{-1}M'(\lambda)]d\lambda$, so we have $f'(\lambda) = \det(M)\,\text{trace}[M(\lambda)^{-1}M'(\lambda)]$. This gives a Newton step:

$$f'(\lambda)^{-1}f(\lambda) = \boxed{\text{trace}[M(\lambda)^{-1}M'(\lambda)]^{-1}},$$

   where the $\det M$ factors cancel.

3. See attached Julia notebook. The correct answer to 73 decimal places is

$$\boxed{\lambda \approx 13.3245807095348859129767072645742889307342307370431253587838638594879456582},$$

   but you were only asked to get the first 4 decimal places right.

## Problem 6 (3+4+3+3 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 $\delta A$ in $A$ to the corresponding change $\delta f = f(A + \delta A) - f(A) \approx f'(A)[\delta A]$, to first order in $\delta A$.

In this problem, you will study and prove a remarkable identity (Mathias, 1996): if $f(A)$ is sufficiently smooth,[1] then for *any* $\delta A$ (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-triangular" 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 $3 \times 3$ `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.)

---

[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.

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 $I$.

**Solution:**

1. See attached Julia notebook.

2. These three cases are:

   (a) $f(A) = A^0 = I$: in this case $f(M) = I$ ($2m \times 2m$), and so the upper-right block is **zero**. This, of course, is the correct result $d(I) = 0$.

   (b) $f(A) = A$: in this case, $f(M) = M$, and the upper-right block is $\delta A$. Again, this is the correct result: $df = f'(A)[dA] = d(A) = dA$, so $f'(A)[\delta A] = \delta A$.

   (c) $f(A) = A^2$. In this case,

   $$f(M) = M^2 = \begin{pmatrix} A & \delta A \\ & A \end{pmatrix} \begin{pmatrix} A & \delta A \\ & A \end{pmatrix} = \begin{pmatrix} A^2 & A\,\delta A + \delta A\,A \\ & A^2 \end{pmatrix}$$

   by the usual "rows-times-columns" rule (which works for matrix *blocks* as well as for scalar elements). But then the upper-right block $A\,\delta A + \delta A\,A$ is precisely $f'(A)[\delta A]$ as derived in class by the product rule.

   (d) $f(A) = A^3$. Building off the previous part, we have

   $$f(M) = M^3 = M^2 M = \begin{pmatrix} A^2 & A^2\,\delta A + \delta A\,A \\ & A \end{pmatrix} \begin{pmatrix} A & \delta A \\ & A \end{pmatrix} = \begin{pmatrix} A^3 & A^2\,\delta A + A\,\delta A\,A + \delta A\,A^2 \\ & A^3 \end{pmatrix}$$

   again by the usual "rows-times-columns" rule. The upper-right block $\delta A + A\,\delta A\,A + \delta A\,A^2$ is again $f'(A)[\delta A]$ as derived in class, corresponding to $d(A^3) = dA + A\,dA\,A + dA\,A^2$

3. For an inductive proof, we assume (for $n > 0$) that the identity holds for $n - 1$, i.e. that:

   $$M^{n-1} = \begin{pmatrix} A^{n-1} & (A^{n-1})'[\delta A] \\ & A^{n-1} \end{pmatrix}.$$

   It then follows that

   $$M^n = M^{n-1}M = \begin{pmatrix} A^{n-1} & (A^{n-1})'[\delta A] \\ & A^{n-1} \end{pmatrix} \begin{pmatrix} A & \delta A \\ & A \end{pmatrix} = \begin{pmatrix} A^n & A^{n-1}\,\delta A + (A^{n-1})'[\delta A]\,A \\ & A^n \end{pmatrix},$$

   again by the usual "rows-times-columns" rule. But the upper-right block corresponds exactly to the product rule $d(A^n) = d(A^{n-1}A) = A^{n-1}dA + d(A^{n-1})A$, so it is indeed $(A^n)'[\delta A]$ as desired.

   As noted in the problem, the inductive base case $n = 0$ was already shown in the previous part, so our result must now hold for all $n \geq 0$.

4. For $f(A) = A^{-1}$, we know from class that $f'(A)[\delta A] = -A^{-1}\,\delta A\,A^{-1}$. We now want to show that

   $$M^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}\,\delta A\,A^{-1} \\ & A^{-1} \end{pmatrix},$$

   which we can establish by explicit multiplication with $M$ (on either the left or right):

   $$\begin{pmatrix} A^{-1} & -A^{-1}\,\delta A\,A^{-1} \\ & A^{-1} \end{pmatrix} \begin{pmatrix} A & \delta A \\ & A \end{pmatrix} = \begin{pmatrix} A^{-1}A & A^{-1}\,\delta A - A^{-1}\,\delta A\,A^{-1}A \\ & A^{-1}A \end{pmatrix} = I$$

   as desired: $f(M) = M^{-1}$ indeed must have the correct derivative $-A^{-1}\,\delta A\,A^{-1}$ in the upper-right block and $A^{-1}$ on the diagonal blocks.