MIT 18.06 Exam 2 Solutions, Fall 2022

Johnson

Problem 1 [(5+5)+10 points]:

These two parts are **answered independently:**

(a) Consider the 2d "plane" S spanned by

$$a_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \; a_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

(i) Give an **orthonormal basis** for S.

Solution: We just need to do Gram–Schmidt:

$$q_1 = \frac{a_1}{\cancelto{2}{\|a_1\|}} = \boxed{\frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}$$

and

$$q_2 = \frac{a_2 - q_1 \cancelto{1}{q_1^T a_2}}{\| \cdots \|} = \frac{\frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}}{\cancelto{1}{\| \cdots \|}} = \boxed{\frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}}.$$

(Although this is the most obvious approach, there are infinitely many other orthonormal bases we chould have chosen. For example, we could have done Gram–Schmidt in the opposite order, on a_2, a_1.)

(ii) Find the **closest point** in S to the (column vector) $y = [-2, 4, -6, 8]$.

Solution: This is just the orthogonal projection p of y onto S, which is easy to do using the orthonormal basis from (a):

$$p = q_1 \cancelto{2}{q_1^T y} + q_2 \cancelto{4}{q_2^T y} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \boxed{\begin{pmatrix} 3 \\ -1 \\ -1 \\ 3 \end{pmatrix}}.$$

Note that we could also have computed the projection matrix $P = QQ^T = q_1 q_1^T + q_2 q_2^T$ and then multiplied it by y, but this is *much* more work (matrices require more arithmetic than vectors)! Even *more* work would be using $A = \begin{pmatrix} a_1 & a_2 \end{pmatrix}$ and then using $A(A^T A)^{-1} A^T$, i..e. solving the normal equations $A^T A \hat{x} = A^T y$ and then finding $p = A\hat{x}$.

(b) Suppose that we have 100 measurements (p_k, v_k) of the volume v of a gas vs. its pressure p, and we want to fit it to a function of the form $v(p) = \frac{c_1}{p} + c_2$ for unknown constants c_1, c_2. Write down the 2×2 **system of equations** you would solve to find c_1, c_2 in order to minimize the sum of the squared errors $\sum_k [v(p_k) - v_k]^2$. You can write your answer (left- and right-hand sides) as products of matrices and/or vectors, as long as you specify what each term is (in terms of the unknowns c_1, c_2 and/or the data $p_1, \ldots, p_{100}$ and $v_1, \ldots, v_{100}$).

Solution: This is a least-square problem, so the answer is to solve the normal equations $\boxed{A^T A c = A^T b}$ for $c = \begin{pmatrix} c_1 & c_2 \end{pmatrix}^T$ where

$$\boxed{A = \begin{pmatrix} \frac{1}{p_1} & 1 \\ \frac{1}{p_2} & 1 \\ \vdots & \vdots \\ \frac{1}{p_{100}} & 1 \end{pmatrix}} \text{ and } \boxed{b = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{100} \end{pmatrix}}$$

so that Ac is the "model" $\begin{pmatrix} v(p_1) \\ v(p_2) \\ \vdots \\ v(p_{100}) \end{pmatrix}$ and b are the data we are fitting to, so that $\sum_k [v(p_k) - v_k]^2 = \|Ac - b\|^2$.

Problem 2 [4+4+4+4+4+4 points]:

These parts can be **answered independently:**

(a) The matrix $\frac{a_1 a_1^T}{a_1^T a_1} + \frac{a_2 a_2^T}{a_2^T a_2}$ is the projection matrix onto the span of $a_1, a_2 \in \mathbb{R}^m$ if a_1 and a_2 are **(circle all true answers)**: *independent, orthogonal, parallel, orthonormal, singular, length-1.*

Solution: $\boxed{\text{orthogonal}}$ or $\boxed{\text{orthonormal}}$. (They *must* be orthogonal for this to be a projection—that's the only way you can project one vector at a time via dot products. Their normalization is irrelevant because we are dividing each term by the length^2, but it's fine if they are normalized to length 1.)

Ideally, this problem should have specified explicitly that the **vectors** a_1, a_2 **are nonzero** (zero vectors are orthogonal to everything, including themselves), but this is implicit in the problem statement since the formula $\frac{a_1 a_1^T}{a_1^T a_1} + \frac{a_2 a_2^T}{a_2^T a_2}$ makes no sense for zero vectors ($\frac{0}{0}$?).

(b) If $\hat{x}$ is the least-square solution minimizing $\|Ax - b\|$ over x, then $A\hat{x} - b$ must lie in **which fundamental subspace** of A?

Solution: $C(A)^\perp = \boxed{N(A^T)}$, i.e. the **left nullspace** of A. $A\hat{x}$ is the projection onto $C(A)$, and the error $b - A\hat{x}$ is orthogonal to $C(A)$.

(c) A, B are 10×3 matrices, and $b \in \mathbb{R}^{10}$. If we want to find the vector $\hat{y} \in \mathbb{R}^3$ for which $A\hat{y} - b \in C(B)^\perp$, then $\hat{y}$ satisfies the 3×3 **system of equations** ________ (in terms of $A, B, b, \hat{y}$).

Solution: $C(B)^\perp = N(B^T)$, so we just need $B^T(A\hat{y} - b) = 0 \implies \boxed{B^T A\hat{y} = B^T b}$.

Note that this is very similar to how we derived the normal equations, by requiring that $A\hat{x} - b$ be orthogonal to $C(A)$; that is, you get the normal equations if you set $B = A$.

(d) A, B are matrices with $C(A) = C(B)$, and we have solved $A^T A\hat{x} = A^T b$ for $\hat{x}$ and $B^T B\hat{y} = B^T b$ for $\hat{y}$. **Circle statements (if any) that *must* be true:** $\hat{x} = \hat{y}$, $A\hat{x} = B\hat{y}$, and/or $\hat{x}^T b = \hat{y}^T b$.

Solution: $\boxed{A\hat{x} = B\hat{y}}$, since these are the orthogonal projections onto $C(A) = C(B)$; the column spaces are the same, so the projections are the same. (But the *coefficients* of the projection $\hat{x}$ in the A basis don't need to match the coefficients $\hat{y}$ in the B basis!)

(e) Q is a 5×3 matrix with orthonormal columns. Circle which **must** be true: $\|Qx\| = \|x\|$ for $x \in \mathbb{R}^3$, $\|Q^T y\| = \|y\|$ for $y \in \mathbb{R}^5$.

Solution: $\boxed{\|Qx\| = \|x\|}$, since $\|Qx\| = \sqrt{(Qx)^T(Qx)} = \sqrt{x^T \cancelto{I}{Q^T Q} x} = \|x\|$. In contrast, $\|Q^T y\| = \sqrt{(Q^T y)^T (Q^T y)} = \sqrt{y^T Q Q^T y}$, but $QQ^T \neq I$ since Q is not square—it is a 5×5 projection matrix onto the 3-dimensional subspace $C(Q)$.

(f) If A is a 3×3 matrix with $\det(A) = 3$, then $\det[A^T A^{-1}] + \det(2A) =$ ____.

Solution: Using the properties of determinants, we find:

$$\det[A^T A^{-1}] + \det(2A) = \underbrace{\det(A^T)}_{\det A = 3} \underbrace{\det(A^{-1})}_{(\det A)^{-1} = \frac{1}{3}} + \underbrace{\det(2A)}_{2^3 \det(A) = 24} = \boxed{25}.$$

Problem 3 [(3+3+3)+5 points]:

These two parts are **answered independently:**

(a) If A is a 10×3 matrix has an SVD $U\Sigma V^T$ with $\Sigma = \begin{pmatrix} 100 & & \\ & 10 & \\ & & 1 \end{pmatrix}$, then

(i) U is a ____ $\times$ ____ matrix, V is a ____ $\times$ ____ matrix, and A has rank ____.

Solution: U is a $\boxed{10 \times 3}$ matrix, V is a $\boxed{3 \times 3}$ matrix (this is the standard size of the "thin" SVD we covered in class, but these are also the only possible sizes that will give the correct 10×3 size for A!), and the rank is $\boxed{3}$ (the number of nonzero singular values $\sigma_1 = 100, \sigma_2 = 10, \sigma_3 = 1$.

(ii) The projection matrix onto $C(A)$ is ________ and the projection onto $C(A^T)$ is ________ (**simplest answers** in terms of U, Σ, V, I).

Solution: U is an orthonormal basis for $C(A)$, so the projection is $\boxed{UU^T}$. V is an orthonormal basis for $C(A^T)$, so the projection is VV^T, but to get full credit you should notice that V is square and hence unitary, so $VV^T = \boxed{I}$. (Alternatively, since A is 10×3 with full column rank, the row space is all of $\mathbb{R}^3$, so the projection must be I.)

Note that we could also compute the projection onto $C(A)$ by the formula $A(A^TA)^{-1}A^T$ if we substitute $A = U\Sigma V^T$ and use the fact that V is square and hence $V^T = V^{-1}$: $U\Sigma V^T(V\Sigma^T \cancel{U^T U}\Sigma V^T)^{-1}V\Sigma^T U^T = U\Sigma \cancel{V^T (V^T)^{-1}}\Sigma^{-2}\cancel{(V)^{-1}V}\Sigma U^T = UU^T$; not only is this a lot more work, but it also doesn't exploit the fact that we *know* that U is an orthonormal basis for $C(A)$. Similarly, we could use $A^T(AA^T)^{-1}A$ to project onto $C(A^T)$ and simplify, but this is even tricker to get the algebra right with because $AA^T = U\Sigma^2 U^T$ but U is not invertible (it isn't square!).

(iii) A good rank-2 approximation for A is ________ (in terms of U, V)

Solution: We get a good rank-2 approximation (in some sense the "best" rank-2 approximation) by setting the third singular value to zero, i.e.

$$\boxed{U \begin{pmatrix} 100 & & \\ & 10 & \\ & & \mathbf{0} \end{pmatrix} V^T} = \boxed{100 u_1 v_1^T + 10 u_2 v_2^T}$$

where u_1, u_2 are the first two columns of U and v_1, v_2 are the first two columns of V.

(b) If$f(x) = (x^T y)^2$ for $x, y \in \mathbb{R}^n$, then give a formula for ∇f (in terms of y and/or x).

Solution: Using the product rule,

$$df = d(x^T y)(x^T y) + (x^T y)d(x^T y) = 2(x^T y)(dx^T y) = \underbrace{2(x^T y)y^T}_{(\nabla f)^T} dx$$

so $\boxed{\nabla f = 2(x^T y)y}$. Alternatively, we could have used the power rule $df = 2(x^T y)d(x^T y)$.

Note that the parentheses are important here. If we write it without parentheses, we might be tempted to write $2x^T yy = 2x^T y^2$, but this is nonsense—you can't multiply $yy = y^2$ because y is a column vector. To get an expression that is associative (i.e., which works regardless of where/whether we put parentheses), we would have to write the gradient as something like $\nabla f = 2yx^T y$ or $\nabla f = 2yy^T x$, using the fact that $x^T y = y^T x$ is a scalar that we can move around freely.