Problem Set 9 due May 14, at 10 PM, on Gradescope

Please list all of your sources: collaborators, written materials (other than our textbook and lecture notes) and online materials (other than Gilbert Strang's videos on OCW).

Give complete solutions, providing justifications for every step of the argument. Points will be deducted for insufficient explanation or answers that come out of the blue.

Problem 1:

Consider the matrix $A = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & 1 \\ 0 & 2 & 2 \end{bmatrix}$.

(1) Compute the SVD of A, and the pseudo-inverse A^+. *(15 points)*

(2) Compute the vector $\boldsymbol{v}^+$ defined by formula (263) in the lecture notes, which will have the property that $\boldsymbol{p} := A\boldsymbol{v}^+$ is as close to $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as possible. *(5 points)*

(3) Compute all solutions to $A\boldsymbol{v} = \boldsymbol{p}$ and prove that $\boldsymbol{v}^+$ is the shortest one. *(10 points)*

Solution: (1) First compute the matrix:

$$A^T A = \begin{bmatrix} 5 & -7 & -2 \\ -7 & 14 & 7 \\ -2 & 7 & 5 \end{bmatrix}$$

and its characteristic polynomial $-\lambda^3 + 24\lambda^2 - 63\lambda = -\lambda(\lambda - 21)(\lambda - 3)$. Therefore, the eigenvalues of $A^T A$ are $\lambda_1 = 21$, $\lambda_2 = 3$ and $\lambda_3 = 0$. Its eigenvectors (scaled so that they have unit length) are:

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \qquad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \boldsymbol{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \tag{1}$$

Therefore A has rank 2 and singular values equal to $\sigma_1 = \sqrt{21}$ and $\sigma_2 = \sqrt{3}$. Its left singular vectors are precisely (1), and to compute the right singular vectors we use the following formulas:

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \frac{1}{\sqrt{14}} \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{u}_2 = \frac{1}{\sigma_2} A\boldsymbol{v}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

Because $A\boldsymbol{v}_3 = 0$, we may just pick $\boldsymbol{u}_3$ so that it is orthonormal to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ above. To find such a $\boldsymbol{u}_3$, just start from an arbitrary vector in $\mathbb{R}^3$ and perform Gram-Schmidt with respect to $\boldsymbol{u}_1$, $\boldsymbol{u}_2$:

$$\boldsymbol{u}_3 = \frac{1}{\sqrt{21}} \begin{bmatrix} -4 \\ -2 \\ 1 \end{bmatrix}$$

So we conclude that:

$$A = U\Sigma V^T$$

where:

$$U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{bmatrix} = \frac{1}{\sqrt{42}} \begin{bmatrix} -\sqrt{3} & \sqrt{7} & -4\sqrt{2} \\ 3\sqrt{3} & -\sqrt{7} & -2\sqrt{2} \\ 2\sqrt{3} & 2\sqrt{7} & \sqrt{2} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sqrt{21} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & \sqrt{3} & -\sqrt{2} \\ 2 & 0 & -\sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \end{bmatrix}$$

Then the pseudo-inverse is:

$$A^+ = V\Sigma^+U^T = \frac{1}{21} \begin{bmatrix} 4 & -5 & 6 \\ -1 & 3 & 2 \\ 3 & -2 & 8 \end{bmatrix}$$

Grading Rubric: 5 points for the singular values of A, 5 points for the left singular vectors and 5 points for the right singular vectors. For each of these three subproblems:

- -2.5 points if the result is correct, but the explanation is insufficient
- -1 point for small computational errors

Solution: (2) The answer is $\boldsymbol{v}^+ = A^+\boldsymbol{b} = \frac{1}{21}\begin{bmatrix} 5 \\ 4 \\ 9 \end{bmatrix}$.

Grading Rubric:

- Correct formula and result (if the student got a different answer for A^+ in part (1), then take their wrong answer as the input in both this part, and the next one) *(5 points)*
- Small computational mistake *(4 points)*
- Wrong formula *(0 points)*

Solution: (3) All other solutions $\boldsymbol{v}$ to $A\boldsymbol{v} = \boldsymbol{p} = A\boldsymbol{v}^+$ have the property that:

$$\boldsymbol{v} = \boldsymbol{v}^+ + \boldsymbol{w}$$

for some $\boldsymbol{w}$ in the nullspace of A. The nullspace of A is one-dimensional and spanned by the eigenvector $\boldsymbol{v}_3$. Therefore, the answer is:

$$\boldsymbol{v} = \frac{1}{21}\begin{bmatrix} 5 \\ 4 \\ 9 \end{bmatrix} + \alpha \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

for all numbers α. Since $\boldsymbol{v}^+ \perp \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$, we have:

$$||\boldsymbol{v}||^2 = ||\boldsymbol{v}^+||^2 + \alpha^2 \cdot 3$$

which implies that $\boldsymbol{v}^+$ is shorter than $\boldsymbol{v}$.

Grading Rubric: 5 points for the complete solution and 5 points for the proof.

Problem 2:

(1) Compute the sixth roots of unity (i.e. the complex numbers z such that $z^6 = 1$) in both Cartesian (i.e. $a + bi$) and polar (i.e. $re^{i\theta}$) form. Draw them all on a picture of the plane.

(10 points)

(2) Prove the double angle and triple angle formulas:

$$\cos(2\theta) = 2(\cos\theta)^2 - 1 \qquad \text{and} \qquad \cos(3\theta) = 4(\cos\theta)^3 - 3\cos\theta \tag{2}$$

by the following logic:

- think of $\cos\theta$ as the real part of the complex number $z = e^{i\theta} = a + bi$ where $a = \cos\theta$, $b = \sin\theta$
- then compute z^2 (respectively z^3) first in polar form, and
- finally compute z^2 (respectively z^3) in Cartesian form.

By equating the results in the last two bullets, you should obtain (2). *(10 points)*

Solution: (1) In polar form, if $z = re^{i\theta}$ satisfies $z^6 = 1$, then:

$$1 = r^6 e^{i6\theta}$$

By taking absolute values in the equality above, we get $1 = r^6$, and because r is a positive real number this implies that $r = 1$. Meanwhile, by taking angular parts in the equality above, we must have:

$$6\theta = \text{integer multiple of } 2\pi$$

because only the integer multiples of 2π have the same angular coordinate as the number 1. Hence:

$$\theta \in \left\{0, \frac{2\pi}{6}, \frac{4\pi}{6}, \frac{6\pi}{6}, \frac{8\pi}{6}, \frac{10\pi}{6}\right\}$$

The reason why we don't take other multiples of $\frac{2\pi}{6}$ is that from $\frac{12\pi}{6}$, they start repeating themselves with period 2π, hence we would just be getting the same angles over and over again. Therefore, the answer is:

$$z \in \left\{1, e^{\frac{\pi}{3}}, e^{\frac{2\pi}{3}}, e^{\frac{3\pi}{3}}, e^{\frac{4\pi}{3}}, e^{\frac{5\pi}{3}}\right\}$$

in polar form. In Cartesian form, these numbers are:

$$z \in \left\{1, \frac{1+i\sqrt{3}}{2}, \frac{-1+i\sqrt{3}}{2}, -1, \frac{-1-i\sqrt{3}}{2}, \frac{1-i\sqrt{3}}{2}\right\}$$

Geometrically, the 6 points corresponding to the 6 complex numbers above are the vertices of a regular hexagon on the unit circle (radius 1 centered at the origin) which has 1 as one of its vertices.

Grading Rubric: 4 points for the answers in polar form, 3 points for the answers in Cartesian form, 3 points for the picture.

Solution: (2) Let $z = e^{i\theta} = a + bi$ where $a = \cos\theta$ and $b = \sin\theta$. On one hand, we have:

$$z^2 = e^{i2\theta} = \cos 2\theta + i \sin 2\theta \qquad \text{and} \qquad z^3 = e^{i3\theta} = \cos 3\theta + i \sin 3\theta$$

But on the other hand, in Cartesian coordinates, we have:

$$z^2 = (a+bi)^2 = a^2 - b^2 + 2abi \qquad \text{and} \qquad z^3 = (a+bi)^3 = a^3 - 3ab^2 + (3a^2b - b^3)i$$

By equating the real parts in the two equations above, we get:

$$\cos 2\theta = a^2 - b^2 = (\cos\theta)^2 - (\sin\theta)^2 \qquad \text{and} \qquad \cos 3\theta = (\cos\theta)^3 - 3(\cos\theta)(\sin\theta)^2$$

If you substitute $(\sin\theta)^2 = 1 - (\cos\theta)^2$, you get precisely the double/triple angle formulas (2).

Grading Rubric:

- Correct solution following the blueprint provided *(10 points)*
- Correct solution following the blueprint provided, with minor computational errors *(8-9 points)*
- Correctly performed computations in the three bullets, but did not conclude answer *(7 points)*
- Correctly obtained the required angle formulas by a different method *(3 points)*
- No answer, or significantly incorrect answer *(0 points)*

Problem 3:

Suppose we have a pair of trick coins which we toss one after the other, and they behave as follows. The first coin is fair, but if the first coin shows heads then the second coin will automatically show head as well, while if the first coin shows tails then the second coin is fair.

(1) Consider the random vector:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

where X_1 (respectively X_2) is 0 or 1 depending on whether the first (respectively second) coin shows heads or tails. Compute the mean $\boldsymbol{\mu}$ and covariance matrix K of $\boldsymbol{X}$. *(10 points)*

(2) Diagonalize the covariance matrix $K = QDQ^T$ where Q is orthogonal and D is diagonal. *(10 points)*

(3) Do principal component analysis to find two linear combinations of X_1 and X_2 which are uncorrelated (i.e. have covariance 0). Find the variances of these linear combinations. *(10 points)*

Solution: (1) Let's first work out all the outcomes of the coin toss, with their probabilities:

$$\boldsymbol{X} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ means tails-tails, and happens with probability } \frac{1}{4}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ means tails-heads, and happens with probability } \frac{1}{4}$$

$$\boldsymbol{X} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ means heads-tails, and happens with probability } 0$$

$$\boldsymbol{X} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ means heads-heads, and happens with probability } \frac{1}{2}$$

Therefore, the mean is:

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \frac{1}{4} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \frac{1}{4} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot 0 + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \frac{1}{2} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix}$$

and the covariance matrix is:

$$K = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix} \right) \left(\begin{bmatrix} 1 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \end{bmatrix} \right) \cdot \frac{1}{4} + \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix} \right) \left(\begin{bmatrix} 1 & 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \end{bmatrix} \right) \cdot \frac{1}{4} +$$

$$+ \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix} \right) \left(\begin{bmatrix} 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \end{bmatrix} \right) \cdot 0 + \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix} \right) \left(\begin{bmatrix} 0 & 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \end{bmatrix} \right) \cdot \frac{1}{2} = \frac{1}{16} \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$$

Grading Rubric:

- Set up and solved the problem correctly (including indicating the correct probabilities, and the correct formulas for the mean and covariance) *(10 points)*
- Same as above, but minor algebra errors *(8-9 points)*
- Computed the correct probabilities, but substantially wrong formula for the mean <u>or</u> covariance *(7 points)*
- Computed the correct probabilities, but substantially wrong formula for the mean <u>and</u> covariance *(4 points)*
- -1 point for each probability incorrectly computed (if you subtract a point from here, use their wrong answer from now on, so for instance do not deduct under "minor algebra errors")

Solution: (2) We've done this so many times that I won't bore you with the process anymore (although you are expected to show your work). The method is: find the characteristic polynomial, solve for the eigenvalues, and then find eigenvectors. The answer is:

$$K = QDQ^T \quad \text{where} \quad Q = \begin{bmatrix} \frac{1-\sqrt{17}}{\sqrt{34-2\sqrt{17}}} & \frac{1+\sqrt{17}}{\sqrt{34+2\sqrt{17}}} \\ \frac{4}{\sqrt{34-2\sqrt{17}}} & \frac{4}{\sqrt{34+2\sqrt{17}}} \end{bmatrix} \quad \text{and} \quad D = \frac{1}{32}\begin{bmatrix} 7-\sqrt{17} & 0 \\ 0 & 7+\sqrt{17} \end{bmatrix}$$

Grading Rubric: 5 points for the eigenvalues, 5 points for the eigenvectors. Of these, give full points only if the explanations are convincing, and half points if the result is correct but the explanations are lacking (deduct at most 1 point for minor algebra errors). Deduct 2 points if the columns of Q are not normalized to have length 1.

Solution: (3) If we consider the random vector:

$$\boldsymbol{Y} = Q^T \boldsymbol{X} = \begin{bmatrix} \frac{1-\sqrt{17}}{\sqrt{34-2\sqrt{17}}} & \frac{4}{\sqrt{34-2\sqrt{17}}} \\ \frac{1+\sqrt{17}}{\sqrt{34+2\sqrt{17}}} & \frac{4}{\sqrt{34+2\sqrt{17}}} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

then $\boldsymbol{Y}$ has diagonal covariance matrix D. Hence the entries of $\boldsymbol{Y}$, namely:

$$Y_1 = \frac{1-\sqrt{17}}{\sqrt{34-2\sqrt{17}}} \cdot X_1 + \frac{4}{\sqrt{34-2\sqrt{17}}} \cdot X_2$$

and:

$$Y_2 = \frac{1+\sqrt{17}}{\sqrt{34+2\sqrt{17}}} \cdot X_1 + \frac{4}{\sqrt{34+2\sqrt{17}}} \cdot X_2$$

are uncorrelated, because their covariance should match the off-diagonal entries of D. The variances of Y_1 and Y_2 are the same as the diagonal entries of D, so:

$$\frac{7-\sqrt{17}}{32} \quad \text{and} \quad \frac{7+\sqrt{17}}{32}$$

respectively.

Grading Rubric:

- Correct application of PCA *(10 points)*
- Mostly correct application of PCA, but small errors (e.g. using Q instead of Q^T) *(7 points)*
- Significant errors in identifying Y_1, Y_2 and their variances *(4 points)*
- Missing or extremely incorrect answer *(0 points)*

Problem 4:

(1) Prove that the total probability of a normal distribution is 1, by the following algorithm:

- Let $\alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$. Fill in the blank: the goal is to show that ____________

- Prove that $\alpha^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy$.

- Prove that $\alpha^2 = 1$ by converting the double integral to polar coordinates: $x = r\cos\theta$, $y = r\sin\theta$. Recall that the differential changes according to $dxdy = rdrd\theta$. Conclude that $\alpha = 1$.

Hint: The last bullet requires you to perform a little integration trick. *(10 points)*

(2) Let $\boldsymbol{X}$ be a random vector with mean $\boldsymbol{\mu}$. Show that the covariance matrix, which is defined by:

$$E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X}^T - \boldsymbol{\mu}^T)]$$

can also be given by the formula $E[\boldsymbol{X}\boldsymbol{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T$ (this is handy if you're ever trying to compute a covariance matrix, and don't want to deal with all those $\boldsymbol{\mu}$'s being subtracted). *(10 points)*

Solution: (1) Since double integrals can be computed one at a time, we have:

$$\alpha^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy$$

Changing to polar coordinates gives us:

$$\alpha^2 = \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} rdrd\theta$$

Since the integrand is simply constant in θ, we get:

$$\alpha^2 = \int_{0}^{\infty} e^{-\frac{r^2}{2}} rdr$$

Change the variable to $u = \frac{r^2}{2}$, and the formula above reads:

$$\alpha^2 = \int_{0}^{\infty} e^{-u} du = -e^{-u}\Big|_{0}^{\infty} = 1$$

Grading Rubric:

- Successfully did all 3 bullets *(10 points)*
- Successfully did the first 2 bullets *(6 points)*
- Successfully did the first bullet *(3 points)*
- Did not solve any part correctly *(0 points)*

Solution: (2) Since the mean of a constant random vector (such as $\boldsymbol{\mu}$) is precisely that same random vector, we have:

$$E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X}^T - \boldsymbol{\mu}^T)] = E[\boldsymbol{X}\boldsymbol{X}^T] - \boldsymbol{\mu}E[\boldsymbol{X}^T] - E[\boldsymbol{X}]\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T$$

Since $E[\boldsymbol{X}] = \boldsymbol{\mu}$ and $E[\boldsymbol{X}^T] = \boldsymbol{\mu}^T$, the formula above is equal to:

$$E[\boldsymbol{X}\boldsymbol{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T = E[\boldsymbol{X}\boldsymbol{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T$$

Grading Rubric:

- Correct proof *(10 points)*
- Incorrect proof *(0 points)*