HW#1 Solution

Submission: Submit your assignment as a single zip file containing your code and report (in PDF). Name the file as `HW1_LastName.zip` and upload it to the CW portal.

Problem 1: Linear Independence and Orthogonality

Given the vector $v_1 = \begin{bmatrix} 2 & -5 & 3 \end{bmatrix}^T$,

a) Express vector v_1 as a linear combination of the vectors $u_1 = \begin{bmatrix} 1 & -3 & 2 \end{bmatrix}^T$, $u_2 = \begin{bmatrix} 2 & -4 & -1 \end{bmatrix}^T$ and $u_3 = \begin{bmatrix} 1 & -5 & 7 \end{bmatrix}^T$.

b) Determine the value of k such that the vectors $u = \begin{bmatrix} 3 & 3k & -4 \end{bmatrix}^T$ and v are orthogonal.

Solution:

a)

$$\begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix} = x \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} + y \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix} + z \begin{bmatrix} 1 \\ -5 \\ 7 \end{bmatrix} = \begin{bmatrix} x + 2y + z \\ -3x - 4y - 5z \\ 2x - y + 7z \end{bmatrix}$$

Thus, we have:

$$\begin{bmatrix} 1 & 2 & 1 \\ -3 & -4 & -5 \\ 2 & -1 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix}$$

We can find the Reduced Row Echelon Form (RREF) or since the coefficient matrix is square we can check the determinant. Given that the determinant is zero, the vectors are dependent and we can not express vector v_1 as a linear combination of the vectors u.

b)

$$\langle u, v \rangle = u^T v = 6 - 15k - 12 = 0 \rightarrow k = -\frac{6}{15}$$

Problem 2: Eigenvalues and Eigenvectors

Consider the matrix A:

$$A = \begin{bmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{bmatrix}$$

a) Find the eigenvalues and eigenvectors of A.

b) Is the matrix A diagonalizable? In other words, does there exist an invertible matrix P such that, with the following relation, the resulting matrix D is diagonal:

$$D = P^{-1} A P$$

c) Determine the rank of the matrix.

d) Show that the sum of the eigenvalues is equal to the trace of the matrix (the sum of its diagonal elements).

e) Show that the product of the eigenvalues is equal to the determinant of the matrix.

Solution:

a)

$$\det(\lambda I - A) = \lambda^3 - 11\lambda^2 + 39\lambda - 45 = (\lambda - 3)^2(\lambda - 5) \rightarrow \lambda_{1,2} = 3, \lambda_3 = 5$$

We have a repeated eigenvalue $\lambda = 3$, so we must determine that there are independent eigenvectors for this eigenvalue or we should find generalized eigenvectors. To this end, we check the algebraic multiplicity k which is 2 and geometric multiplicity of $\lambda = 3$ which is $\alpha = n - \text{rank}(3I - A) = 3 - 1 = 2$. The eigenvectors correspond to $\lambda = 3$ can be found as:

$$(3I - A)x = 0 \rightarrow \begin{bmatrix} -1 & -1 & 1 \\ -2 & -2 & 2 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \xrightarrow{RREF} \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

Thus, we can write the equation as $x_1 + x_2 - x_3 = 0$ and by choosing x_2 and X_3 as free variables, the corresponding eigenvectors can be found as follows:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} x_3 \rightarrow v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Moreover, for $\lambda_3 = 5$, we have:

$$(5I - A)x = 0 \rightarrow \begin{bmatrix} 1 & -1 & 1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

Thus, we can write the equations as $x_1 - x_3 = 0$ and $x_2 - 2x_3 = 0$, and by choosing x_3 as the free variable, the corresponding eigenvector can be found as follows:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} x_3 \rightarrow v_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

b) Since two multiplicities are equal there are independent eigenvectors corresponding the the repeated eigenvalue and we can find a fully diagonal form of the matrix. To this end, we put the eigenvectors in the matrix as

$$P = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}$$

that result in the diagonal form as follows

$$D = P^{-1}AP = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

c) We can find the rank based on the RREF of the matrix as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which indicates that the matrix is full rank.

d) The characteristic equation for a square matrix A of order n is $\det(\lambda I - A) = 0$, i.e.

$$(-1)^n\lambda^n + S_{n-1}\lambda^{n-1} + \cdots + S_1\lambda + S_0 = 0$$

where S_{n-1} is the negative sum of the diagonal elements of A.

$$S_{n-1} = -\text{tr}(A)$$

From the fundamental theorem of algebra, the eigenvalues are precisely the roots of the characteristic polynomial

$$\lambda^n + S_{n-1}\lambda^{n-1} + \cdots + S_1\lambda + S_0 = 0$$

From polynomial theory, the sum of the roots (i.e., the sum of eigenvalues) of a monic polynomial of degree n

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = -S_{n-1}$$

Since $S_{n-1} = -\text{tr}(A)$, we get:

$$\text{tr}(A) = \sum \lambda_i$$

Check for the example:

$$4 + 5 + 2 = 3 + 3 + 5$$

e) Since the matrix can be diagonalizable $D = P^{-1}AP$ we have:

$$A = PDP^{-1} \to \det(A) = \det(PDP^{-1}) = \lambda_1\lambda_2\lambda_3 = 45$$

Problem 3: Linear Transformations

Each point $(x, y) \in \mathbb{R}^2$ can be identified with the point $(x, y, 1)$ on the plane in $\mathbb{R}^3$ that lies one unit above the xy-plane. We say that (x, y) has *homogeneous coordinates* $(x, y, 1)$.

a) Based on the homogeneous coordinates, find the corresponding 3×3 transformation matrix that translates a 2D point (x, y) to a shifted point $(x + h, y + k)$.

b) Any linear transformation on $\mathbb{R}^2$ is represented with respect to homogeneous coordinates by a partitioned matrix of the form

$$\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}$$

where A is a 2×2 matrix. At first, find the corresponding 3×3 transformation matrix that scales x by s and y by t. Afterward, find the 3×3 transformation matrix that result in a counterclockwise rotation of ϕ about the origin.

c) Find the 3×3 matrix that corresponds to the composite transformation of a scaling by .3, a rotation of 90 degrees about the origin, and finally, a translation that adds $(-0.5, 2)$ to the point (x, y).

d) (Bonus) Write a Python program that creates the original figure and applies the given transformations in part c to obtain the final figure.

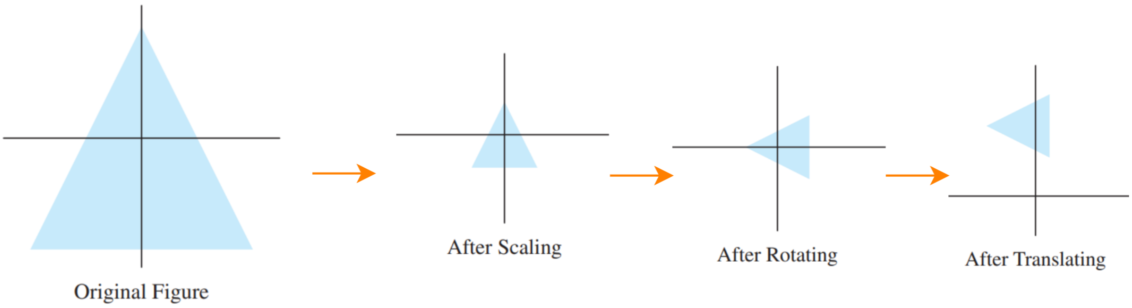

Figura 1: Visualization of the original figure and the sequential transformations

Solution:

a)

$$A_{sh} = \begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix}$$

b)

$$A_s = \begin{bmatrix} s & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_r = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

c)

$$A = \underbrace{\begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Translate}} \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Rotate}} \underbrace{\begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Scale}} = \begin{bmatrix} 0 & -0.3 & -0.5 \\ 0.3 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

Problem 4: Calculus

Compute the derivatives df/dx of the following functions by using the chain rule. Provide the dimensions of every single partial derivative. Describe your steps in detail.

a) $f(z) = log(1+z), \quad z = x^T x, \quad x \in \mathbb{R}^N$

b) $f = tanh(z), \quad z = Ax + b, \quad x \in \mathbb{R}^N, \quad A \in \mathbb{R}^{M\times N}, \quad b \in \mathbb{R}^M, \quad f \in \mathbb{R}^M$

Solution:

a)

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} \in \mathbb{R}^{1\times D}$$

$$\begin{aligned}\frac{\partial f}{\partial z} &= \frac{1}{1+z} = \frac{1}{1+x^T x} \\ \frac{\partial z}{\partial x} &= 2x^T\end{aligned}$$

$$\rightarrow \frac{\mathrm{d}f}{\mathrm{d}x} = \frac{2x^T}{1+x^T x}$$

b)

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} \in \mathbb{R}^{M\times N}$$

$$\begin{aligned}\frac{\partial f}{\partial z} &= \mathrm{diag}(1-\tanh^2(z)) \in \mathbb{R}^{M\times M} \\ \frac{\partial z}{\partial x} &= \frac{\partial Ax}{\partial x} = A \in \mathbb{R}^{M\times N}\end{aligned}$$

We get the latter result by defining $y = Ax$, such that

$$y_i = \sum_j A_{ij}x_j \rightarrow \frac{\partial y_i}{\partial x_k} = A_{ik} \rightarrow \frac{\partial y_i}{\partial x} = \begin{bmatrix} A_{i1}, \cdots, A_{iN}\end{bmatrix} \in \mathbb{R}^{1\times N} \rightarrow \frac{\partial y}{\partial x} = A$$

Problem 5: Linear Algebra, Least Squares and Calculus

Consider the optimization problem of least squares with $\ell 2$-regularization:

$$w^* = \mathrm{argmin}_w f(w)$$

$$f(w) = \frac{1}{2n}\left\|X^T w - y\right\|_2^2 + \frac{\lambda}{2}\|w\|_2^2$$

To solve the above problem, we use the gradient descent method step by step. In each step, we move in the opposite direction of the gradient to reach a local minimum for the optimization problem. The initial value is chosen randomly.

$$w_{k+1} = w_k - \alpha\nabla f(w_k)$$

The value of α is an arbitrary number and is considered as:

$$\alpha = \frac{1}{\sigma_{max}(A)}$$

where $\sigma_{max}(A)$ is the largest singular value of matrix A, which is defined as follows:

$$A = \frac{1}{n}XX^T + \lambda I$$

a) Prove that:

$$\nabla f(w) = Aw - \frac{1}{n}Xy = A(w - w^*)$$

b) Prove that matrix A is positive semi-definite.

c) Prove that:

$$\|w_{k+1} - w^*\| \leq \left(1 - \frac{\sigma_{\min}(A)}{\sigma_{\max}(A)}\right) \|w_k - w^*\|$$

Solution:

a)

$$\nabla f(w) = \frac{1}{2n} \times 2X(X^T w - y) + \lambda w = \left(\frac{XX^T}{n} + \lambda I\right) w - \frac{Xy}{n} = Aw - \frac{1}{n} Xy$$

If $w = w^*$ the derivative is zero and we have $Aw^* = Xy/n$, therefore $A(w - w^*) = \nabla f(w)$

b) Simply it can be shown that $A^T = A$ and we have:

$$x^T \left[\frac{1}{n} XX^T + \lambda I\right] x = \frac{1}{n} \left\|X^T x\right\|^2 + \lambda \|x\|^2 \geq 0$$

c)

$$\begin{aligned}
\|w_{k+1} - w^*\| &= \|w_k - \alpha \nabla f(w_k) - w^*\| \\
&= \|w_k - \alpha A (w_k - w^*) - w^*\| \\
&= \|(w_k - w^*) (I - \alpha A)\| \\
&\leq \|(w_k - w^*)\| \, \sigma_{max}(I - \alpha A) \\
&= \|(w_k - w^*)\| \, (1 - \alpha \sigma_{min}(A)) \\
&= \|(w_k - w^*)\| \left(1 - \frac{\sigma_{min}(A)}{\sigma_{max}(A)}\right)
\end{aligned}$$

Problem 6: Least Square Method for Curve Fitting

Find the least-squares line $y = \beta_0 + \beta_1 x$ that best fits the data $(-2, 3), (-1, 5), (0, 5), (1, 4)$ and $(2, 3)$. Suppose the errors in measuring the y-values of the last two data points are greater than for the other points. Weight these data half as much as the rest of the data. (Hint: Formulate this as a weighted least squares problem in the form $WAx = Wy$, and find the solution x^* by determining the weight matrix W and applying the least squares method.)

Solution: The weighted least-squares problem is to find approximate solution $\hat{\mathbf{x}}$ that makes $\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{y}}$ as close to $\mathbf{y}$ as possible and the corresponding equation is

$$\mathbf{WAx} = \mathbf{Wy}$$

The normal equation for the weighted least-squares solution is

$$(\mathbf{WA})^T \mathbf{WAx} = (\mathbf{WA})^T \mathbf{Wy}$$

The least-squares line is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals. Therefore, we define the matrices as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 4 \\ 3 \end{bmatrix}$$

For a weighting matrix, we choose $\mathbf{W} = \text{diag}([2, 2, 2, 1, 1])$. Left multiplication by $\mathbf{W}$ scales the rows of $\mathbf{X}$ and $\mathbf{y}$:

$$\mathbf{WX} = \begin{bmatrix} 2 & -4 \\ 2 & -2 \\ 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{Wy} = \begin{bmatrix} 6 \\ 10 \\ 10 \\ 4 \\ 3 \end{bmatrix}$$

For the normal equation, we compute

$$\mathbf{WX})^T \mathbf{WX} = \begin{bmatrix} 14 & -9 \\ -9 & 25 \end{bmatrix} \quad \text{and} \quad (\mathbf{WX})^T \mathbf{Wy} = \begin{bmatrix} 59 \\ -34 \end{bmatrix}$$

Thus, we have:

$$\begin{bmatrix} 14 & -9 \\ -9 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix} = \begin{bmatrix} 59 & -34 \end{bmatrix}$$

The solution of the normal equation is (to two significant digits) $\beta_0 = 4.3$ and $\beta_1 = 0.2$. The desired line is

$$y = 4.3 + 0.2x$$

Problem 7: Optimization

A manufacturing company produces two products, x and y. The profit function is given by:

$$P(x, y) = 100x + 150y - 0.1x^2 - 0.2y^2 - 0.05xy$$

The company has the constraint of that the total resources used cannot exceed 1000 units:

$$2x + 3y \leq 1000$$

The goal is to maximize the profit. Solve the problem and obtain the optimal production levels and the corresponding profit in the following cases:

a) Unconstrained optimization

b) Constrained optimization (using Lagrangian multipliers)

Solution:

a) To find the critical points, take partial derivatives and set them to zero:

$$\frac{\partial P}{\partial x} = 100 - 0.2x - 0.05y = 0$$
$$\frac{\partial P}{\partial y} = 150 - 0.4y - 0.05x = 0$$

Rewrite the equations:

$$0.2x + 0.05y = 100$$
$$0.05x + 0.4y = 150$$

Therefore the approximate solution (floating number of products) is:

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} 0.2 & 0.05 \\ 0.05 & 0.4 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 150 \end{bmatrix} \approx \begin{bmatrix} 419.355 \\ 322.581 \end{bmatrix}$$

We consider the integer values $x^* = 419$ and $y^* = 322$ and the profit at this point is $P^* = 45161$.

b) Rewrite the constraint as equality (since the maximum will lie on the boundary of the constraints). The Lagrangian function is:

$$\mathcal{L}(x, y, \lambda) = 100x + 150y - 0.1x^2 - 0.2y^2 - 0.05xy - \lambda(2x + 3y - 1,000)$$

Take partial derivatives and set them to zero:

$$\frac{\partial \mathcal{L}}{\partial x} = 100 - 0.2x - 0.05y - 2\lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 150 - 0.4y - 0.05x - 3\lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(2x + 3y - 1,000) = 0$$

Rewrite the equations:

$$0.2x + 0.05y + 2\lambda = 100$$
$$0.05x + 0.4y + 3\lambda = 150$$
$$2x + 3y = 1000$$

Therefore the approximate solution (floating number of products) is:

$$\begin{bmatrix} x^* \\ y^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} 0.2 & 0.05 & 2 \\ 0.05 & 0.4 & 3 \\ 2 & 3 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 150 \\ 1000 \end{bmatrix} \approx \begin{bmatrix} 232.143 \\ 178.571 \\ 22.321 \end{bmatrix}$$

We consider the integer values $x^* = 232$ and $y^* = 178$ and the profit at this point is $P^* = 36116$.

Problem 8: Probability and Statistics

Consider normal random variables X and Y with means and variances $\mu_X = 0$, $\sigma_X^2 = 1$, $\mu_Y = -1$, $\sigma_Y^2 = 4$ and correlation coefficient $\rho = -1/2$.

(a) Find $P(X + Y > 0)$.

(b) Find a given that $X + 2Y$ and $aX + Y$ are independent.

(c) Find the correlation between $X + Y$ and $2X - Y$.

Solution:

(a) In order to find the probability $P(X + Y > 0)$, we define the random variable $W = X + Y$ which has a normal distribution (due to the linear combination of two normal variables) and find its mean value and variance as follows:

$$E[W] = E[X] + E[Y] = -1.$$

$$\begin{aligned}
\mathrm{Var}[W] &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\sigma_X \sigma_Y \rho(X, Y) \\
&= 1 + 4 + 2(1)(2)(-\frac{1}{2}) = 3.
\end{aligned}$$

So the distribution of W is $W \sim N(-1, 3)$. To find the probability $P[W > 0]$, we define standard normal variable $Z = \frac{W+1}{\sqrt{3}}$ and use standard cdf as follows:

$$P[W > 0] = 1 - P\left[Z \leq \frac{1}{\sqrt{3}}\right] = 0.2819.$$

(b) The covariance matrix for n variables is shown below:

$$\mathrm{Cov}(\mathbf{x}) = P_{XX} = E\begin{bmatrix} (x_1 - \bar{x}_1)^2 & \cdots & (x_1 - \bar{x}_1)(x_n - \bar{x}_n) \\ \vdots & \ddots & \vdots \\ (x_n - \bar{x}_n)(x_1 - \bar{x}_1) & \cdots & (x_n - \bar{x}_n)^2 \end{bmatrix}$$

We consider the off-diagonal elements as $\mathrm{Cov}(X_i, X_j) = E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)]$ where $i, j = 1, 2, \cdots n$ and $i \neq j$. Thus we have:

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= E[(X_i - E[X_i])(X_j - E[X_j])] \\
&= E[X_i X_j - X_i E[X_j] - E[X_i] X_j + E[X_i] E[X_j]] \\
&= E[X_i X_j] - 2E[X_i]E[X_j] + E[X_i]E[X_j] \\
&= E[X_i X_j] - E[X_i]E[X_j].
\end{aligned}$$

By defining $Z_1 = X + 2Y$ and $Z_2 = aX + Y$ and using the fact that Z_1 and Z_2 are independent, we can note this as $\mathrm{Cov}(Z_1, Z_2) = 0$. Therefore using the above result we can find a by setting the $\mathrm{Cov}(Z_1, Z_2)$ to zero as follows:

$$\begin{aligned}
\mathrm{Cov}(Z_1, Z_2) &= \mathrm{Cov}(X + 2Y, aX + Y) \\
&= E[(X + 2Y)(aX + Y)] - E[X + 2Y]E[aX + Y] \\
&= E[aX^2 + (1 + 2a)XY + 2Y^2] - (E[X] + 2E[Y])(aE[X] + E[Y]) \\
&= a(E[X^2] - E[X]^2) + (1 + 2a)(E[XY] - E[X]E[Y]) + 2(E[Y^2] - E[Y]^2) \\
&= a\mathrm{Var}(X) + (1 + 2a)\mathrm{Cov}(X, Y) + 2\mathrm{Var}(Y) \\
&= a + (1 + 2a)\sigma_X \sigma_Y \rho(X, Y) + 8 = 7 - a = 0 \rightarrow a = 7.
\end{aligned}$$

(c) In order to find the correlation between $X + Y$ and $2X - Y$, we define random variables $Z_3 = X + Y$ and $Z_4 = 2X - Y$ which have normal distributions (due to the linear combination of two normal variables) and find their mean value and variance as follows:

$$\begin{aligned}
E[Z_3] &= E[X] + E[Y] = -1 \\
E[Z_4] &= 2E[X] - E[Y] = 1.
\end{aligned}$$

$$
\begin{aligned}
\text{Var}[Z_3] &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\
&= \text{Var}(X) + \text{Var}(Y) + 2\sigma_X \sigma_Y \rho(X, Y) \\
&= 1 + 4 + 2(1)(2)(-\frac{1}{2}) = 3.
\end{aligned}
$$

and

$$
\begin{aligned}
\text{Var}[Z_4] &= 4\text{Var}(X) + \text{Var}(Y) - 4\text{Cov}(X, Y) \\
&= 4\text{Var}(X) + \text{Var}(Y) - 4\sigma_X \sigma_Y \rho(X, Y) \\
&= 4 + 4 - 4(1)(2)(-\frac{1}{2}) = 12.
\end{aligned}
$$

Afterwards, the $\text{Cov}(Z_3, Z_4)$ can be obtained as follows:

$$
\begin{aligned}
\text{Cov}(Z_3, Z_4) &= \text{Cov}(X + Y, 2X - Y) \\
&= E[(X + Y)(2X - Y)] - E[X + Y]E[2X - Y] \\
&= E[2X^2 + XY - Y^2] - (E[X] + E[Y])(2E[X] - E[Y]) \\
&= 2(E[X^2] - E[X]^2) + (E[XY] - E[X]E[Y]) - (E[Y^2] - E[Y]^2) \\
&= 2Var(X) + \text{Cov}(X, Y) - Var(Y) \\
&= 2 - 1 - 4 = -3.
\end{aligned}
$$

Finally, the correlation between $X + Y$ and $2X - Y$ is

$$
\rho(Z_3, Z_4) = \frac{\text{Cov}(Z_3, Z_4)}{\sqrt{\text{Var}(Z_3)\text{Var}(Z_4)}} = -\frac{1}{2}.
$$

Problem 9: Image Noise Reduction using SVD

In this exercise, you should reduce the noise in the noisy image included in the homework files on CW using singular value decomposition (SVD) and then reconstruct the image. To this end, you can read the image using `matplotlib`, extract the RGB matrices, and apply SVD using the `np.linalg` library to compute the U, S, and V arrays. Then, construct the diagonal S matrix, where the singular values are placed along the diagonal:

$$
S = \begin{bmatrix}
\sigma_1 & 0 & 0 & 0 & 0 \\
0 & \sigma_2 & 0 & 0 & 0 \\
0 & 0 & \ddots & 0 & 0 \\
0 & 0 & 0 & \sigma_k & 0 \\
0 & 0 & 0 & 0 & \ddots
\end{bmatrix}
$$

Given that, starting from the threshold σ_k, the smaller singular values in matrix S have a significant influence on noise and are much lower in value compared to the first few singular values (e.g., σ_1, σ_2, etc.), we can reduce noise by ignoring these smaller singular values (replacing them with zero) and reconstructing the image using the larger singular values, which contain the main features of the image. It should be noted that the appropriate value of k can be determined through trial and error, balancing the trade-off between noise reduction and preserving the main features of the image. If k is too large, the noise may remain unchanged, whereas if k is too small, the image will be overly compressed, potentially losing resolution and quality. Save the final image as a PNG file using `matplotlib`. A sample output of the processed image is shown below:

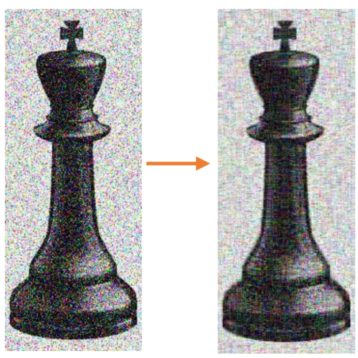

Figura 2: Noisy image (left) and processed image (right)

Solution:

```
from PIL import Image
import numpy as np

k = 30

img = plt.imread('noisy.jpg')
I = np.array(img)

R = I[:, :, 0]
G = I[:, :, 1]
B = I[:, :, 2]

matrices = [R, G, B]
results = []


for matrix in matrices:
   U, s, V = np.linalg.svd(matrix)

   s[k:] = 0
   S = np.zeros(matrix.shape)
   S[:matrix.shape[1], :matrix.shape[1]] = np.diag(s)


   results.append(np.uint8(U.dot(S.dot(V))))

I[:, :, 0] = results[0]
I[:, :, 1] = results[1]
I[:, :, 2] = results[2]

plt.imsave('clear.png', I)
```

Good luck!