

EE263 Homework 8 Solutions

Fall 2025

4.581. Orthogonal projection matrices, LS, LN, and SVD. A matrix $P \in \mathbb{R}^{m \times m}$ is called an *orthogonal projection matrix* if $P = P^\top$ and $P^2 = P$.

- a) Show that if P is an orthogonal projection matrix then so is $I - P$.
- b) Suppose that the columns of $U \in \mathbb{R}^{m \times r}$ are orthonormal. Show that UU^\top is an orthogonal projection matrix. Onto which of the four fundamental subspaces of U ($\text{range}(U)$, $\text{range}(U^\top)$, $\text{null}(U)$, $\text{null}(U^\top)$) does UU^\top project? What about the orthogonal projection matrix $I - UU^\top$?
- c) Suppose $A \in \mathbb{R}^{m \times n}$ is full rank, with $n \leq m$. The fitted vector for least-squares problem

$$\min_x \|Ax - y\|^2$$

is given by $y_{\text{ls}} = A(A^\top A)^{-1}A^\top y$. Show that $P_{\text{ls}} = A(A^\top A)^{-1}A^\top$ is an orthogonal projection matrix. Onto which subspace associated with A does it project?

- d) Now suppose $A \in \mathbb{R}^{m \times n}$ has full row rank ($m \leq n$). The x solving

$$\min_x \|x\|^2 \quad \text{s.t.} \quad Ax = y$$

is given by $x_{\text{ln}} = A^\top(AA^\top)^{-1}y$. Show that $P_{\text{ln}} = A^\top(AA^\top)^{-1}A$ is an orthogonal projection matrix. Onto which subspace associated with A does it project?

- e) Let $A \in \mathbb{R}^{m \times n}$ have rank r and singular value decomposition

$$A = U\Sigma V^\top, \quad U = [U_1 \ U_2], \quad V = [V_1 \ V_2], \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix},$$

where $\Sigma_1 \in \mathbb{R}^{r \times r}$ is diagonal with positive entries, $U_1 \in \mathbb{R}^{m \times r}$, $U_2 \in \mathbb{R}^{m \times (m-r)}$, $V_1 \in \mathbb{R}^{n \times r}$, and $V_2 \in \mathbb{R}^{n \times (n-r)}$. Write each of the four fundamental subspaces of A ($\text{range}(A)$, $\text{range}(A^\top)$, $\text{null}(A)$, $\text{null}(A^\top)$) in terms of the spans of the columns of the matrices U_1, U_2, V_1, V_2 .

- f) The pseudo-inverse of A is defined (in terms of the SVD above) by

$$A^\dagger = V_1 \Sigma_1^{-1} U_1^\top.$$

Show that AA^\dagger and $A^\dagger A$ are orthogonal projection matrices, and identify the subspaces onto which they project.

- g) Let $\mathcal{X}_{\text{ls}} = \arg \min_w \|Aw - y\|$, i.e. the set of all least-squares minimizers. Show that

$$\mathcal{X}_{\text{ls}} = \{x \mid Ax = (AA^\dagger)y\},$$

and that all least-squares minimizers have the same residual $(I - AA^\dagger)y$. Hint: use one of the projection matrices from the previous part to write y in terms of its projection onto $\text{range}(A)$ and its projection onto $\text{null}(A^\top)$.

h) Explain why every $x \in \mathcal{X}_{\text{ls}}$ can be written as

$$x = A^\dagger y + (I - A^\dagger A)z,$$

for some $z \in \mathbb{R}^n$.

i) By considering the norm of the above expression for x show that $x_{\text{pinv}} = A^\dagger y$ is the unique vector of minimum norm among all least-squares minimizers, i.e. x_{pinv} is the minimum-norm least-squares solution.

Solution.

a) To show that $I - P$ is an orthogonal projection matrix we need to check two properties:

- i. $I - P = (I - P)^\top$
- ii. $(I - P)^2 = I - P$.

The first one is easy: $(I - P)^\top = I - P^\top = I - P$ because $P = P^\top$ (P is an orthogonal projection matrix.) To show the second property we have

$$\begin{aligned} (I - P)^2 &= I - 2P + P^2 \\ &= I - 2P + P \quad (\text{since } P = P^2) \\ &= I - P \end{aligned}$$

and we are done.

The projector P projects onto $\text{range}(P)$ and $I - P$ projects onto orthogonal complement of $\text{range}(P)$ which is equal to $\text{null}(P^\top)$.

b) Since the columns of U are orthonormal we have $U^\top U = I$. Using this fact it is easy to prove that UU^\top is an orthogonal projection matrix, i.e., $(UU^\top)^\top = UU^\top$ and $(UU^\top)^2 = UU^\top$. Clearly, $(UU^\top)^\top = (U^\top)^\top U^\top = UU^\top$ and

$$\begin{aligned} (UU^\top)^2 &= (UU^\top)(UU^\top) \\ &= U(U^\top U)U^\top \\ &= UU^\top \quad (\text{since } U^\top U = I). \end{aligned}$$

Thus UU^\top is the orthogonal projector onto $\text{range}(U)$ and $I - UU^\top$ is the orthogonal projector onto $\text{range}(U)^\perp = \text{null}(U^\top)$.

c) We can show that $P_{\text{ls}} = P_{\text{ls}}^\top$ and $P_{\text{ls}}^2 = P_{\text{ls}}$.

$$\begin{aligned} P_{\text{ls}}^\top &= A(A^\top A)^{-1}A^\top = P_{\text{ls}} \\ P_{\text{ls}}^2 &= A(A^\top A)^{-1}(A^\top A)(A^\top A)^{-1}A^\top = P_{\text{ls}}. \end{aligned}$$

Thus P_{ls} is an orthogonal projector. It projects onto $\text{range}(A)$.

d) We can show that $P_{\text{ln}} = P_{\text{ln}}^T$ and $P_{\text{ln}}^2 = P_{\text{ln}}$.

$$P_{\text{ln}}^T = A^T(AA^T)^{-1}A = P_{\text{ln}}$$

$$P_{\text{ln}}^2 = A^T(AA^T)^{-1}(AA^T)(AA^T)^{-1}A = P_{\text{ln}}.$$

Thus P_{ln} is an orthogonal projector. It projects onto $\text{range}(A^T)$.

e) Multiplying out the full SVD we have

$$A = U_1 \Sigma_1 V_1^T,$$

since V_1 is orthogonal and Σ_1 is diagonal with positive entries we can conclude that $\text{range}(A) = \text{span}(U_1)$. The matrix U is orthogonal so $\text{span}(U_2)^\perp = \text{span}(U_1)$ meaning $\text{null}(A^T) = \text{span}(U_2)$. By applying similar reasoning to A^T we can deduce that $\text{range}(A^T) = \text{span}(V_1)$ and $\text{null}(A) = \text{span}(V_2)$.

f) The pseudoinverse is $A^\dagger = V_1 \Sigma_1^{-1} U_1^T$. Then

$$AA^\dagger = U \Sigma V^T V_1 \Sigma_1^{-1} U_1^T = U_1 U_1^T,$$

$$A^\dagger A = V_1 \Sigma_1^{-1} U_1^T U \Sigma V^T = V_1 V_1^T.$$

Using the results from part (b) we can conclude that both are orthogonal projectors and AA^\dagger projects onto $\text{range}(A) = \text{span}(U_1)$ and $A^\dagger A$ projects onto $\text{range}(A^T) = \text{span}(V_1)$.

g) Let $\mathcal{X}_{\text{ls}} = \arg \min_w \|Aw - y\|$. Decompose y orthogonally as

$$y = (AA^\dagger)y + (I - AA^\dagger)y.$$

For any x ,

$$\begin{aligned} \|Ax - y\|^2 &= \|Ax - (AA^\dagger)y - (I - AA^\dagger)y\|^2 \\ &= \|Ax - (AA^\dagger)y\|^2 + \|(I - AA^\dagger)y\|^2, \end{aligned}$$

since $Ax \in \text{range}(A)$ is orthogonal to $(I - AA^\dagger)y \in \text{null}(A^T)$. The second term is independent of x , and the first achieves its minimum value 0 if $Ax = (AA^\dagger)y$. Therefore

$$\mathcal{X}_{\text{ls}} = \{x \mid Ax = (AA^\dagger)y\},$$

and all least-squares minimizers have the residual

$$y - Ax = (I - AA^\dagger)y.$$

h) All solutions to $Ax = (AA^\dagger)y$ can be written as $x = x_p + v$ for some x_p satisfying $Ax_p = (AA^\dagger)y$ and some $v \in \text{null}(A)$. Choose $x_p = A^\dagger y$. From part (f) we know that $A^\dagger A$ is the orthogonal projector onto $\text{range}(A^T)$ so $I - A^\dagger A$ is the orthogonal projector onto $\text{null}(A)$, hence every $x \in \mathcal{X}_{\text{ls}}$ can be written as

$$x = A^\dagger y + (I - A^\dagger A)z,$$

for some $z \in \mathbb{R}^n$.

- i) Since $A^\dagger = V_1 \Sigma_1^{-1} U_1^\top$ we have $A^\dagger y \in \text{span}(V_1) = \text{range}(A^\top)$ and we have already noted that $(I - A^\dagger A)z \in \text{null}(A)$. The two subspaces $\text{range}(A^\top)$ and $\text{null}(A)$ orthogonal complements of each other, therefore norm of any $x \in \mathcal{X}_{\text{ls}}$ is given by

$$\|x\|^2 = \|A^\dagger y\|^2 + \|(I - A^\dagger A)z\|^2,$$

for some $z \in \mathbb{R}^n$. The norm is minimized when when $z = 0$. Thus

$$x_{\text{pinv}} = A^\dagger y$$

is the unique minimum-norm vector among all least-squares minimizers.

16.2560. Blind signal detection. A binary signal s_1, \dots, s_T , with $s_t \in \{-1, 1\}$ is transmitted to a receiver, which receives the (vector) signal $y_t = as_t + v_t \in \mathbb{R}^n$, $t = 1, \dots, T$, where $a \in \mathbb{R}^n$ and $v_t \in \mathbb{R}^n$ is a noise signal. We'll assume that $a \neq 0$, and that the noise signal is centered around zero, but is otherwise unknown. (This last statement is vague, but it will not matter.)

The receiver will form an approximation of the transmitted signal as

$$\hat{s}_t = w^\top y_t, \quad t = 1, \dots, T,$$

where $w \in \mathbb{R}^n$ is a weight vector. Your job is to choose the weight vector w so that $\hat{s}_t \approx s_t$. If you knew the vector a , then a reasonable choice for w would be $w = a^\dagger = a/\|a\|^2$. This choice is the smallest (in norm) vector w for which $w^\top a = 1$.

Here's the catch: You don't know the vector a . Estimating the transmitted signal, given the received signal, when you don't know the mapping from transmitted to received signal (in this case, the vector a) is called *blind signal estimation* or *blind signal detection*.

Here is one approach. Ignoring the noise signal, and assuming that we have chosen w so that $w^\top y_t \approx s_t$, we would have

$$(1/T) \sum_{t=1}^T (w^\top y_t)^2 \approx 1.$$

Since $w^\top v_t$ gives the noise contribution to \hat{s}_t , we want w to be as small as possible. This leads us to choose w to minimize $\|w\|$ subject to $(1/T) \sum_{t=1}^T (w^\top y_t)^2 = 1$. This doesn't determine w uniquely; we can multiply it by -1 and it still minimizes $\|w\|$ subject to $(1/T) \sum_{t=1}^T (w^\top y_t)^2 = 1$. So we can only hope to recover either an approximation of s_t or of $-s_t$; if we don't know a we really can't do any better. (In practice we'd use other methods to determine whether we have recovered s_t or $-s_t$.)

- Explain how to find w , given the received vector signal y_1, \dots, y_T , using concepts from the class.
- Apply the method to the signal in the file `bs_det_data.json`, which contains a matrix Y , whose columns are y_t . Give the weight vector w that you find. Plot a histogram of the values of $w^\top y_t$ using `using Plots; histogram(w'*Y, bins=60)`. You'll know you're doing well if the result has two peaks, one negative and one positive. Once you've chosen w , a reasonable guess of s_t (or, possibly, its negative $-s_t$) is given by

$$\tilde{s}_t = \text{sign}(w^\top y_t), \quad t = 1, \dots, T,$$

where $\text{sign}(u)$ is $+1$ for $u \geq 0$ and -1 for $u < 0$. The file `bs_det_data.json` contains the original signal, as a row vector \mathbf{s} . Give your error rate, *i.e.*, the fraction of times for which $\tilde{s}_t \neq s_t$. (If this is more than 50%, you are welcome to flip the sign on w .)

Solution. We can write

$$\frac{1}{T} \sum_{t=1}^T (w^\top y_t)^2 = (1/T) \|Y^\top w\|^2,$$

where $Y = [y_1 \cdots y_T]$. We must minimize $\|w\|$ subject to $\|Y^\top w\| = \sqrt{T}$. Both of these are homogeneous in w , so we could just as well maximize $\|Y^\top w\|$, subject to $\|w\| = 1$, and then scale the solution so that $\|Y^\top w\| = \sqrt{T}$. To maximize $\|Y^\top w\|$ subject to $\|w\| = 1$ is easy: we take w to be v_1 , the right singular vector associated with the largest singular value of Y^\top . (This is also, by the way, the left singular vector associated with the largest singular value of Y .) We then scale v_1 by α , so that $\|Y^\top(\alpha v_1)\| = \alpha \sigma_1 = \sqrt{T}$, where σ_1 is the largest singular value (*i.e.*, the norm) of Y (or Y^\top). This yields

$$w = (\sqrt{T}/\sigma_1)v_1.$$

Note that we could just as well take the negative of this vector, which would also minimize $\|w\|$ subject to $\|Y^\top w\| = \sqrt{T}$.

Here is the code to do this:

```
include("readclassjson.jl")
using LinearAlgebra
using Plots

data = readclassjson("bs_det_data.json")
Y = data["Y"]
T = data["T"]
s = data["s"]

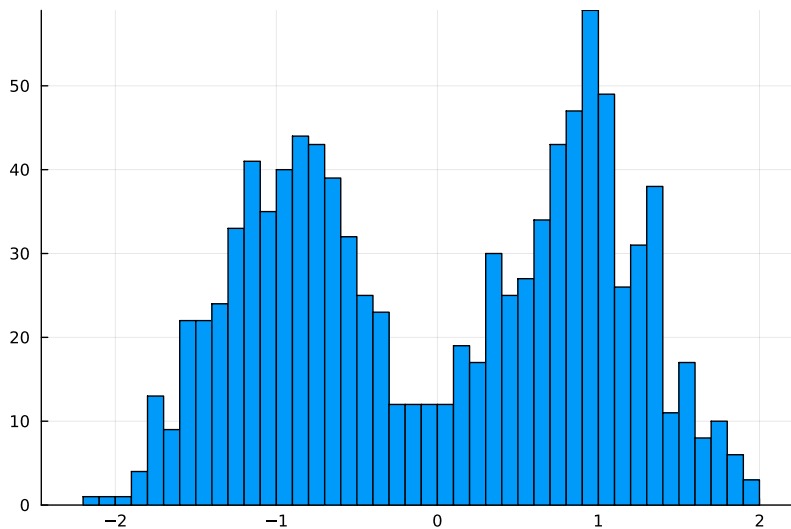
U, S, V = svd(Y')
w = (sqrt(T) / S[1]) * V[:, 1]

shat = (Y' * w)

histogram(shat, bins=60, legend=false)
savefig("bs_det_hist.pdf")

stilde = sign.(shat)
error_rate = sum(s .!= stilde) / T
println("error_rate = ", error_rate)
```

The error rate is 2.9%. The resulting histogram is shown below.



16.2670. Regularization and SVD. Let $A \in \mathbb{R}^{n \times n}$ be full rank, with SVD

$$A = \sum_{i=1}^n \sigma_i u_i v_i^T.$$

(We consider the square, full rank case just for simplicity; it's not too hard to consider the general nonsquare, non-full rank case.) Recall that the regularized approximate solution of $Ax = y$ is defined as the vector $x_{\text{reg}} \in \mathbb{R}^n$ that minimizes the function

$$\|Ax - y\|^2 + \mu \|x\|^2,$$

where $\mu > 0$ is the regularization parameter. The regularized solution is a linear function of y , so it can be expressed as $x_{\text{reg}} = By$ where $B \in \mathbb{R}^{n \times n}$.

a) Express the SVD of B in terms of the SVD of A . To be more specific, let

$$B = \sum_{i=1}^n \tilde{\sigma}_i \tilde{u}_i \tilde{v}_i^T$$

denote the SVD of B . Express $\tilde{\sigma}_i, \tilde{u}_i, \tilde{v}_i$, for $i = 1, \dots, n$, in terms of $\sigma_i, u_i, v_i, i = 1, \dots, n$ (and, possibly, μ). Recall the convention that $\tilde{\sigma}_1 \geq \dots \geq \tilde{\sigma}_n$.

b) Find the norm of B . Give your answer in terms of the SVD of A (and μ).

c) Find the worst-case relative inversion error, defined as

$$\max_{y \neq 0} \frac{\|AB y - y\|}{\|y\|}.$$

Give your answer in terms of the SVD of A (and μ).

Solution.

a) The regularized least-squares solution is given by $x_{\text{rls}(\mu)} = (A^\top A + \mu I)^{-1} A^\top y$, and thus

$$\begin{aligned}
 B &= (A^\top A + \mu I)^{-1} A^\top \\
 &= \left((U \Sigma V^\top)^\top U \Sigma V^\top + \mu I \right)^{-1} (U \Sigma V^\top)^\top \\
 &= (V \Sigma U^\top U \Sigma V^\top + \mu I)^{-1} V \Sigma U^\top \\
 &= (V (\Sigma^2 + \mu I) V^\top)^{-1} V \Sigma U^\top \\
 &= (V (\Sigma^2 + \mu I)^{-1} V^\top) V \Sigma U^\top \\
 &= V (\Sigma^2 + \mu I)^{-1} \Sigma U^\top \\
 &= V \text{diag} \left(\frac{\sigma_i}{\sigma_i^2 + \mu} \right) U^\top.
 \end{aligned}$$

This is almost the SVD of B , except for one detail: the numbers

$$\frac{\sigma_i}{\sigma_i^2 + \mu}$$

aren't necessarily ordered from largest to smallest. Thus we have

$$\tilde{\sigma}_i = \frac{\sigma_{[i]}}{\sigma_{[i]}^2 + \mu} \quad \tilde{u}_i = v_{[i]}, \quad \tilde{v}_i = u_{[i]},$$

where the notation $x_{[i]}$ means the i th largest element of x . (We accepted all sorts of descriptions of this!) One common misconception was that the numbers

$$\frac{\sigma_i}{\sigma_i^2 + \mu}$$

were simply in reverse order, so all that had to be done was to reverse the ordering. That isn't true; just sketch the function $\sigma/(\sigma^2 + \mu)$ as a function of μ to see that it is not always decreasing. (It increases first, then decreases.)

b) The norm of B is its largest singular value, *i.e.*,

$$\|B\| = \max_i \frac{\sigma_i}{\sigma_i^2 + \mu}.$$

c) The worst-case relative inversion error is the matrix norm of $AB - I$:

$$\begin{aligned}
 AB - I &= U \Sigma V^\top V (\Sigma + \mu \Sigma^{-1})^{-1} U^\top - I \\
 &= U \Sigma (\Sigma + \mu \Sigma^{-1})^{-1} U^\top - I \\
 &= U \left((I + \mu \Sigma^{-2})^{-1} - I \right) U^\top \\
 &= U \text{diag} \left(\frac{1}{1 + \mu/\sigma_i^2} - 1 \right) U^\top \\
 &= -U \text{diag} \left(\frac{\mu}{\sigma_i^2 + \mu} \right) U^\top
 \end{aligned}$$

This is the SVD of $AB - I$ (to within reordering). Its largest singular value, *i.e.*, the norm of $AB - I$, is given by

$$\|AB - I\| = \frac{\mu}{\sigma_n^2 + \mu}.$$

Note that we had to absorb the negative sign in the lefthand orthogonal matrix; one common error was to keep the negative sign in the norm. Obviously that couldn't be right because norms are always nonnegative!

16.2900. SVD-based image compression. In this problem we examine how singular value decomposition (SVD) and low-rank approximations can be used to compress images.

Images are naturally represented by matrices. For example, we represent a 512×512 pixel image with a matrix $X \in \mathbb{R}^{512 \times 512}$; here we assume that the entries are real numbers (doubles) instead of, for example, 8 bit binary numbers. To store this image we would require storage for 512^2 doubles, which given that a double (say) occupies 8 bytes, sums up to $512^2 \times 8$ bytes = 2.1 Mbytes.

In `lena_data.json`, you will find a matrix X which is the famous test image in the signal processing community. We will investigate compression of this image using SVD.

- a) Compute and plot the singular values of X . Comment on the distribution of the singular values.
- b) We can save image storage space by using a low-rank approximation of the matrix X , such that we do not lose essential information from the matrix. Find matrix \tilde{X} with $\text{rank}(\tilde{X}) \leq 75$, which is as close as possible to the original image X in the Frobenius norm, *i.e.*, in the $\|X - \tilde{X}\|_F$ sense.

Compute and give $\|X - \tilde{X}\|_F / \|X\|_F$, where $\|X\|_F = \sqrt{\text{trace } X^T X}$ (Frobenius norm). How many doubles do you need to store the low-rank approximation? What is the compression ratio with respect to the original 512^2 doubles?

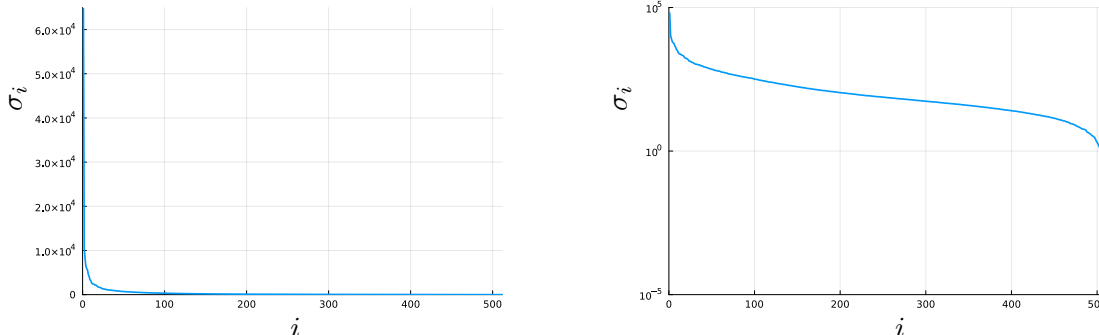
Hint: To display images in this problem, you can use the following Julia commands (using `Plots.jl`):

```
using Plots
heatmap(X, c=:grays, yflip=true,
        axis=false, aspect_ratio=:equal)
```

Solution.

- a) We take SVD of matrix X that corresponds to `lena` image ($m = n = 512$). The singular

values are plotted below (first on a linear and then on a logarithmic vertical scale).



We see that there is a large disparity between values of the first 100 or so singular values and the rest. This means that we could drop the terms in the SVD expansion after the 100th index without affecting the image very much.

- b) We take the first 75 terms in the SVD expansion to form the matrix \tilde{X} . We have

$$\tilde{X} = \sum_{i=1}^{75} \sigma_i u_i v_i^T,$$

where $u_i \in \mathbb{R}^{512}$ are the left singular vectors and $v_i \in \mathbb{R}^{512}$ are the right singular vectors. We know that the given matrix \tilde{X} is the optimal rank 75 approximator of the matrix X in the $\|X - \tilde{X}\|$ sense.

We can use the low-rank approximation as the compression scheme for the image, since in this case we only need to store

$$\#\text{doubles} = 512 \times 75 + 75 + 512 \times 75 = 76875$$

compared to 512^2 doubles in the original image. This gives us 3.41:1 compression ratio (and equivalently, $76875/512^2 = 0.293$). Here, we could actually save even more space, by only storing u_1, \dots, u_{75} and $\sigma_1 v_1, \dots, \sigma_{75} v_{75}$ (scaled right singular vectors), which gives total of $2(512 \times 75) = 76800$ doubles. In this problem, we took off a few points for mistakes such as: storing the whole Σ matrix with unnecessary zeros, storing only left singular vectors, etc.

Now let's see what is the error in the image. We compute $\|X - \tilde{X}\|_F / \|X\|_F = 0.042$, which is very small meaning that the images are quite similar. Next we show the original and the 'compressed' image for visual comparison.

The following Julia code was used to find the low-rank approximation.

```
include("readclassjson.jl")
using LinearAlgebra, Plots

data = readclassjson("lena_data.json")
X = data["X"]
```



Figure 1: *Left.* Original image. *Right.* Compressed image with 3.41:1 compression ratio.

```

m, n = size(X)

F = svd(X)
U, S, V = F.U, F.S, F.Vt

K = 75
C = U[:,1:K] * Diagonal(S[1:K]) * V[1:K,:]

error = norm(X - C) / norm(X)

```

17.940. Independent Gaussians are special.

- a) Suppose $x \in \mathbb{R}$ and $y \in \mathbb{R}$ are independent random variables, each uniformly distributed on $[-1, 1]$. Define

$$z = \begin{bmatrix} x \\ y \end{bmatrix}$$

so that z is a random variable in \mathbb{R}^2 . What is the pdf of z ?

- b) Suppose $p \in \mathbb{R}$ and $q \in \mathbb{R}$ are independent random variables, with $x \sim \mathcal{N}(0, 1)$ and $y \sim \mathcal{N}(0, 1)$. Define

$$w = \begin{bmatrix} p \\ q \end{bmatrix}$$

so that w is a random variable on \mathbb{R}^2 . What is the pdf of w ?

- c) These two random variables z and w are very different. In particular, create two different plots, one showing 10000 samples of z and the other showing 10000 samples of w .

Notice that the spattering of z has a square shape; this is what you might expect for two identically distributed random variables. But the spattering of w is circularly symmetric.

- d) Prove that w is circularly symmetric. To do this, let $T \in \mathbb{R}^{2 \times 2}$ be a matrix that rotates the plane by angle θ . Show that the random variable r defined by

$$r = Tw$$

has the same pdf as w .

Solution.

- a) The pdf of z is

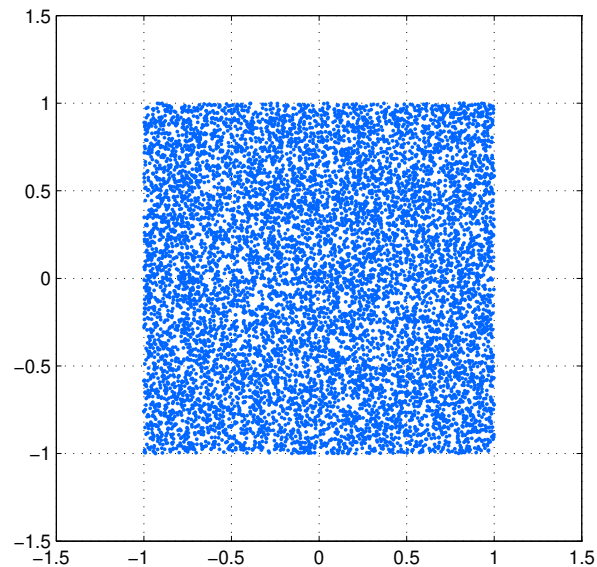
$$\begin{aligned} f_z(z) &= f_x(z_1)f_y(z_2) \\ &= \begin{cases} \frac{1}{4} & \text{if } -1 \leq z_1 \leq 1 \text{ and } -1 \leq z_2 \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

- b) The pdf of w is

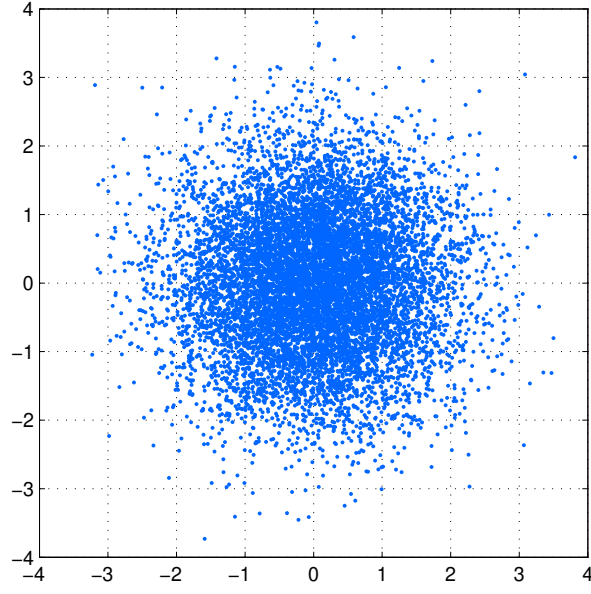
$$\begin{aligned} f_w(w) &= f_p(w_1)f_q(w_2) \\ &= \frac{1}{2\pi}e^{-\frac{1}{2}w_1^2}e^{-\frac{1}{2}w_2^2} \\ &= \frac{1}{2\pi}e^{-\frac{1}{2}w^T w} \end{aligned}$$

which means $w \sim \mathcal{N}(0, I)$.

- c) The samples of z are shown below.



The samples of w are shown below.



d) We have

$$T = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

rotates the plane \mathbb{R}^2 by angle θ . Then since $w \sim \mathcal{N}(0, I)$, we have

$$Tw \sim \mathcal{N}(0, TT^T)$$

and since $TT^T = I$ we have

$$Tw \sim \mathcal{N}(0, I)$$

which is the same as w .