EE263 Homework 8

Fall 2025

due Thursday 11/20, at 11:59 PM

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 $P = P^\mathsf{T}$ 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^T is an orthogonal projection matrix. Onto which of the four fundamental subspaces of U (range(U), range(U^T), null(U), null(U^T)) does UU^T project? What about the orthogonal projection matrix $I - UU^\mathsf{T}$?

c) Suppose $A \in \mathbb{R}^{m\times n}$ is full rank, with $n \le m$. The fitted vector for least-squares problem

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

is given by $y_{\mathrm{ls}} = A(A^\mathsf{T}A)^{-1}A^\mathsf{T}y$. Show that $P_{\mathrm{ls}} = A(A^\mathsf{T}A)^{-1}A^\mathsf{T}$ 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 ($m \le n$). The x solving

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

is given by $x_{\mathrm{ln}} = A^\mathsf{T}(AA^\mathsf{T})^{-1}y$. Show that $P_{\mathrm{ln}} = A^\mathsf{T}(AA^\mathsf{T})^{-1}A$ 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

$$A = U\Sigma V^\mathsf{T}, \qquad U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix},$$

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 (range(A), range(A^T), null(A), null(A^T)) 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

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

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

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

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

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 range(A) and its projection onto null(A^T).

h) Explain why every $x \in \mathcal{X}_{\mathrm{ls}}$ 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 $x_{\mathrm{pinv}} = A^\dagger y$ is the unique vector of minimum norm among all least-squares minimizers, i.e. x_{pinv} is the minimum-norm least-squares solution.

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

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

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 $w^\mathsf{T} a = 1$.

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 $w^\mathsf{T} y_t \approx s_t$, we would have

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

Since $w^\mathsf{T} v_t$ 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 $(1/T) \sum_{t=1}^{\mathsf{T}} (w^\mathsf{T} y_t)^2 = 1$. This doesn't determine w uniquely; we can multiply it by -1 and it still minimizes $\|w\|$ subject to $(1/T) \sum_{t=1}^{\mathsf{T}} (w^\mathsf{T} y_t)^2 = 1$. 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 $-s_t$.)

a) Explain how to find w, given the received vector signal $y_1, \dots, y_T$, using concepts from the class.

b) 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 $w^\mathsf{T} y_t$ 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

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

where $\mathrm{sign}(u)$ 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 `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.)

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

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

(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

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

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

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

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

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

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 $\|X\|_F = \sqrt{\operatorname{trace} X^T X}$ (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)
```

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.