Problem Sets Related to Lectures and Readings

LEC #	TITLE	Reading	Assignment
1	The Column Space of *A* Contains All Vectors *Ax*	Section I.1	Problem Set I.1
2	Multiplying and Factoring Matrices	Section I.2	Problem Set I.2
3	Orthonormal Columns In *Q* Give *Q'Q= I*	Section I.5	Problem Set I.5
4	Eigenvalues and Eigenvectors	Section I.6	Problem Set I.6
5	Positive Definite and Semidefinite Matrices	Section I.7	Problem Set I.7
6	Singular Value Decomposition (SVD)	Section I.8	Problem Set I.8
7	Eckart-Young: The Closest Rank *k* Matrix to *A*	Section I.9	Problem Set I.9
8	Norms of Vectors and Matrices	Section I.11	Problem Set I.11
9	Four Ways to Solve Least Squares Problems	Section II.2	Problem Set II.2 Problems 2, 8, 9
10	Survey of Difficulties with *Ax = b*	Intro Ch. 2	Problem Set II.2 Problems 12 and 17
11	Minimizing ‖*x*‖ Subject to *Ax = b*	Section I.11	Problem Set I.11 Problem 6 Problem Set II.2 Problem 10
12	Computing Eigenvalues and Singular Values	Section II.1	Problem Set II.1
13	Randomized Matrix Manipulation	Section II.4	Problem Set II.4
14	Low Rank Changes in *A* and Its Inverse	Section III.1	Problem Set III.1
15	Matrices *A(t)* depending on *t* / Derivative = *dA/dt*	Sections III.1–2	Problem Set III.2 Problems 1, 2, 5
16	Derivatives of Inverse and Singular Values	Sections III.1–2	Problem Set III.2 Problems 3, 12
17	Rapidly Decreasing Singular Values	Section III.3	Problem Set III.3
18	Counting Parameters in SVD, LU, QR, Saddle Points	Append., Sec. III.2	Problem Set III.2
19	Saddle Points Continued / Maxmin Principle	Sections III.2, V.1	Problem Set V.1 Problems 3, 8
20	Definitions and Inequalities	Sections V.1, V.3	Problem Set V.1 Problems 10. 12 Problem Set V.3 Problem 3

21	Minimizing a Function Step by Step	Sections VI.1, VI.4	Problem Set VI.1
22	Gradient Descent: Downhill to a Minimum	Section VI.4	Problem Set VI.4 Problems 1, 6
23	Accelerating Gradient Descent (Use Momentum)	Section VI.4)	Problem Set VI.4 Problem 5
24	Linear Programming and Two-Person Games	Sections VI.2–VI.3	Problem Set VI.2 Problem 1 Problem Set VI.3 Problems 2, 5
25	Stochastic Gradient Descent	Section VI.5	Problem Set VI.5
26	Structure of Neural Nets for Deep Learning	Section VII.1	Problem Set VII.1
27	Backpropagation to Find Derivative of the Learning Function	Section VII.2	Problem Set VII.2
28	Computing in Class	Section VII.2 and Appendix 3	[No Problems Assigned]
29	[No Video Recorded]	No Readings	[No Problems Assigned]
30	Completing a Rank-One Matrix / Circulants!	Sections IV.8, IV.2	Problem Set IV.8 Problem Set IV.2
31	Eigenvectors of Circulant Matrices: Fourier Matrix	Section IV.2	Problem Set IV.2
32	ImageNet is a CNN / The Convolution Rule	Section IV.2	Problem Set IV.2
33	Neural Nets and the Learning Function	Sections VII.1, IV.10	Problem Set VII.1 Problem Set IV.10
34	Distance Matrices / Procrustes Problem / First Project	Sections IV.9, IV.10	Problem Set IV.9
35	Finding Clusters in Graphs / Second Project: Handwriting	Sections IV.6–IV.7	Problem Set IV.6
36	Third Project / Alan Edelman and Julia Language	Sections III.3, VII.2	[No Problems Assigned]

Problems for Lecture 1 (from textbook Section I.1)

1 Give an example where a combination of three nonzero vectors in $\mathbf{R}^4$ is the zero vector. Then write your example in the form $A\boldsymbol{x} = \mathbf{0}$. What are the shapes of A and $\boldsymbol{x}$ and $\mathbf{0}$?

4 Suppose A is the 3 by 3 matrix **ones**$(3,3)$ of all ones. Find two independent vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ that solve $A\boldsymbol{x} = \mathbf{0}$ and $A\boldsymbol{y} = \mathbf{0}$. Write that first equation $A\boldsymbol{x} = \mathbf{0}$ (with numbers) as a combination of the columns of A. Why don't I ask for a third independent vector with $A\boldsymbol{z} = \mathbf{0}$?

9 Suppose the column space of an m by n matrix is all of $\mathbf{R}^3$. What can you say about m ? What can you say about n ? What can you say about the rank r ?

18 If $A = CR$, what are the CR factors of the matrix $\begin{bmatrix} 0 & A \\ 0 & A \end{bmatrix}$?

Problems for Lecture 2 (from textbook Section I.2)

2 Suppose $\boldsymbol{a}$ and $\boldsymbol{b}$ are column vectors with components $a_1, \ldots, a_m$ and $b_1, \ldots, b_p$. Can you multiply $\boldsymbol{a}$ times $\boldsymbol{b}^{\mathrm{T}}$ (yes or no)? What is the shape of the answer $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$? What number is in row i, column j of $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$? What can you say about $\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$?

6 If A has columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ and $B = I$ is the identity matrix, what are the rank one matrices $\boldsymbol{a}_1\boldsymbol{b}_1^*$ and $\boldsymbol{a}_2\boldsymbol{b}_2^*$ and $\boldsymbol{a}_3\boldsymbol{b}_3^*$? They should add to $AI = A$.

Problems for Lecture 3 (from textbook Section I.5)

2 Draw unit vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ that are *not* orthogonal. Show that $\boldsymbol{w} = \boldsymbol{v} - \boldsymbol{u}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v})$ is orthogonal to $\boldsymbol{u}$ (and add $\boldsymbol{w}$ to your picture).

4 Key property of every orthogonal matrix: $||Q\boldsymbol{x}||^2 = ||\boldsymbol{x}||^2$ for every vector $\boldsymbol{x}$. More than this, show that $(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ for every vector $\boldsymbol{x}$ and $\boldsymbol{y}$. So *lengths and angles are not changed by* Q. **Computations with Q never overflow!**

6 A **permutation matrix** has the same columns as the identity matrix (in some order). *Explain why this permutation matrix and every permutation matrix is orthogonal*:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \text{ has orthonormal columns so } P^{\mathrm{T}}P = \underline{\qquad} \text{ and } P^{-1} = \underline{\qquad}.$$

When a matrix is symmetric or orthogonal, **it will have orthogonal eigenvectors**. This is the most important source of orthogonal vectors in applied mathematics.

Problems for Lecture 4 (from textbook Section I.6)

2 Compute the eigenvalues and eigenvectors of A and A^{-1}. Check the trace !

$$A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad A^{-1} = \begin{bmatrix} -1/2 & 1 \\ 1/2 & 0 \end{bmatrix}.$$

A^{-1} has the _____ eigenvectors as A. When A has eigenvalues λ_1 and λ_2, its inverse has eigenvalues _____ .

11 ***The eigenvalues of A equal the eigenvalues of A^{T}.*** This is because $\det(A - \lambda I)$ equals $\det(A^{\mathrm{T}} - \lambda I)$. That is true because _____. Show by an example that the eigenvectors of A and A^{T} are *not* the same.

15 (a) Factor these two matrices into $A = X\Lambda X^{-1}$:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}.$$

(b) If $A = X\Lambda X^{-1}$ then $A^3 = (\quad)(\quad)(\quad)$ and $A^{-1} = (\quad)(\quad)(\quad)$.

Problems for Lecture 5 (from textbook Section I.7)

3 For which numbers b and c are these matrices positive definite?

$$S = \begin{bmatrix} 1 & b \\ b & 9 \end{bmatrix} \qquad S = \begin{bmatrix} 2 & 4 \\ 4 & c \end{bmatrix} \qquad S = \begin{bmatrix} c & b \\ b & c \end{bmatrix}.$$

With the pivots in D and multiplier in L, factor each A into LDL^{T}.

14 Find the 3 by 3 matrix S and its pivots, rank, eigenvalues, and determinant:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} & & \\ & S & \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 4(x_1 - x_2 + 2x_3)^2.$$

15 Compute the three upper left determinants of S to establish positive definiteness. Verify that their ratios give the second and third pivots.

$$\textbf{Pivots} = \textbf{ratios of determinants} \qquad S = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 5 & 3 \\ 0 & 3 & 8 \end{bmatrix}.$$

Problems for Lecture 6 (from textbook Section I.8)

1 A symmetric matrix $S = S^{\mathrm{T}}$ has orthonormal eigenvectors $\boldsymbol{v}_1$ to $\boldsymbol{v}_n$. Then any vector $\boldsymbol{x}$ can be written as a combination $\boldsymbol{x} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$. Explain these two formulas:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = c_1^2 + \cdots + c_n^2 \qquad\qquad \boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2.$$

6 Find the σ's and $\boldsymbol{v}$'s and $\boldsymbol{u}$'s in the SVD for $A = \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix}$. Use equation (12).

Problems for Lecture 7 (from textbook Section I.9)

2 Find a closest rank-1 approximation to these matrices (L^2 or Frobenius norm):

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

10 If A is a 2 by 2 matrix with $\sigma_1 \geq \sigma_2 > 0$, find $||A^{-1}||_2$ and $||A^{-1}||_F^2$.

Problems for Lecture 8 (from textbook Section I.11)

1 Show directly this fact about ℓ^1 and ℓ^2 and ℓ^∞ vector norms: $||\boldsymbol{v}||_2^2 \leq ||\boldsymbol{v}||_1 \, ||\boldsymbol{v}||_\infty$.

7 A short proof of $||AB||_F \leq ||A||_F \, ||B||_F$ starts from multiplying rows times columns:

$|(AB)_{ij}|^2 \leq ||\text{row } i \text{ of } A||^2 \, ||\text{column } j \text{ of } B||^2$ is the Cauchy-Schwarz inequality

Add up both sides over all i and j to show that $||AB||_F^2 \leq ||A||_F^2 \, ||B||_F^2$.

Problems for Lecture 9 (from textbook Section II.2)

2 Why do A and A^+ have the same rank? If A is square, do A and A^+ have the same eigenvectors? What are the eigenvalues of A^+?

8 What multiple of $\boldsymbol{a} = \left[\begin{smallmatrix} \mathbf{1} \\ \mathbf{1} \end{smallmatrix}\right]$ should be subtracted from $\boldsymbol{b} = \left[\begin{smallmatrix} \mathbf{4} \\ \mathbf{0} \end{smallmatrix}\right]$ to make the result $\boldsymbol{A_2}$ orthogonal to $\boldsymbol{a}$? Sketch a figure to show $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{A_2}$.

9 Complete the Gram-Schmidt process in Problem 8 by computing $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ and $\boldsymbol{q}_2 = \boldsymbol{A_2}/\|\boldsymbol{A_2}\|$ and factoring into QR:

$$\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix} \begin{bmatrix} \|\boldsymbol{a}\| & ? \\ 0 & \|\boldsymbol{A_2}\| \end{bmatrix}.$$

The backslash command $A\backslash\boldsymbol{b}$ is engineered to make A block diagonal when possible.

Problems for Lecture 10 (from textbook Introduction Chapter 2)

Problems 12 and 17 use four data points $\boldsymbol{b} = (0, 8, 8, 20)$ to bring out the key ideas.

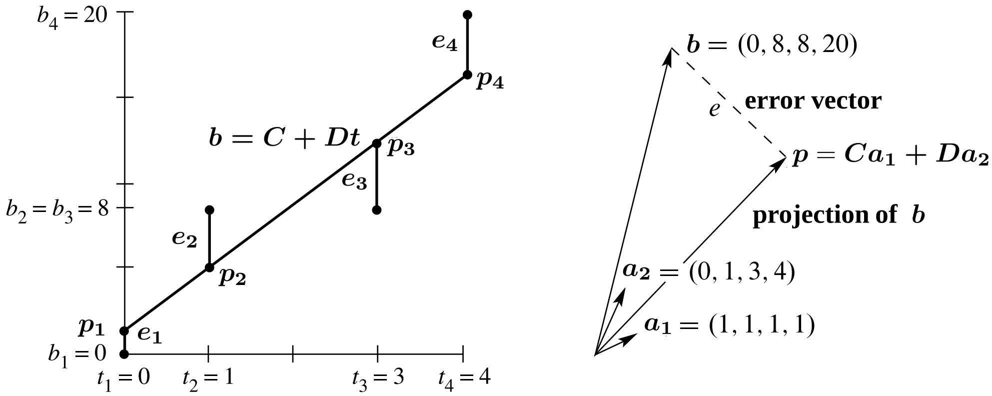

Figure II.3: The closest line $C + Dt$ in the $t - b$ plane matches $C\boldsymbol{a}_1 + D\boldsymbol{a}_2$ in $\mathbf{R^4}$.

12 With $b = 0, 8, 8, 20$ at $t = 0, 1, 3, 4$, set up and solve the normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. For the best straight line in Figure II.3a, find its four heights p_i and four errors e_i. What is the minimum squared error $E = e_1^2 + e_2^2 + e_3^2 + e_4^2$?

17 Project $\boldsymbol{b} = (0, 8, 8, 20)$ onto the line through $\boldsymbol{a} = (1, 1, 1, 1)$. Find $\widehat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ and the projection $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$. Check that $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is perpendicular to $\boldsymbol{a}$, and find the shortest distance $\|\boldsymbol{e}\|$ from $\boldsymbol{b}$ to the line through $\boldsymbol{a}$.

Problems for Lecture 11 (from textbook Section I.11)
Problem Set I.11

6 The first page of I.11 shows *unit balls* for the ℓ^1 and ℓ^2 and ℓ^∞ norms. Those are the three sets of vectors $\boldsymbol{v} = (v_1, v_2)$ with $||\boldsymbol{v}||_1 \leq 1, ||\boldsymbol{v}||_2 \leq 1, ||\boldsymbol{v}||_\infty \leq 1$. *Unit balls are always convex because of the triangle inequality for vector norms*:

$$\text{If } ||\boldsymbol{v}|| \leq 1 \text{ and } ||\boldsymbol{w}|| \leq 1 \text{ show that } ||\frac{\boldsymbol{v}}{2} + \frac{\boldsymbol{w}}{2}|| \leq 1.$$

Problem Set II.2

10 What multiple of $\boldsymbol{a} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}$ should be subtracted from $\boldsymbol{b} = \begin{bmatrix} \mathbf{4} \\ \mathbf{0} \end{bmatrix}$ to make the result $\boldsymbol{A_2}$ orthogonal to $\boldsymbol{a}$? Sketch a figure to show $\boldsymbol{a}, \boldsymbol{b}$, and $\boldsymbol{A_2}$.

Problem for Lecture 12 (from textbook Section II.1)

These problems start with a bidiagonal n by n backward difference matrix $D = I - S$. Two tridiagonal second difference matrices are DD^{T} and $A = -S + 2I - S^{\mathrm{T}}$. The shift S has one nonzero subdiagonal $S_{i,i-1} = 1$ for $i = 2, \ldots, n$. A has diagonals $-1, 2, -1$.

1 Show that DD^{T} equals A except that $1 \neq 2$ in their $(1,1)$ entries. Similarly $D^{\mathrm{T}}D = A$ except that $1 \neq 2$ in their (n,n) entries.

Problems for Lecture 13 (from textbook Section II.4)

1 Given positive numbers $a_1, \ldots, a_n$ find positive numbers $p_1 \ldots, p_n$ so that

$$p_1 + \cdots + p_n = 1 \quad \text{and} \quad V = \frac{a_1^2}{p_1} + \cdots + \frac{a_n^2}{p_n} \text{ reaches its minimum } (a_1 + \cdots + a_n)^2.$$

The derivatives of $L(p, \lambda) = V - \lambda(p_1 + \cdots + p_n - 1)$ are zero as in equation (8).

4 If $M = \mathbf{1}\,\mathbf{1}^{\mathrm{T}}$ is the n by n matrix of 1's, prove that $nI - M$ is positive semidefinite. Problem 3 was the energy test. For Problem 4, find the eigenvalues of $nI - M$.

Problems for Lecture 14 (from textbook Section III.1)

1 Another approach to $(I - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})^{-1}$ starts with the formula for a geometric series: $(1-x)^{-1} = 1 + x + x^2 + x^3 + \cdots$ Apply that formula when $x = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} =$ matrix:

$$\begin{aligned}(I - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})^{-1} &= I + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \cdots \\ &= I + \boldsymbol{u}\left[1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u} + \cdots\right]\boldsymbol{v}^{\mathrm{T}}.\end{aligned}$$

Take $x = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}$ to see $I + \dfrac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}}{1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}}$. This is exactly equation (1) for $(I - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})^{-1}$.

4 Problem 3 found the inverse matrix $M^{-1} = (A - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})^{-1}$. In solving the equation $M\boldsymbol{y} = \boldsymbol{b}$, we compute **only the solution $\boldsymbol{y}$** and not the whole inverse matrix M^{-1}. You can find $\boldsymbol{y}$ in two easy steps:

Step 1 Solve $A\boldsymbol{x} = \boldsymbol{b}$ and $A\boldsymbol{z} = \boldsymbol{u}$. Compute $D = 1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{z}$.

Step 2 Then $\boldsymbol{y} = \boldsymbol{x} + \dfrac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{D}\boldsymbol{z}$ is the solution to $M\boldsymbol{y} = (A - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})\boldsymbol{y} = \boldsymbol{b}$.

Verify $(A - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})\boldsymbol{y} = \boldsymbol{b}$. **We solved two equations using $\boldsymbol{A}$, no equations using $\boldsymbol{M}$.**

Problems for Lecture 15 (from textbook Sections III.1-III.2)

1 A unit vector $\boldsymbol{u}(t)$ describes a point moving around on the unit sphere $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$. Show that the velocity vector $d\boldsymbol{u}/dt$ is orthogonal to the position: $\boldsymbol{u}^{\mathrm{T}}(d\boldsymbol{u}/dt) = 0$.

2 Suppose you add a positive semidefinite **rank two** matrix to S. What interlacing inequalities will connect the eigenvalues λ of S and α of $S + \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} + \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$?

5 Find the eigenvalues of A_3 and A_2 and A_1. Show that they are interlacing:

$$A_3 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \qquad A_1 = \begin{bmatrix} 1 \end{bmatrix}$$

Problems for Lecture 16 (from textbook Sections III.1-III.2)

3 (a) Find the eigenvalues $\lambda_1(t)$ and $\lambda_2(t)$ of $A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} + t \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

(b) At $t = 0$, find the eigenvectors of $A(0)$ and verify $\dfrac{d\lambda}{dt} = \boldsymbol{y}^{\mathrm{T}} \dfrac{dA}{dt} \boldsymbol{x}$.

(c) Check that the change $A(t) - A(0)$ is positive semidefinite for $t > 0$. Then verify the interlacing law $\lambda_1(t) \geq \lambda_1(0) \geq \lambda_2(t) \geq \lambda_2(0)$.

12 If $\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ and C is invertible, why is $(C\boldsymbol{y})^{\mathrm{T}} S (C\boldsymbol{y})$ also positive? This shows again that if S has all positive eigenvalues, so does $C^{\mathrm{T}} S C$.

Problems for Lecture 17 (from textbook Section III.3)

2 Show that the evil **Hilbert matrix** H passes the Sylvester test $AH - HB = C$

$$H_{ij} = \frac{1}{i+j-1} \qquad A = \frac{1}{2}\text{diag}\,(1, 3, \ldots, 2n-1) \qquad B = -A \qquad C = \textbf{ones}(n)$$

6 If an invertible matrix X satisfies the Sylvester equation $AX - XB = C$, find a Sylvester equation for X^{-1}.

Problems for Lecture 18 (from textbook Section III.2)

4 S is a symmetric matrix with eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_n$ and eigenvectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$. Which i of those eigenvectors are a basis for an i-dimensional subspace Y with this property: The minimum of $\boldsymbol{x}^{\mathrm{T}} S\boldsymbol{x}/\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ for $\boldsymbol{x}$ in Y is λ_i.

10 Show that this $2n \times 2n$ KKT matrix H has n positive and n negative eigenvalues:

$$\begin{array}{c}\boldsymbol{S}\ \textbf{positive definite}\\ \boldsymbol{C}\ \textbf{invertible}\end{array} \qquad \boldsymbol{H} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{C} \\ \boldsymbol{C}^{\mathbf{T}} & \mathbf{0} \end{bmatrix}$$

The first n pivots from S are positive. The last n pivots come from $-C^{\mathrm{T}}S^{-1}C$.

Problems for Lecture 19 (from textbook Sections III.2 and V.1)

3 We know: $\frac{1}{3}$ of all integers are divisible by 3 and $\frac{1}{7}$ of integers are divisible by 7. What fraction of integers will be divisible by 3 or 7 or both?

8 Equation (4) gave a second equivalent form for S^2 (the variance using samples):

$$\boldsymbol{S^2} = \frac{1}{N-1}\text{ sum of } (x_i - m)^2 = \frac{1}{N-1}\left[\left(\text{sum of } x_i^2\right) - Nm^2\right].$$

Verify the matching identity for the expected variance σ^2 (using $m = \Sigma\, p_i\, x_i$):

$$\boldsymbol{\sigma^2 = \textbf{ sum of } p_i\,(x_i - m)^2 = \left(\textbf{sum of } p_i\, x_i^2\right) - m^2}.$$

Problems for Lecture 20 (from textbook Section V.1)

10 Computer experiment: Find the average $A_{1000000}$ of a million random 0-1 samples! What is your value of the standardized variable $X = \left(A_N - \frac{1}{2}\right)/2\sqrt{N}$?

12 For any function $f(x)$ the expected value is $\mathrm{E}[f] = \sum p_i\, f(x_i)$ or $\int p(x)\, f(x)\, dx$ (discrete or continuous probability). The function can be x or $(x-m)^2$ or x^2.

If the mean is $\mathrm{E}[x] = m$ and the variance is $\mathrm{E}[(x-m)^2] = \sigma^2$ **what is E$[\boldsymbol{x^2}]$** ?

Problem for Lecture 20 (from textbook Section V.3)

3 A fair coin flip has outcomes $X = 0$ and $X = 1$ with probabilities $\frac{1}{2}$ and $\frac{1}{2}$. What is the probability that $X \geq 2\overline{X}$? Show that Markov's inequality gives the exact probability $\overline{X}/2$ in this case.

Problems for Lecture 21 (from textbook Sections VI.1 and VI.4)

1 When is the union of two circular discs a convex set ? Or two squares ?

5 Suppose K is convex and $F(x) = 1$ for x in K and $F(x) = 0$ for x not in K. Is F a convex function ? What if the 0 and 1 are reversed ?

Problems for Lecture 22 (from textbook Section VI.4)

1 For a 1 by 1 matrix in Example 3, the determinant is just $\det X = x_{11}$. Find the first and second derivatives of $F(X) = -\log(\det X) = -\log x_{11}$ for $x_{11} > 0$. Sketch the graph of $F = -\log x$ to see that this function F is convex.

6 What is the gradient descent equation $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - s_k \nabla f(\boldsymbol{x}_k)$ for the least squares problem of minimizing $f(\boldsymbol{x}) = \frac{1}{2}||A\boldsymbol{x} - \boldsymbol{b}||^2$?

Problem for Lecture 23 (from textbook Section VI.4)

5 Explain why projection onto a convex set K is a *contraction* in equation (24). Why is the distance $||\boldsymbol{x} - \boldsymbol{y}||$ never increased when $\boldsymbol{x}$ and $\boldsymbol{y}$ are projected onto K ?

Problem for Lecture 24 (from textbook Section VI.2)

1 Minimize $F(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x} = \frac{1}{2}x_1^2 + 2x_2^2$ subject to $A^{\mathrm{T}}\boldsymbol{x} = x_1 + 3x_2 = b$.

(a) What is the Lagrangian $L(\boldsymbol{x}, \lambda)$ for this problem ?

(b) What are the three equations "derivative of $L =$ zero" ?

(c) Solve those equations to find $\boldsymbol{x}^* = (x_1^*, x_2^*)$ and the multiplier λ^*.

(d) Draw Figure VI.4 for this problem with constraint line tangent to cost circle.

(e) Verify that the derivative of the minimum cost is $\partial F^* / \partial b = -\lambda^*$.

Problems for Lecture 24 (from textbook Section VI.3)

2 Suppose the constraints are $x_1 + x_2 + 2x_3 = 4$ and $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$. Find the three corners of this triangle in $\mathbf{R}^3$. Which corner minimizes the cost $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = 5x_1 + 3x_2 + 8x_3$?

5 Find the optimal (minimizing) strategy for X to choose rows. Find the optimal (maximizing) strategy for Y to choose columns. What is the payoff from X to Y at this optimal minimax point $\boldsymbol{x}^*, \boldsymbol{y}^*$?

Payoff matrices $\begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix} \qquad \begin{bmatrix} 1 & 4 \\ 8 & 2 \end{bmatrix}$

Problem for Lecture 25 (from textbook Section VI.5)

1 Suppose we want to minimize $F(x,y) = y^2 + (y-x)^2$. The actual minimum is $F = 0$ at $(x^*, y^*) = (0,0)$. Find the gradient vector $\boldsymbol{\nabla F}$ at the starting point $(x_0, y_0) = (1,1)$. For full gradient descent (*not stochastic*) with step $s = \frac{1}{2}$, where is (x_1, y_1) ?

Problems for Lecture 26 (from textbook Section VII.1)

4 Explain with words or show with graphs why each of these statements about Continuous Piecewise Linear functions (CPL functions) is true :

- $\boldsymbol{M}$ The maximum $M(x, y)$ of two CPL functions $F_1(x, y)$ and $F_2(x, y)$ is CPL.
- $\boldsymbol{S}$ The sum $S(x, y)$ of two CPL functions $F_1(x, y)$ and $F_2(x, y)$ is CPL.
- $\boldsymbol{C}$ If the one-variable functions $y = F_1(x)$ and $z = F_2(y)$ are CPL, so is the composition $C(x) = z = (F_2(F_1(x))$.

Problem 7 uses the blue ball, orange ring example on playground.tensorflow.org with one hidden layer and activation by ReLU (not Tanh). When learning succeeds, a white polygon separates blue from orange in the figure that follows.

7 Does learning succeed for $N = 4$? What is the count $r(N, 2)$ of flat pieces in $F(\boldsymbol{x})$? The white polygon shows where flat pieces in the graph of $F(\boldsymbol{x})$ change sign as they go through the base plane $z = 0$. How many sides in the polygon ?

Problems for Lecture 27 (from textbook Section VII.2)

2 If A is an m by n matrix with $m > n$, is it faster to multiply $A(A^{\mathrm{T}}A)$ or $(AA^{\mathrm{T}})A$?

5 Draw a computational graph to compute the function $f(x, y) = x^3(x - y)$. Use the graph to compute $f(2, 3)$.

Problem for Lecture 30 (from textbook Section IV.8)

3 For a connected graph with M edges and N nodes, what requirement on M and N comes from each of the words *spanning tree*?

Problem for Lecture 30 (from textbook Section IV.2)

1 Find $\boldsymbol{c} * \boldsymbol{d}$ and $\boldsymbol{c} \mathbin{\boldsymbol{\ast}} \boldsymbol{d}$ for $\boldsymbol{c} = (2, 1, 3)$ and $\boldsymbol{d} = (3, 1, 2)$.

Problems for Lecture 31 (from textbook Section IV.2)

3 If $\boldsymbol{c} * \boldsymbol{d} = \boldsymbol{e}$, why is $(\sum c_i)(\sum d_i) = (\sum e_i)$? Why was our check successful? $(1+2+3)\,(5+0+4) = (\mathbf{6})\,(\mathbf{9}) = \mathbf{54} = 5+10+19+8+12$.

5 What are the eigenvalues of the 4 by 4 circulant $C = I + P + P^2 + P^3$? Connect those eigenvalues to the discrete transform $F\boldsymbol{c}$ for $\boldsymbol{c} = (1,1,1,1)$. For which three real or complex numbers z is $1 + z + z^2 + z^3 = 0$?

Problem for Lecture 32 (from textbook Section IV.2)

4 Any two circulant matrices of the same size commute: $CD = DC$. They have the same eigenvectors $\boldsymbol{q}_k$ (the columns of the Fourier matrix F). Show that the eigenvalues $\lambda_k(CD)$ are equal to $\lambda_k(C)$ times $\lambda_k(D)$.

Problem for Lecture 33 (from textbook Section VII.1)

5 How many weights and biases are in a network with $m = N_0 = 4$ inputs in each feature vector $\boldsymbol{v}_0$ and $N = 6$ neurons on each of the 3 hidden layers? How many activation functions (ReLU) are in this network, before the final output?

Problem for Lecture 33 (from textbook Section IV.10)

2 $||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2 = 9$ and $||\boldsymbol{x}_2 - \boldsymbol{x}_3||^2 = 16$ and $||\boldsymbol{x}_1 - \boldsymbol{x}_3||^2 = 25$ do satisfy the triangle inequality $3 + 4 > 5$. Construct G and find points $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ that match these distances.

Problem for Lecture 34 (from textbook Sections IV.9 and IV.10)

1 Which orthogonal matrix Q minimizes $||X - YQ||_F^2$? Use the solution $Q = UV^{\mathrm{T}}$ above and also minimize as a function of θ (set the θ-derivative to zero) :

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Problem for Lecture 35 (from textbook Sections IV.6-IV.7)

1 What are the Laplacian matrices for a triangle graph and a square graph ?