1 Chapter 1: Introduction

1.1 Problems

Exercise 1.2.1. Using the hint: $(x+y)^t(x+y) = |x|_2^2 + |y|_2^2 + 2x^ty \leq |x|_2^2 + |y|_2^2 + 2|x||y|$ by the Cauchy-Schwarz inequality. Therefore, $|x+y|_2 \leq |x|_2 + |y|_2$
Exercise 1.2.2. Let $(v_1x_j, v_2x_j, ..., v_dx_j)$ be the j-th row and $(v_1x_i, v_2x_i, ..., v_dx_i)$ be the i-th row of our matirx. Then, if $x_j \neq 0$ take $x_i = \frac{x_i}{x_j} * x_j$ and i-th row will be the multiple of the j-th row. The same way if $x_i \neq 0$ we can get j-th row as a multiple of the i-th row. If $x_i = 0, x_j = 0$, then they are trivially multiples of each other.
Let $(v_jx_1, v_jx_2, ...v_jx_n)$ be the j-th column and $(v_ix_1, v_ix_2, ...v_ix_n)$ be the i-th column of our matrix. Then, we can apply the same logic as above to the v_i, v_j.
Exercise 1.2.3. $\sum_i \sum_r A_iB_rC_r$???
Exercise 1.2.4. $(A^TA)A^T$ is better to compute since $shape(A^TA) = 2 \times 2$, and $shape(AA^T) = 1000000 \times 1000000$.
Exercise 1.2.5. $(DA)_{ij} = \sum_r d_{ir}a_{rj}$. So, the sum of the j-th column components of DA will be $D_{1j} + D_{2j} + ... + D_{nj} = \sum_r d_{1r}a_{rj} + \sum_r d_{2r}a_{rj} + ... + \sum_r d_{nr}a_{rj} = \sum_r (d_{1r}a_{rj} + d_{2r}a_{rj} + ... + d_{nr}a_{rj}) = \sum_r (d_{1r} + d_{2r} + ... + d_{nr})a_{rj} = 0$

1.2 Exercises

Exercise 1. If the vectors are orthogonal then their dot-product is equal to 0 (since $cos\theta = 0$).
(i) $(x-y)^t(x+y) = x^tx + x^ty - y^tx - y^ty = x^tx - y^ty = |x|_2^2 - |y|_2^2 = a^2 - a^2 = 0$.
(ii) $(x-3y)^t(x+3y) = x^tx - 3y^ty = a^2(1-3) = -2a^2$ is negative. **Exercise 2.** $dim(A) = 10 \times 2, dim(B) = 2 \times 10, dim(C) = 10 \times 10$. a) $dim(AB) = 10 \times 10, dim(BC) = 2 \times 10$. So, $A(BC)$ is more efficient to compute since instead of temporary 10 by 10 matrix we will hold in memory 2×10 matrix.
b) $dim(CA) = 10 \times 2, dim(AB) = 10 \times 10$. The very same logic applies.
Exercise 3. $A = -A^T$. Suppose there exists some diagonal element $a_{jj} \neq 0$. Then $a_{jj} = -a_{jj}$ since while taking the transpose of the matrix the diagonal elements remain in their respective positions. Therefore, we get a contradiction.
Exercise 4. Since $A = -A^T$ then $x^tAx = -x^tATx$ which means that $x^tAx = 0$, since x^tAx is a scalar.
Exercise 5.? $A = D^T$ for some $n \times d$ matrix D. First of all, two matrices are considered as equal if they have the same number of rows, the same number of columns, and element-wise they are equal to each other. So, the only possibility for d is $d = n$. Since $D^T = A, D = A^T$. $x^TAx = x^TD^Tx = (Dx)^tx = (A^tx)^tx$????????????
Exercise 6. $AB_{ij} = \sum_r a_{ir}b_{rj} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{in}b_{nj}$. Suppose, that we scaled by $c \in R \setminus 0$ the i-th column and row respectively. Then, $ca_{ik} * 1/cb_{k,j} = a_{ik} * b_{k,j}$. Therefore, the sum did not change and the ij-term remains the same.
Exercise 7. https://math.stackexchange.com/questions/3993988/show-that-any-matrix-product-ab-can-be-expressed-in-the-form-u-lambda-v-wher/39940573994057

Exercise 8. If the permutation matrix P has a dimension $d \times d$, then by at most d row interchange operations we can get the identity matrix. 1-interchange to bring the row with 1 at the first column to the first row. Second interchange to bring the row with 1 at the second column to the second row and so on.
Exercise 9. A^{-1} is the inverse of A. Let the reordered matrix be PA. Then $(PA)^{-1} = A^{-1}P^{-1} = A^{-1}P^{T}$.
Exercise 10. I assume, by multiplying U by scalar c and V by scalar $1/c$. So, $D = UV^T = cU1/cV^T$.
Exercise 11. a) The order of multiplication matters. Let's multiply the identity matrix I by an elementary matrix B that scales the second row of I by c. Then multiply by the second elementary matrix A that sums the second row to the first one. As a result we will get $(1, 0, 0..., c, ..., 0)$ in the first row. If we interchange the order of matrices and apply A, then B we will get in the first row of the resulting matrix $(c, 0, 0..., c, ..., 0)$.
Exercise 12. For any permutation matrix there can be only a finite number of permutations to get the identity matrix (exercise 8). Therefore, there are always some i, j such that $P^i = P^j = I$. So, $P^{i-j} = P^iP^{-j} = P^jP^{-j} = I$.
Exercise 13. ?
Exercise 14. If it exists then, $a_1a_1 + a_2a_3 = 0$ $a_1a_2 + a_2a_4 = 1$ $a_3a_1 + a_4a_3 = 0$ $a_3a_2 + a_4a_4 = 0$

$a_3(a_1+a_4) = 0$ $a_3 = 0$ or $(a_1+a_4) = 0$ $a_2(a_1+a_4) = 0$ $a_2 = 0$ or $(a_1+a_4) = 0$

If $a_3 = 0$ or $a_2 = 0$, then $a_1 = 0$ from the first equation and $a_4 = 0$ from the fourth equation. Therefore, second equation can not hold.
If $(a_1 + a_4) = 0$. Then of course, the second equation can not hold.
Exercise 15. By the sum of the vectors diagonals of parallelogram are equal to $x + y$ and $x - y$

$2(|x|^2 + |y|^2) = |x+y|^2 + |x-y|^2 = |x|^2 + |y|^2 + 2|x||y| + |x|^2 + |y|^2 - 2|x||y|$
Exercise 16. (i) $f(x) = ln(x), df(x)/dx = 1/x, df(x)/dx^2 = -1/x^2, df(x)/dx^3 = 2/x^3, df(x)/dx^4 = -6/x^4$. So, the Taylor expansion will be

$$f(x) = \ln a + \frac{(x-a)}{x} - \frac{(x-a)^2}{2x^2} + \frac{(x-a)^3}{3x^3} - \frac{(x-a)^4}{4x^4}$$

(ii) $f(x) = \sin x, df(x)/dx = \cos x, df(x)/dx^2 = -\sin x, df(x)/dx^3 = -\cos x, df(x)/dx^4 = \sin x$. Taylor expansion will be

$$f(x) = \sin a + \frac{(x-a)\cos x}{1} - \frac{(x-a)^2 \sin x}{2} - \frac{(x-a)^3 \cos x}{6} + \frac{(x-a)^4 \sin x}{24}$$

Exercise 17. $f(x, y) = sin(x + y), df(x, y)/dx = cos(x + y), df(x, y)/dy = cos(x+y), df(x, y)/dx^2 = -sin(x+y), df(x, y)/dy^2 = -sin(x+y), df(xy)dxdy = -sin(x + y)$ The Taylor expansion will be:

$$f(x, y) = sin(0) + (x-a)(y-a)G + (x-a)^2H(y-a)^2$$

Exercise 18.?

Exercise 19. https://math.stackexchange.com/questions/2961686/approximating-the-inverse-of-a-perturbed-matrix

Exercise 20. $P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

Exercise 21. $8 + 18 = 16 + x^2, |x + y|_2^2 = 10$ by the Parallelogram law.

Exercise 22. $I = (AA^{-1)}) = (AA^{-1)})^T = A^{-1T}A^T = A^{-1T}A$. $I = (A^{-1)}A) = (A^{-1)}A)^T = A^T(A^{-1})^T = A(A^{-1})^T$. Therefore, $(A^{-1})^T$ is the inverse of A by definition. However, since inverse is unique then, $(A^{-1})^T = A^{-1}$.

Exercise 23. Let's multiply two such matrices AB. Then, the first column will remain a column of 0's. Moreover, 12-th element will become 0 since by matrix column multiplication it is $B_{12} * 0 + B_{22} * A_{:,2} + B_{32} * A_{:,3} + ... = 0$ since $B_{i2} = 0$ where $i \geq 2$ because of the nature of the matrix. So, the second column will be column of 0's.

Exercise 24. $AB + BD = AD, BD = AD - AB \; DC + AD = AC, DC = AD - AC$

$AC^2 = AD^2 + BC^2/4 - ADBC \cos b$ so, $\cos b = \frac{AD^2+BC^2/4-AC^2}{ADBC}$

$AB^2 = AD^2 + BC^2/4 - ADBC \cos a$ so, $cosa = \frac{AD^2+BC^2/4-AB^2}{ADBC}$

$\frac{AD^2+BC^2/4-AB^2}{ADBC} = -\frac{AD^2+BC^2/4-AC^2}{ADBC}$. After simplifaction we get the desired result.

Exercise 25.? $\cos x = \frac{a^t b}{|a||b|}$ $\sin^2(x) + \frac{b^t a a^t b}{|a|^2|b|^2} = 1, \sin x = +- \sqrt{1 - \frac{b^t a a^t b}{|a|^2|b|^2}}$

$\sin A = \sqrt{1 - frac^t bb^t v|v|^2|b|^2}$

Exercise 26.?

Exercise 27. $A = \begin{pmatrix} -3 & 1 \\ -5 & 1 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$

Exercise 28. $A = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \end{pmatrix} = I_{10\times 10} + 0111111111$

1011111111
1101111111
1110111111
1111011111
1111101111
1111110111
1111111011

1111111101
1111111110 $= I_{10\times 10} + uu^t$, where $u = (1,1,1,1,1,1,1,1,1,1)$

$A^{-1} = (I+uu^t)^{-1} = I - \frac{uu^t}{1+1}$

Exercise 29. $2x+3y-z=0$. Let $v_1 = (3,2,0), v_2 = (2,-1,1)$.
$c_1\begin{pmatrix}3\\2\\0\end{pmatrix} + c_2\begin{pmatrix}2\\-1\\1\end{pmatrix}$. The two vectors v_1, v_2 span the plane in the 3 dimensional space.

Exercise 30. Let b_{ij} be the components of B. $AB^T_{ij} = \sum_{r=1}^n a_{ir}b_{jr} = \sum_{r=1}^k a_{ir}b_{jr} + \sum_{r=k+1}^n a_{ir}b_{jr} = A_1B_1^T + A_2B_2^T$.

Exercise 31.??

Exercise 32. $|AP|_F^2 = tr(APP^TA^T) = tr(AA^T) = |A|_F^2$.

Exercise 33.$|uv^T|_F = \sqrt{\sum_{i=1}^n\sum_{j=1}^d a_{ij}^2} = \sqrt{\sum_{i=1}^n\sum_{j=1}^d (u_iv_j)^2} = \sqrt{\sum_{i=1}^n (u_i)^2 \sum_{j=1}^d (v_j)^2} = \sqrt{\sum_{i=1}^n (u_i)^2}\sqrt{\sum_{j=1}^d (v_j)^2} = |u||v|$.

Exercise 34. $|A+B|_F^2 = tr((A+B)^T(A+B)) = tr(A^TA+B^TA+A^TB+B^TB)$.

Keep in mind that $tr(A+B) = \sum_{i=1}^n (A+B)_{ii} = \sum_{i=1}^n (A)_{ii} + (B)_{ii} = tr(A)+tr(B)$. So, $|A+B|_F^2 = tr((A+B)^T(A+B)) = tr(A^TA+B^TA+A^TB+B^TB) = |A|_F^2 + |B|_F^2$.

Exercise 35. $tr((xa^t+yb)^t(xa^t+yb)^t) = tr((ax^t+by^t)(xa^t+yb^t)) = tr(ax^txa^t + ax^tyb^t + by^txa^t + by^tyb^t) = tr(ax^txa^t) + tr(by^tyb^t) = |xa^t|_F^2 + |yb^t|_F^2$.

Exercise 36. What does it mean by "merged"?

Exercise 37. ?

Exercise 38. $A_{ij} = (x_i - x_j)^T(x_i - x_j) = x_i^Tx_i - 2x_i^tx_j + x_j^tx_j = 1+1-2x_i^tx_j$. Therefore, $-A_{ij}/2 + 1 = 2x_i^tx_j$.

Exercise 39. By the observation 1.2.1 we have that for any $f, g, g(A)f(A) = f(A)g(A)$. Multiply by $f(A)^{-1}$ from both sides to get $f(A)^{-1}g(A) = g(A)f(A)^{-1}$.

Exercise 40.?

Exercise 41. $A^tf(AA^T) = A^t[a_0I + a_1AA^T + a_2(AA^T)^2 + ...] = a_0A + a_1A^TAA^T + a_2A^T(AA^T)^2 + ... = [a_0I + a_1A^TA + a_2(A^TA)^2A^T] = f(A^TA)A^T$

Prove by induction that $A^T(AA^T)^n = (A^TA)^nA^T$.
Base case: $n = 0 : A^T = A^T$.
Induction step: Assume, $A^T(AA^T)^k = (A^TA)^kA^T$ $A^T(AA^T)^{k+1} = A^T(AA^T)^k(AA^T) = (A^TA)^kA^TAA^T = (A^TA)^k(A^TA)A^T = (A^TA)^{k+1}A^T$.

Exercise 42. ?

Exercise 43. $(I+A)^{-1} = I - A + A^2 - A^3$.

Exercise 44. $A = I + U$ where U is a strictly triangular matrix which by the result on the page 14 has a property that $U^3 = 0$. Therefore, $A^{-1} = (I+U)-1 = I - U + U^2$.

Exercise 45. Let v be a matrix of 1's of dimension $d \to 1$. So, $M = vv^T$. By the Lemma 1.2.5 we have that $(I+M)^{-1} = (I+vv^T)^{-1} = I - \frac{vv^T}{2}$.

Exercise 46. $AB = BA$. $f(A) = a_0I + a_1A + a_2A^2 + ...$ $f(B) = b_0I + b_1B + b_2B^2 + ...$ $f(A)f(B) = (a_0I + a_1A + a_2A^2 + ...)(b_0I + b_1B + b_2B^2 + ...) = \sum_{i=0} a_0b_iB^i + \sum_{i=0} a_1Ab_iB^i + \sum_{i=0} a_2A^2b_iB^i + ...$

2 Chapter 2:

2.1 Problems

2.2 Exercises

Exercise 1. False. For example take the matrix $A = \begin{pmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{pmatrix}$. Then, $A^2 = I$ but $A \neq + - I$.
Exercise 2. Following the hint: If $Ax = 0$ then $A^T Ax = A^T 0 = 0$. If $A^T Ax = 0$, then $Ax \in null(A^T)$ and suppose $Ax \neq 0$ so, $Ax \in range(A)$ too. Then, $null(A^T)$ and the $range(A)$ are not disjoint, which can not be true since they are orhogonal complements. Therefore, $Ax = 0$. Therefore, they have the same dimension of the null space. Let it be k. Hence, according to the dimension of the row space of Ax and $A^T A$ is also the same and is equal to $d - k$. The same logic can be applied to A^T and AA^T. But since A and A^T have the same rank since the during the transposition column become rows i.e. number of independent columns of the A become the number of independent rows of A^T and the similar process for the rows of A and columns of A^T. Therefore, the $rank(A)$ remains unchanged.
Exercise 3. It means we rotate the matrix by $60 * 9 = 540$ degrees. That is we rotate by 360 degrees by returning to the point where we started and then rotating again by 180 degrees.
Exercise 4. B^T has the shape 10×6. According to the $Lemma 2.6.4$ the upper bound is $minrank(A), rank(B^T) = 6$. According to $Lemma 2.6.5$ the lower bound is $m + m - 10 = 2$.
Exercise 5. Row-reduction: $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & -1 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$
The basis columns are $(1, 2, 1)^T, (2, 1, 1)^T$. Gram-Schmidt: $u_1 = v_1 = (1, 2, 1)^T$ so, after normalization we get $e_1 = 1/\sqrt{6}(1, 2, 1)^T$
$u_2 = v_2 - \frac{v_2 u_1}{|u_1|_2} u_1 = (7/6, -2/3, 1/6)$.
Exercise 6. Form a matrix A where the columns are our given vectors that span the vector space. Orthogonalize it using the Gram-Schmidt process. Then, if the two sets of column vectors span the same vector space we will have second set of columns equal to 0 since they will be dependent of the first set of columns.
Exercise 7. The fact that the diagonal elements are equal to 0 was shown in exercise ?.
If $A^T = -A$ then, $x^T Ax = x^T A^T x = -x^T Ax$. Therefore, $x^T Ax = 0$. Note that $x^T Ax$ is a scalar hence the first equation holds.

Suppose $x^T Ax = 0 = x^T A^T x$. Then, $x^T (A^T - A)x = 0$. So, that $x^T (A^T + A)x = 0$.
Exercise 8. $G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 90 & 0 & -\sin 90 \\ 0 & 0 & 1 & 0 \\ 0 & \sin 90 & 0 & \cos 90 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$ Let $X =$

$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$. Then $Ax = \begin{pmatrix} x_1 \\ -x_4 \\ x_3 \\ x_2 \end{pmatrix}$

Reflect around y-axis and around $x = -y$ line. 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 I - 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1

I - 0 1 0 0 1 0 0 0 0 0 1 0 0 0 0 1

Exercise 9.??

Exercise 10. $rank(A) = 5, rank(B) = 2, rank(C) = 4$. According to the Lemma $2.6.2, 2.6.3$ $3 \leq rank(A+B) \leq 7$. Hence by the Lemma $2.6.4, 2.6.5$ the $4+4-5 = 3 \leq rank((A+B)C) \leq 4$ if $rank(A+B) \leq 4$. Otherwise, $2 \leq rank(A+B)+4-5 \leq rank((A+B)C) \leq rank(A+B) \leq 3$.

Exercise 11. Not possible since the $A_1 1 = 0$ which contradictds to the definition 2.5.2 where the column index of the leftmost non-zero entry in each row increases. $A = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \\ 1 & 2 & 1 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 1 & 2 & 1 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \end{pmatrix} \to$ $\begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & -1 \end{pmatrix}$. So, by back substitution we get $x_3 = 1, x_2 = 1, x_1 = 1$.

$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$.

Exercise 12. $u_1 = v_1 = (0,1,1)^T, e_1 = (0, 1/\sqrt{2}, 1/\sqrt{2})^T$
$u_2 = v_2 - \frac{v_2^T u_1}{|u_1|_2} u_1 = (1, -1/2, 1/2)^T, e_2 = (\sqrt{6}/3, -\sqrt{6}/6, \sqrt{6}/6)^T$
$u_3 = v_3 - \frac{v_3^T u_2}{|u_2|_2} u_2 - \frac{v_3^T u_1}{|u_1|_2} u_1 = (1/3, 1/3, -1/3)^T, e_3 = (\sqrt{3}/3, \sqrt{3}/3, -\sqrt{3}/3)$.
$Ax = QRx = b, Rx = Q^T b, x = R^{-1} Q^T b$.

Exercise 13. Let $b \in range(AB)$. Then there exists some x such that $ABx = b$. However, as it was noted in the Lemma 2.6.4 the columns of AB are a linear combination of the columns of A where the columns of B are the coefficients of that combination. Therefore, $b \in range(A)$.

Firs note that for any matrix A the given sets of the null spaces are nested i.e. $null(A) \subset null(A^2) \subset null(A^3)$.... Suppose, it is not true that $null(A^n) = null(A^{n+1})$ then the subsets are proper subsets. So, the null spaces stop growing after some point since the dimension of each null space grows at least by 1 and the $null(A^k)$ can not be more than the d, where d is the dimension of the row space. Therefore, for some $k, null(A^k) = null(A^{k+1})$.
Moreover, we have that $A^{k+1} = AA^k$ its column space is a subspace of the column space of A. Therefore, we have $...range(A^3) \subset range(A^2) \subset range(A)$. Suppose, it is not true that $range(A^n) = range(A^{n+1}$ then the subsets are proper. Therefore, the dimension of the range decays at least by 1 at each power. Therefore, by the same logic as above we have that $range(A^m) = range(A^{m+1})$ for some m. Since the $dimrange(A^n)$ can not be indefinitely big. The same applies to the column space of A^T i.e. the row space of A and the null space of

A^T i.e. the left null space. Taking the maximum of the numbers for which the given relations hold we will get that they hold simultaneously for all spaces.

Exercise 14. $\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

Therefore, both of B_2 basis vectors are dependent of the basis B_1 vectors. So, they lie in the same vector space. By the theorem 2.3.1 the dimensionality is 2. $Ax_a = x \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} * \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = v$. In the B_2 basis $x_b = (1/2, 3/2)$.

Exercise 15. $u_1 = v_1 = (3, 0, 4)^T, e_1 = (3/5, 0, 4/5)^T$.
$u_2 = v_2 - \frac{u_1^T v_2}{|u_1|_2} u_1 = (0, 1, 0)^T, e_2 = (0, 1, 0)^T$.
To detect if the vector b is in the column space of A we can project it to the column space. If the projection is the vector itself than it is in the column space. Otherwise, it is not. Here it is not.

Exercise 16.

Exercise 17. For $s = 1$ using the suggested formula and taking $A_s = q_1$ we would make exactly the same step as in the Gram-Schmidt recursive calculation i.e. $q_2 = a_2 - \frac{q_1^T a_2}{|q_1|_2} q_1$. For the $s = 2$ and A_s having as columns q_1, q_2 we would have $q_3 = a_3 - \frac{\begin{pmatrix} q_1^2 & q_1 q_2 \\ q_1 q_2 & q_2^2 \end{pmatrix}}{q_1^2 + q_2^2} a_3$.

Exercise 18. $null(A) + range(A) = 2null(A) = d, null(A) = d/2$ since $null$ and $range$ are integers then d is even.

Exercise 19.

Exercise 20. Suppose, that $|Ax| < |x|$ but $(A - I)$ is not invertible. Then, there exists some $x \neq 0$ such that $(A - I)x = 0, Ax = x$. Hence, we get a contradiction.

Exercise 21.? $|Pb| = |QQ^T b| = b^T P^T P b = b^T Q Q^T Q Q^T b = b^T Q Q^T b = |Q^T b|$, since Q is an orthogonal matrix.
$Q^T b$ is a mapping of the n-dimensional vector b to a d-dimensional row space of Q.

Exercise 22. $2rank(A) - 10 \leq rank(A^2) = rank(A * A) = 6 \leq rank(A)$. So, the least bound is $rank(A) \geq 6$ with example of $e_1, e_2, ..., e_6$ in the firs 6 columns and 0-s in the remaining columns. The upper bound is $rank(A) \leq 6/2+10/2 = 8$ with example of $0, 0$-in the first two columns and $e_1, e_2, ...e_8$ in the remaining columns.

$(I - 2qq^t) = (Q^T Q - 2qq^T)$

Exercise 23. $\sum Q_{ir} Q^{T_{rj} = \sum Q_{ir} Q_{jr} = 1}$.

Exercise 24.??????

Exercise 25. By exercise 13 we already have that $range(A^k) \subset ...range(A^3) \subset range(A^2) \subset range(A)$.

$range(A^{m+k}) = range(A^{m+k+1}$. We already have that $range(A^{m+k+1}) \subset range(A^{m+k})$. Suppose $x \in range(A^{m+k})$. Then, there exists some v such that $A^{m+k}v = x, A^m A^k v = A^m A^{k+1} v = A^{m+k+1} v = x$. Therefore, $range(A^{m+k}) \subset range(A^{m+k+1})$ for any $m \in N$.

Exercise 26. ??? $rank(B) = n-1, rank(B^k) = n-k$. First of all note that $rank(B^2) \neq rank(B)$ since then by the exercise 25 we would have that $rank(B) = rank(B^2) = n-1 = rank(B^{1+r})$ for all $r \geq 1$. So, eventually $rank(B) = rank(B^k)$ which is a contradiction. Therefore, $rank(B^2)$ reduces at least by 1 in comparison with the rank of B. However, note that at the same time it reduces at most by 1 since otherwise there would be some B^m such that $rank(B^m) = rank(B^{m+1}) = ... = rank(B) =$

Exercise 27. $a_0 v + a_1 Bv + a_2 B^2 v + ... + a_k B^{k-1} v = 0, a_0 B^{k-1} v + a_1 B^k v + a_2 B^{k+1} v + ... + a_k B^{2k-2} v = a_0 B^{k-1} v = 0$. So, $a_0 = 0$. The same way we can get that the remaining coefficients are equal to 0 too.

Exercise 28. $A^{-1} = R^{-1}Q^{-1} = R^{-1}Q^T$. So, $A^{-1}A = R^{-1}Q^T QR = R^{-1}R = I$.

Exercise 29. $(Ax-b)^T(Ax-b) = x^T A^T Ax - 2x^T A^T b + b^T b$. The derivative with respect to x gives $2A^T Ax - 2A^T b = 0, x = (A^T A)^{-1} A^T b = (R^T R)^{-1} R^T Q^T b = R^{-1} Q^T b$.

Exercise 30. ?

Exercise 31.

$|v|^2 = |v - v_r|^2 + |v_r|^2 + 2v_r^T(v - v_r) = v^T v - 2v^T v^r + 2v_r^T v_r + 2v_r^T v - 2v_r^T v_r = v^t v \geq |v_r|^2 + 2v_r^T(v - v_r)$. The last inequality comes from the fact that $|v - v_r|^2 \geq 0$. $v^t v = b^T(AA)^{-T} Ax = b^T(AA^T)^{-T} b = b^T((AA^T)^T)^{-1} b = b^T(AA^T)^{-1} b$

$(b^T(AA)^{-T}A)(A^T(AA^T)^{-1}b) = b^T(AA)^{-1}b$. Therefore, $v_r^T(v - v_r) = 0$.

Exercise 32. ?

Exercise 36. https://web.stanford.edu/class/cs205/homework/hw2$_s olutions.pdf$

Exercise 38. $x^T(P + aI)x = x^T Px + a|x|^2 = x^T P^T PX + a|x|^2 = |Px| + a|x|^2 > 0, a > 0$

Exercise 39. $R + I = 2(I - vv^T)$. Singularity: $(R + I)v = 2(I - vv^T)v = 2(v - v) = 0$. Since v is a unit normal vector.

Exercise 40. $x^T(A^T A - I)x = 0$ for any x. So, $A^T A = I$. So, A is orthogonal.

Exercise 41.?

Exercise 42. $a_1 x_1 + a_2 x_2 + ... + a_d x_d = 0$
$a_1 Ax_1 + a_2 Ax_2 + ... + a_d Ax_d = 0$
$a_1 x_1^T Ax_1 + a_2 x_1^T Ax_2 + ... + a_d x_1^T Ax_d = a_1 x_1^T Ax_1 = 0$ which implies that $a_1 = 0$. The same logic can be applied to the other coefficients to show that they are all equal to 0.

Exercise 43. $< x, y >= x^T Sy$. Note that by the previous exercise we see that $x_1, x_2, ..., x_d$ are linearly independent hence, they form the basis for $\mathbf{R^d}$.

1) $< x_i, x_j + x_k >= x_i^T S(x_j + x_k) = x_i^T Sx_j + x_i^T Sx_k =< x_i, x_j > + < x_i, x_k >$.

2) $< cx_i, x_j >= cx_i^T Sx_j = c(x_i^T Sx_j) = c < x_i, x_j >$ 3) $< x, y >\geq 0$: $< x, y >=< \sum_{i=1}^d a_i x_i, \sum_{i=j}^d b_j x_j >=< \sum_{i=1}^d a_i x_i, b_1 x_1 > + < \sum_{i=1}^d a_i x_i, b_2 x_2 > +...+ < \sum_{i=1}^d a_i x_i, b_d x_d >=< a_1 x_1, b_1 x_1 > + < a_2 x_2, b_1 x_1 > +...+ < a_d x_d, b_1 x_1 >$

$+ < a_1x_1, b_2x_2 > + < a_2x_2, b_2x_2 > +...+ < a_dx_d, b_2x_2 > +...+ < a_1x_1, b_dx_d >$ $+ < a_2x_2, b_dx_d > +...+ < a_dx_d, b_dx_d >= \sum_{i=1}^d a_ib_1x_i^TSx_1 + \sum_{i=1}^d a_ib_2x_i^TSx_2 +$ $... + \sum_{i=1}^d a_ib_dx_i^TSx_d = \sum_{j=1}^d a_jb_jx_jSx_j > 0.$
4) $< x_i, x_j >= x_i^TSx_j = 0 = x_j^TSx_i =< x_j, x_i >$ when $i \neq j$. If $i = j$ then $x_i^TSx_i$ is symmetric. 5) $< 0, 0 >= 0^TS0 = 0$.
Exercise 44. $Ax = b, (AXB-C)^T(AXB-C) = B^TX^TA^TAXB-B^TX^TA^TC-C^TAXB-C^TC$

$B^TA^T(AXB) + B^TX^TA^T(AB) - B^TA^TC = 0$

$(C - A)$

3 Chapter 3:

3.1 Problems

Exercise 3.2.1.?
Exercise 3.2.2. Keeping in mind the orthonormality of Q, P and the Lemma 3.2.2 we have: $det(A) = det(QP^T) = det(Q)det()det(P^T) = sgn(\sigma)det()$.
The absolute value of the determinant is det().
The sign of the determinant can be negative if the determinants of Q, P have the opposite signs.
If $Q = P$, then the sign of the determinant will always be positive.
Exercise 3.2.3. Let the i-th row be different then $det(aA + (1 - a)B) = \sum_j aA_{ij} + (1 - a)B_{ij}det((aA + (1 - a)B)_{ij})$. Note that the $(aA + (1 - a)B)$ without the i-th row and j-th column is the same as A without i-th row and the j-th column since the i-th row is the only place where the matrices differ. This is due to the fact that for any element of $aA + (1 - a)B)$ not in the i-th row we have $aA_{kj} + (1 - a)B_{kj} = aA_{kj} + (1 - a)A_{kj} = A_{kj}$. Hence, we have $det(aA + (1 - a)B) = \sum_j aA_{ij} + (1 - a)B_{ij}det((aA + (1 - a)B)_{ij}) = \sum_j aA_{ij} + (1 - a)B_{ij}det(A_{ij}) = a\sum_j A_{ij}det(A_{ij}) + (1 - a)\sum_j B_{ij}det(A_{ij}) = a*det(A)+(1-a)*det(B)$. The last equation is possible since $det(B_{ij}) = det(A_{ij})$ by the nature of the matrix B.
Exercise 3.2.4.
Exercise 3.2.5. Let $A = QR$, then b the Lemma 3.2.2 we have $det(A) = det(QR) = sgn(\sigma)det(R)$.
Let $A = LU, det(A) = det(LU) = det(L)det(U) = \prod_i L_{ii} \prod_j U_{jj} = \prod_i L_{ii}U_{jj}$.
Exercise 3.2.6. $A = -A^T, det(A) = det(A^T) = det(-A^T) = (-1)^d det(A^T)$. If d is odd then $det(A^T) = -det(A^T) = 0$.
Exercise 3.2.7. ?? Lets prove by induction. Base case $d = 1$: $a_{ij} \leq 1 = d^{d/2}$. Inductive step: Suppose the relation holds for any $d-1 \times d-1$ matrix. $det(A) = \sum_{j=1}^d a_{ij}det(A_{ij})$ since A_{ij} is a $(d - 1) \times d - 1$ matrix then by the inductive assumption we have $det(A) = \sum_{j=1}^d a_{ij}det(A_{ij}) \leq \sum_{j=1}^d a_{ij}(d - 1)^{(d-1)/2} \leq \sum_{j=1}^d (d - 1)^{(d-1)/2} \leq d(d)^{(d-1)/2}$
Exercise 3.3.1. Let $Av = \lambda v, (A + aI)v = Av + av = \lambda v + av = (\lambda + a)v$. Therefore, the eigenvectors remain the same. The eigenvalues of the new matrix

are the scaled versions of the previous matrix.
Exercise 3.3.2.
Exercise 3.3.3. $A = -A^T, Av = \lambda v, -A^T v = \lambda v, A^T v = -\lambda v$. So, if λ is the eigenvalue of A then $-\lambda$ is the eigenvalue of A^T.
Exercise 3.3.4. One can show it using the Cayley-Hamiltion Theorem and expressing high degree terms as a polynomial of the $d-1$ degree.
Exercise 3.3.5. $Av = \lambda v$, where $\lambda = a+bi$ for some $a, b \in R$. $(A-\lambda I)(u+iv) = (A-(a+bi)I)(u+iv) = (Au-au+bv)+i(Av-av-bu)$ and $(A-\lambda^* I)(u-iv) = (A-(a-bi)I)(u-iv) = (Au-au+bv)-i(Av-av-bu)$. Now, lets remember that a complex number is equal to zero iff both its real and imaginary parts are equal to zero. From this we can say that if λ is a eigenvalue of A with the corresponding eigenvector $u+iv$, then the conjugate λ^* should also be a conjugate of A with the corresponding eigenvector $u-iv$.
Exercise 3.3.6. From the definition of the inverse we have that $VV^{-1} = V^{-1}V = I$ and in particular $(V^{-1}V)_{ij} = \sum_m V^{-1}_{im} V_{mj} = 0$ if $i \neq j$ and 1 otherwise. Therefore, it can be concluded that the rows of V^{-1} and the columns of V are orthonormal i.e. the dot product of the left and right eigenvectors corresponding to the same eigenvalue is 1, otherwise 0.
Exercise 3.3.7. $H = I - 2uu^T, H = UAU^{-1}, R = I - 2vv^T, R = VBV^{-1}$. The householder matrices have eigenvales $1, -1$. If the dimensions of the matrices is the same than the $A = B$. Hence, we have $R = VBV^{-1} = VAV^{-1} = VU^{-1}HUV^{-1} = QHQ^{-1}$, where $Q = VU^{-1}$.
Exercise 3.3.8. ?
Exercise 3.3.9. We can always find a sequence of permutation matrices $P_1, P_2, ...$ such that they permute the rows and columns of the Givens rotation matrices by making them similar to each other. Suppose, $cos(\alpha)$ is at the positions $(i,j), (k,r)$ for matrix A, and is in position $(i^{',j^{'}),(k^{',r^{'}})}$ for the matrix B. Then, we will apply permutation matrix P_1 on the rows of B to get $cos(\alpha)$ be in $(i, j^{'}), (k, r')$. P_1 will have 1 at the i' position in the i-th row and similarly in the i-th position of the $i^{'}$ row. The same logic applies to $k, k^{'}$. Moreover, its transpose will permute the corresponding columns of B. Just what was needed. So, $A = P_1 B P_1^T$.
Exercise 3.3.10. The isomorphic graphs are the graphs that have a bijection between vertices and the adjacent vertices of first graph remain adjacent in the second graph too. From the bijection part we get that the rows/columns of graph G_B are permuted in respect with the graph G_A. From the adjacency part we get that there are no new elements or combinations of the elements in the G_B matrix in comparison with the matrix of G_A. Therefore, the matrices are related by the permutations.
Exercise 3.3.11.??
Exercise 3.3.12. $AB = V\Lambda_1 V^T V \Lambda_2 V^T$. So, AB takes the vector to the basis formed by the eigenvectors of B/A stretch it in the directions of the eigenvectors and translates it back to the original system, then once again it took the vectors to the eigenvector's basis stretch them by some other factors and translates it back.

Exercise 3.3.13. Geometrically, let A, B for fixed pair i, j and with angles α, θ respectively. Then, AB will rotate the given vector x's i, j-th coordinates first by the angle θ and then it will rotate already rotated coordinates once more with angle α. So, the overall the rotation of the i, j-th coordinates will be $\alpha + \theta$. Using similar logic one can show that BA does the same thing.
Exercise 3.3.14. $A = V\Lambda V^{-1}$. The Frobenious norm is $\sqrt{\sum_{i=1}\sum_{j=1} A_{ij}^2} = \sqrt{\sum_{i=1}\sum_{j=1}(V\Lambda V^{-1})_{ij}^2}$.
Using that $(V\Lambda V^{-1})_{ij} = \sum_m \lambda_m V_{im} V_{mj}^{-1}$ we can get the squared Frobenious norm as $\sum_{i=1}\sum_{j=1} A_{ij}^2 = \sum_{i=1}\sum_{j=1}(V\Lambda V^{-1})_{ij}^2 = \lambda_m V_{im} V_{mj}^{-1}$

3.2 Exercises

Exercise 1. ?
Exercise 2. ? I believe, the even number of reflections will result in a positive sign of the determinant and the odd number of reflections will result in a negative sign of a determinant. However, I am not able to show it mathematically yet.
Exercise 3. $A^2 = 4I, A^2 - 4I, (A^2 - 4I)v = 0$. Therefore, 4 is an eigenvalue of A^2. On the other hand all eigenvalue of I are 1 therefore, all eigenvalues of $4I$ are 4 so, all eigenvalues are 4. Let λ be an eigenvalue of A, then $Av = \lambda v$ for some v. $AAv = A^2v = \lambda Av = \lambda^2 v$. Therefore, the eigenvalues of A^2 are the squares of the eigenvalues of A. So, $\lambda = +-2$.
Exercise 4. In the second method use the fact that if two matrices commute then they have the same set of eigenvectors and so, the eigenvalues of the sum is the sum of eigenvalues. https://math.stackexchange.com/questions/4020409/given-some-of-the-eigenvalues-and-corresponding-eigenvectors-reconstruct-the-ent/40204414020441
Exercise 5. The eigenvalues will remain the same. To show this observe the characteristic equations of the two matrices A, A'. $det(A - \lambda I) = 0$ and $det(A' - \lambda I) = det(PAP^T - \lambda I) = det(P)det(A - \lambda I)det(P^T) = det(A - \lambda I)$. Here we used the fact that $det(AB) = det(A)det(B)$. The given operations can are elementary operations and can be represented by elementary matrices P, P^T and $PP^T = P^TP = I$.
Exercise 6. The eigenvectors corresponding to the distinct eigenvalues are linearly independent.
a) Yes. b) No c) Yes d) No e) No. The matrices with repeated eigenvectors will be diagonalizable only if the algebraic and geometric multiplicities of the repeated eigenvalues will be equal.
Exercise 7. Let A be an odd dimension matrix. Suppose, it does not have real-valued eigevenvalues. Then, we would get a contradiction since in the problem 4 we already showed that complex eigenvalues as well as complex eigenvectors come in pairs. Therefore, the number of eigenvalues will be even.
The determinant of a matrix is equal to the product of the eigenvalues. Since all the eigenvalues are complex and are in pairs we would get $det(A) = (a_1a_1^*)(a_2a_2^*)...(a_na_n^*)$. Where a^* is the complex conjugate and $(xx^*) = (a + bi)(a - bi) = a^2 - abi + abi + b^2 = a^2 + b^2$. Therefore, the product of a complex number with its conju-

gate is a non-negative real number. So, the determinant being the product of a non-negative real numbers is non-negative.
If $det(A) < 0$ should have at least two distinct real eigenvalues. Otherwise the determinant will be non-negative if the eigenvalues are all complex as it was shown. It can not have one real value since the complex eigenvalues come in pairs. Hence, there are at least two distinct (because if they are the same the determinant can not be negative) real eigenvalues.
Exercise 8.?
Exercise 9.? $A = VUV^{-1}, AA = VUV^{-1}VUV^{-1} = VU^2V^{-1}$
Exercise 10. $tr(AB - BA) = tr(AB) - tr(BA) = 0 = \sum_i \lambda_i$. Where λ_i are the eigenvalues of $AB - BA$. If $AB - BA$ is positive semidefinite then it is symmetric and all its eigenvalues are 0. Therefore, the only way is 0 matrix.
Exercise 11. $\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}$

its square is $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$
Exercise 12. $AB = (AB)^T = B^T A^T = BA$. So, the matrices A, B commute. Therefore, they are simultaneously diagonalizable.
Exercise 13. According to the Lemma 3.3.14 we have that a matrix is positive semidefnite iff it can be expressed as $B^T B$ for some matrix B. Let $S = B^T B, DSD^T = DB^T BD^T = (BD^T)^T (BD^T)$. Hence, it is positive semidefinite.

Let's examine $x^T DSD^T x$. $(x^T D)^T = D^T x, y = x^T D$. So, $y^T Sy \geq 0$ is a positive semidefinite.
Exercise 14. $x^T Sx \geq 0$ for any $x \in R^n$. Take $x = e_i$ where e_i is a vector with i-th element 1 and the other elements being 0. Then, $e_i^T Se_i = S_{ii} \geq 0$.
Exercise 15. $P^2 = PP = P$. Let λ be an eigenvalue of P with eigenvector v. Then $Pv = \lambda v, P^2 v = \lambda Pv = \lambda^2 v = \lambda v = Pv, (\lambda - 1)\lambda v = 0$. So, $\lambda = 0, \lambda = 1$.
Exercise 16. A
Exercise 17. $Ax = \lambda_r x$. $yA = \lambda_l y$

$yAx = \lambda_r yx = \lambda_l yx$. Since $\lambda_r \neq \lambda_l$ then $(\lambda_r - \lambda_l)yx = 0$. Implies that $yx = 0$.
Exercise 18. a) False. Any strictly upper triangular matrix has 0 eigenvalues.

b) True. $det(A) = 0$
Exercise 19. Let $Bv = w$, then $ABv = Aw = \lambda v$

$\lambda Bv = \lambda w = BAw$.
Exercise 20. $f(x_1, x_2, x_3) = 2x_1^2 + 3x_2^2 + 2x_3^2 - 3x_1x_2 - x_2x_3 - 2x_1x_3$. To see if it is convex let's try to write it in a matrix polynomial form.

$$A = \begin{pmatrix} 2 & -3/2 & -1 \\ -3/2 & 3 & -1/2 \\ -1 & -1/2 & 2 \end{pmatrix}$$

Now according to the remark in page 124 right after (cf. Chapter 5) we know that the function in convex if A is a positive semidefinite matrix. To find out if A is positive semidefinite note that it is a symmetric matrix and let's compute

its eigenvalues. The eigenvalues of A are $\lambda_1 = 0.228, \lambda_2 = 2.672, \lambda_3 = 4.100$. Hence, the matrix is indeed positive semidefinite and our function is convex.
Exercise 21. ?? $f(x_1, x_2) = x_1^2 + 3x_1x_2 + 6x_2^2$

$A = \begin{pmatrix} 1 & 3/23/2 & 6 \end{pmatrix}$ Apply eigendecomposition.
Exercise 22. $A = A^T, B = B^T, A = VBV^{-1}$.

$tr(A - B) = tr(VBV^{-1} - B) = tr(V)tr(B - B)tr(V^{-1}) = 0$. Thus, the sum of the eigenvalues of $A - B$ is zero. Which means that either eigenvalues cancel out each other or they are all equal to 0. In the latter case we have that $A - B = 0$.
Exercise 23. ?

If the matrix is diagonalizable then the n-th root's geometric intuition will be that it rescales the vector in the same directions (eigenvectors) as the original matrix however the factor by which it does so is the n-th root of the original matrix factors.
Exercise 24.
Exercise 25. ?? I think that the geometric intuition is the going like this. $Ax = V\Lambda V^T x$ will rotate the vector x and align it in the direction of the eigenvectors then it will rescale it and rotate back. The similar effect with some other direction and scale factors will do Bx. Now suppose instead of just any vector we took the eigenvector v of $A - B$. Then for some $c \in R$ we have $(A - B)v = cv$ i.e. the vector is just rescaled by some factor c.

Just a good resource: https://www.youtube.com/watch?v=PFDu9oVAE-g
Exercise 26. $Av = \lambda v, A^2 v = \lambda^2 v$ and in general by induction it can be shown that for any $k \in N$ $A^k v = \lambda^k v$. If $A^k = 0$ then $\lambda^k = 0$ since $v \neq 0$. Therefore, $\lambda = 0$ is the only value that eigenvalues of nilpotent matrices can attain.
Exercise 27. $A^2 = A$. Let https://math.stackexchange.com/questions/3608660/if-a2-a-then-a-is-diagonalizable
Exercise 28. First of all observe that the characteristic equation of a $d \times d$ row addition matrix A is $det(A - \lambda I) = (1 - \lambda)^d$ since it is upper/lower triangular matrix. Therefore, the algebraic multiplicity of the only eigenvalue 1 of A is d. Now let's have a look at the geometric multiplicity by finding the eigenectors of A. $A - I$ is a matrix where all columns are zero except a single column with, say, $a \in R$ in one place and zeros in other places. Then the nullspace of such a matrix can be trivially computed by taking a vector which has a single 0 at the position corresponding to the column number where the off-diagonal entry is located and 1's otherwise. Hence, by this procedure we will get $d - 1$ linearly independent eigenvectors corresponding to the eigenvalue 1. Since the algebraic and geometric multiplicities does not equal the matrix is defective.
Exercise 29. P is symmetric. By the exercise $27, (i)$ P is diagonalizable, Therefore, $P = V\Lambda V^T$. Since $P^2 = P$, then by the exercise 15 all eigenvalues are 1 or 0. $P = V(\Lambda)^{1/2}(\Lambda)^{1/2}V^T$. Let $Q = V(\Lambda)^{1/2}, Q^T = (\Lambda)^{1/2}V^T$.
Exercise 30. According to the Schur decomposition $A = PUP^T$ where U is a triangular matrix. Any triangular matrix $U = D + S$ where D is a diagonal matrix and S is a strictly upper triangular matrix (nilpotent matrix). Then, $A = PUP^T = PDP^T + PSP^T$. Moreover, PSP^T is nilpotent too (note the powers of it).

Exercise 31. ??
Exercise 32. ??
Exercise 33. The rank of the Jordan form is the rank of the original matrix. To show this we should show that the multiplication by invertible matrices does not change the rank of the matrix. Since the matrix is invertible it can be represented as a product of elementary matrices. Therefore, our task changes to showing that multiplication by elementary matrix does not change the rank.
Certainly, the row exchange and row multiplication does not change the span$(v_1, v_2, ..., v_n)$ $(v_1, ..., v_n$ are the column vectors) and hence does not change the rank...

$rank(A) = rank(VUV^{-1}) = rank(U)$.

Exercise 34. $A = \begin{pmatrix} 1 & a/2 & 0 \\ a/2 & 2 & 1/2 \\ 0 & 1/2 & 1 \end{pmatrix}$

We should compute the characteristic polynomial $det(A - \lambda I) = \frac{a^2\lambda}{4} - \frac{a^2}{4} - \lambda^3 + 4\lambda^2 - \frac{19\lambda}{4} + \frac{7}{4} = 0$. Solving this we would get $\lambda_1 = 1, \lambda_2 = 3/2 - \sqrt{a^2+2}/2, \lambda_3 = (\sqrt{a^2+2}+3)/2$. So, we want $\lambda_2 \geq 0, 3 \geq \sqrt{a^2+2}, 9 \geq a^2+2, 7 \geq a^2, \sqrt{7} \geq a$.
Exercise 35.
Exercise 36.
Exercise 37. $\lim_{n\to\infty}(I + V\Lambda V^{-1})^n = \lim_{n\to\infty}(I + \Lambda/n)^n$. Therefore, the diagonal entries are the powers of e.

4 Chapter 4:

4.1 Problems

4.2 Exercises

Exercise 1. a) $F(x) = x^2 - 2x + 2, F'(x) = 2x - 2 = 0, x = 1, F(1) = 1 - 2 + 2 = 1$. The minimum is 1, the maximum is ∞ as $x \to \infty$.
b) $F(x, y) = x^2 - 2x - y^2, \nabla F(x, y) = \begin{pmatrix} 2x - 2 \\ -2y \end{pmatrix} = 0, (x, y)^T = (1, 0)^T$. $\nabla^2 F(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$. The hessian is indefinite. Therefore
Exercise 2. $J(\hat{w}_0 + \epsilon\frac{\hat{y}}{\|\hat{y}\|}) = J(\hat{w}_0) + \epsilon(\frac{\hat{y}}{\|\hat{y}\|})^T \nabla J(\hat{w}) + \frac{\epsilon^2}{2}(\frac{\hat{y}}{\|\hat{y}\|})^T H \frac{\hat{y}}{\|\hat{y}\|} = J(\hat{w}_0) + \frac{\epsilon^2}{2}(\frac{\hat{y}}{\|\hat{y}\|})^T H \frac{\hat{y}}{\|\hat{y}\|}, 2(J(\hat{w}_0 + \epsilon\frac{\hat{y}}{\|\hat{y}\|}) - J(\hat{w}_0)) = (\frac{\hat{y}}{\|\hat{y}\|})^T H \frac{\hat{y}}{\|\hat{y}\|}$
Exercise 3. Since the H is indefinite then w_0 is a saddle point. Let v be our direction. If the function is increasing in the direction v then all the elements of v that correspond to the positive eigenvalues should be non-zero and all the other elements should be equal to 0. A similar logic applies to the decrease of the function
Exercise 4. The minimum: The minimum of two convex functions is not necessarily convex. Take $f(x) = x, g(x) = -x, h = \min\{x, -x\}$. $h(2/3 * 3 + 1/3 * (-3)) = 1, 2/3h(3) + 1/3h(-3) = 2/3 * -3 + 1/3 * -3 = -3$. Is not convex.

The intersection: $C_1 \cap C_2$ is convex. Take $x, y \in C_1 \cap C_2$, then $x + (1-\theta)y \in C_1$ since C_1 is convex and also $x + (1-\theta)y \in C_2$ by the same reason. Therefore $x + (1-\theta)y \in C_1 \cap C_2$
The union: $C_1 \cup C_2$ is not convex. For example take C_1 be a unit ball centered at $0 \in R^2$. Take C_2 be a unit ball centered at $0.5 \in R^2$. Then you can visually see that their union is not convex.
Exercise 5. (i) Note that the epigraph of a convex function is convex. (refer to Convex analysis by Rockafellar or Convex optimization theory by Bertsekas for the definition of epigraph and the mentioned property)
The sum of two convex functions is convex, since the epigraph of the sum is the intersection of the epigraphs of the functions which is convex as shown in the previous exercise.
(i) False. Let $f(x) = x, g(x) = -x, h(x) = f(x)g(x) = -x^2$ is not convex.
Exercise 6.
Exercise 7.
Exercise 8. $f(w) = w^T A w + b^T w + c$ $f'(w) = 2Aw + b = 0, Aw = -b$. From which follows that the global minimum exists iff $b \in R(A)$.
Exercise 9.
Exercise 10. $f(x, y) = x^2 + 2y^2 + axy$ $\nabla f(x, y) = \begin{pmatrix} 2x + ay 4y + ax \end{pmatrix}$ $\nabla^2 f(x, y) = \begin{pmatrix} 2 & a \\ a & 4 \end{pmatrix}$. Is concave if $8 - a^2 \geq 0, 8 \geq a^2, a \in [-\sqrt{8}, \sqrt{8}]$. The matrix is negative definite if and only if its odd principal minors are negative and its even principal minors are positive. This matrix can not be negative definite because the 2 is not negative. Therefore, $f(x, y)$ is never concave. In all other cases it is indefinite.
Exercise 11. $f(x, y) = x^3/6 + x^2/2 + y^2/2 + xy$. $\nabla f(x, y) = \begin{pmatrix} x^2/2 + x + y y + x \end{pmatrix}$ $\nabla^2 f(x, y) = \begin{pmatrix} x+1 & 1 \\ 1 & 1 \end{pmatrix}$. The hessian is positive semidefinite if $x \geq -1$ and $(x+1) - 1 \geq 0, x \geq 0$ i.e. $x \in [0, \infty), y \in R$.
Exercise 12.
Exercise 13.
Exercise 14.
Exercise 15. $f : R^m \to R$ $L = f(Wx)$, $\frac{dL}{dx} = \nabla f(Wx) x^T$ is a outer product therefore it has a rank-1.
Exercise 16. If we get an additional d dimensional training point then the number of rows of D will increase but the dimension of $D^T D$ will not change. The update will become $D^T D + XX^T$. Using the hint we can compute the inverse as follows $(D^T D + XX^T)^{-1} = (D^T D)^{-1} - \frac{(D^T D)^{-1} X X^T (D^T D)^{-1}}{1 + X^T (D^T D)^{-1} X}$.

Exercise 17. a) $(\frac{dJ}{dV})_{ij} = \frac{dJ}{dV_{ij}} = \begin{pmatrix} \frac{dJ}{dV_{11}} & \frac{dJ}{dV_{12}} \cdots \frac{dJ}{dV_{1k}} \\ \cdots & \\ \frac{dJ}{dV_{d1}} & \frac{dJ}{dV_{d2}} \cdots \frac{dJ}{dV_{dk}} \end{pmatrix}$

$$(\frac{dJ}{dV_j})_i = \frac{dJ}{d(V_j)_i} = \frac{dJ}{dV_{ij}} = \begin{pmatrix} \frac{dJ}{dV_{1j}} \\ \cdots \\ \frac{dJ}{dV_{dj}} \end{pmatrix}$$

$(\frac{dJ}{dv_i})_j = \frac{dJ}{d(v_i)j} = \frac{dJ}{dV_{ij}} = \begin{pmatrix} \frac{dJ}{dV_{i1}} \\ ... \\ \frac{dJ}{dV_{ik}} \end{pmatrix}^T$

b) $J = \|V\|_F^2 = \sum_{i=1}^d \|v_i\|_2^2$ $(\frac{dJ}{dv_i})_j = (\frac{\sum_{i=1}^d \|v_i\|_2^2}{dv_i})_j = 2v_{ij}$

Exercise 19. Because there are lots of local optimums. Every global optimium for $J_i(w_i)$ can be a global optimum or a local optimum of the new function.

Exercise 20.

Exercise 21.

Exercise 22. $\sqrt{< \theta x + (1-\theta)y, \theta x + (1-\theta)y >} = \sqrt{< \theta x, \theta x + (1-\theta)y > + < (1-\theta y, \theta x + (1-\theta)y) >}$ $\sqrt{< \theta x, \theta x > + < \theta x, (1-\theta)y > + < (1-\theta)y, \theta x > + < (1-\theta)y, (1-\theta)y >} \leq$ $\sqrt{\theta^2\|x\|_2^2 + (1-\theta)^2\|y\|_2^2 + 2\theta(1-\theta) < x, y >} = \theta < x, x > +(1-\theta) < y, y >$.

$g(x) = h(f(x))$ where $h(x) = x^2$ is an increasing convex function on R_+.

Exercise 23. $J = \|C - AXB\|_2^2$. $\frac{dJ}{dX} = -2A^T(C - AXB)B^T = 0, A^TCB - A^TAXBB^T = 0, A^TCB = A^TAXBB^T$

$(A^TA)^{-1}A^TCB(BB^T)^{-1} = A^+C(B^+)^T = XBB^T$

Exercise 24.

Exercise 25. $f(x) = ax^3 + bx^2 + cx + d, f'(x) = 3ax^2 + 2bx + c = 0, D = 4b^2 - 12ac < 0, b^2 < ac$

Exercise 26. $b^2 = ac$.

Exercise 27.

Exercise 28.

Exercise 29. $h' = W_1W_2 - W_2^2W_3 + W_1W_2W_3$

5 Chapter 5:

5.1 Problems

5.2 Exercises

Exercise 1. See notebook.

Exercise 2. $x_{k+1} = x_k + p_k = x_k - \nabla^2 f_k^T \nabla f_k$

$B^{-1}y_{k+1} = B^{-1}y_k + p_k$

$\nabla f(x_k) = \nabla f(B^{-1}y_k) = B^{-1}\nabla f(y_k)$ $\nabla^2 f(x_k) = \nabla^2 f(B^{-1}y_k) = B^{-1}\nabla^2 f(y_k)$

$B^{-1}y_{k+1} = B^{-1}y_k + p_k = B^{-1}y_k - (B^{-1})^2\nabla^2 f_k^T \nabla f_k$

$y_{k+1} = y_k - B^{-1}\nabla^2 f(y_k)^T \nabla f(y_k) = Bx_k + B^{-1}B^2\nabla^2 f(x_k)^T\nabla f(x_k) = Bx_k + B\nabla^2 f(x_k)^T \nabla f(x_k) = Bx_{k+1}$

Exercise 3. The general form for the second-order Taylor expansion at the point a is: $f(x) = f(a) + f'(a)(x-a) + 1/2 * f''(a)(x-a)^2$.

So, using above around $a = 0$ we have: a) $x^2 = 0 + 0 + x = x$ b) $x^3 = 0 + 0 + 0 = 0$ c) $x^4 = 0$ d) $cos(x) = 1 + 0 - x/2 = 1 - x/2$

Exercise 4. $f'(x) = 2ax + b, f''(x) = 2a$

$x_1 = x_0 - f''(x_0)^{-1}f'(x_0) = x_0 - (x_0 + b/2a) = -b/2a$

Exercise 5,6,8. See the notebook.

Exercise 9. Let $q_1, ..., q_d$ be the conjugate directions. Let's take their linear combination and check for dependence:

$a_1 q_1 + ... + a_d q_d = 0$

Multiply by the hessian and some conjugate vector q_i.

$a_1 q_i H q_1 + ... + a_i q_i H q_i + ... + a_d q_i H q_d = 0$

$a_i q_i H q_i = 0$

However, because the function is strongly convex the only possibility is to have $a_i = 0$. This can be shown for any $i \in [1, d]$. Hence, they are independent.

Exercise 10. Let's also assume that the d linearly independent vectors are orthogonal. If no, we can make them to be orthogonal. Then, $v = a_1 e_1 + ... + a_d e_d$, $v^T e_1 = a_1 \|e_1\|^2 = 0, a_1 = 0$. The same can be shown for the other coefficients.

Exercise 11. Premultiplying both sides by $q_t H$ we can get that the q_i's are conjugate directions if b is equal to:

$q_j H q_{t+1} = q_j H v_{t+1} + \sum_i q_j H b_{ti} q_i$

$\sum_i b_{ti} q_j H q_i = b_{tj} q_j H q_j = -q_j H v_{t+1}$

So, $b_{tj} = -\frac{q_j H v_{t+1}}{q_j H q_j}$.

The main difference is that in (5.22) we are computing updates where H is included only t times. Now we are doing it for each i t times.

Exercise 12. a) Using (5.21), (5.23) and the way Hq_i is defined in the exercise: $\sigma_i[q_i^T H q_{t+1}] = [\nabla J(W_{i+1}) - \nabla J(W_i)]^T q_{t+1} = [\nabla J(W_{i+1}) - \nabla J(W_i)]^T[-\nabla J(W_{t+1}) + b_t q_t] = -[\nabla J(W_{i+1}) - \nabla J(W_i)]^T[\nabla J(W_{t+1})] + [\nabla J(W_{i+1}) - \nabla J(W_i)]^T[b_t q_t] = -[\nabla J(W_{i+1}) - \nabla J(W_i)]^T[\nabla J(W_{t+1})] + \sigma_i b_t(q_i^T H q_t)$.

Also, $\sigma_t q_i^T H q_t = \sigma_t q_i^T[\nabla J(W_{t+1}) - \nabla J(W_t)]/\sigma_t = [\nabla J(W_{t+1}) - \nabla J(W_t)]q_i$.

b) Using the hint let's induct on t.

Base case: $t = 0$. Then, $\sigma_0 q_0^T H q_1 = -[\nabla J(W_1) - \nabla J(W_0)]^T[\nabla J(W_1)] + \sigma_0 b_0(q_0^T H q_0) = -[\nabla J(W_1) - \nabla J(W_0)]^T[\nabla J(W_1)] + [\nabla J(W_1) - \nabla J(W_0)]q_0 = -\nabla J(W_1)^T \nabla J(W_1) + \nabla J(W_0)^T \nabla J(W_1) - \nabla J(W_1)^T \nabla J(W_0) + \nabla J(W_0)^T \nabla J(W_0) = \nabla J(W_0)^T \nabla J(W_0) - \nabla J(W_1)^T \nabla J(W_1)$

If we use normalized gradients of the loss function at each step then we would have $\sigma_0 q_0^T H q_1 = 0$.

$q_0 = -\nabla J(W_0)$

$W_1 = W_0 + a_0 q_0$

$J(w_1) = b_0 q_0 - q_1$

Exercise 13. Note that the exercise is about regularized L_2-loss SVM. Here we will use the hints given to get:

$J(W) = \sum_i J_i(W) = \frac{1}{2}\sum_i(\max\{0, (1 - y_i[W X_i^T])\})^2$

$(D^T A_w D + aI_d)^{-1} D^T A_w y = (D^T \sqrt{A_w}\sqrt{A_w} D + aI_d)^{-1} D^T A_w y = (D_w^T D_w + aI_d)^{-1} D^T A_w y = D^T A_w y (aI_+ D_w D_w^T)^{-1}$

Since D is wide then $n < d$.

Exercise 14. $F(x_1, ..., x_d) = f(x_1) + ... + f(x_d) = x_1^3 - 3x_1^3 + ... + x_d^3 - 3x_d^3$.

$\nabla F = (3x_1^2 - 3, ..., 3x_d^2 - 3)^T$.

Hence the only critical points in R^d are $x' = (\sqrt{3}/3, ..., \sqrt{3}/3)$ and $x'' = -1 * (\sqrt{3}/3, ..., \sqrt{3}/3)$.

We have the hessian is: $H = \begin{pmatrix} 6x_1 & 0 & ... & 0 \\ 0 & 6x_2 & ... & 0 \\ ... & & & \\ 0 & 0 & ... & 6x_d \end{pmatrix}$.

For any vector $v = (v_1, ..., v_d) \in R^d, v^T H v = 6 \sum_{i=1}^{d} v_i^2 x_i$.

So, at $x', H(x')$ is a positive definite matrix and at $x^{''}, H(x^{''})$ is a negative definite matrix.

Exercise 17. $f(x) = x^3 - 3x$

$f'(x) = 3x^2 - 3$

$f''(x) = 6x$

The minimum: $x_0 = 0.5$

$x_1 = 0.5 - [1/(6 * 0.5)]f'(0.5) = 1.25$

$x_2 = 1.25 - [1/(6 * 1.25)]f'(1.25) = 1.025$

$x_3 = 1.025 - [1/(6 * 1.025)]f'(1.025) = 1$

The maximum: $x_0 = -0.8$

$x_1 = -0.8 - [1/(6 * -0.8)]f'(-0.5) = -1.025$

$x_2 = -1.025 - [1/(6 * -1.025)]f'(-1.025) = -1$

Exercise 18.

$\|DW - y\|_1$ is the sum of the absolute values of the components of the vector by definition. Since the function $|x|$ is not differentiable at 0 then we can't apply the Newton-method to it.

6 Chapter 6:

6.1 Problems

6.2 Exercises

Exercise 1. The problem can be formulated as follows:

maximize xy

subject to:

$x^2 + y^2 = 1$.

Exercise 2. The gradient is

$\nabla f(x, y) = \begin{pmatrix} 2x + 2 \\ 2y + 3 \end{pmatrix}$.

Let $v_0 = (x_0, y_0)$ and let $V = \{(x, y) : x + y = 1\}$.

By page (257) point 2 the projection operator is $P(v_0) = \arg\min_{v \in V} \|v - v_0\|^2$. However, since $v_0 \in V$ already, then the projected gradient will be just our usual gradient evaluated at the point v_0. So,

$\nabla f(x_0, y_0) = \begin{pmatrix} 2x_0 + 2 \\ 2y_0 + 3 \end{pmatrix}$.

Exercise 3. Let $y = 1 - x$

Then the objective becomes:

$f(x) = x^2 + 2x + 1 + x^2 - 2x + 3 - 3x = 2x^2 - 3x + 4$.

$f'(x) = 4x - 3 = 0, x = 3/4$. Moreover, since it is a quadratic equation with positive coefficient for the x^2 term then $x = 3/4$ is the minimum point and $f(3/4)$ is the minimal value.

Exercise 4. $L(x, y, a) = x^2 + 2x + y^2 + 3y + a(x + y - 1)$.

$\frac{dL}{dx} = 2x + 2 + a = 0, x = -(2 + a)/2$

$\frac{dL}{dy} = 2y + 3 + a = 0, y = -(3 + a)/2$

maximize $g(a) = (-\frac{2+a}{2})^2 - (2 + a) + (-\frac{3+a}{2})^2 + a(-\frac{2+a}{2} - \frac{3+a}{2} - 1) = \frac{4+a^2+4a}{4} - 2 - a + \frac{9+a^2+6a}{4} + \frac{-2a^2-5a}{2} - a$

$g'(a) = \frac{2a+4}{4} - 1 + \frac{2a+6}{4} + \frac{-4a-5}{2} - 1 = 0$

$2a + 4 - 4 + 2a + 6 - 8a - 10 - 4 = 0, -4a = 8, a = -2, y = -1/2, x = 0$

Exercise 5. See the notebook.

Exercise 6. The primal problem can be written as:

minimize $\sum_{i=1}^{d} c_i w_i$

subject to:

$\sum_j A_{ij} w_j \leq b_i$ for $i = 1, ..., d$.

$L(c, w, v) = c^T w + v^T (Aw - b) = (c + A^T v)^T w - b^T v$.

$g(v) = -b^T v$ if $A^T v + c = 0$ and

$g(v) = -\infty$ otherwise.

So, the dual problem is given as:

maximize $-b^T v$

subject to:

$A^T v + c = 0, v \geq 0$

Exercise 7. $L(w, v) = 0.5w^T Qw + c^T w + v^T (Aw - b)$.

$dL/dw = Qw + c + A^T v = 0, w = Q^{-1}(A^T v + c)$.

$g(v) = -0.5(A^T v + c)^T Q^{-1}(A^T v + c) - b^T v$.

The dual problem is: $-0.5(A^T v + c)^T Q^{-1}(A^T v + c) - b^T v$

subject to:

$v \geq 0$.

Exercise 8. Minimize $J(W) = \frac{\lambda}{2} \|W\|^2 + C \sum_i e_i$

subject to:

$e_i \geq 1 - y_i[W X_i^T] + b$

$e_i \geq 0$ for all $i = 1, ..., n$.

So, the Lagrange will be: $L(W, b, a, v) = \frac{\lambda}{2} \|W\|^2 + C \sum_i e_i - \sum_i a_i(e_i - 1 + y_i(W X_i^T) - b) - \sum_i v_i e_i$.

$\frac{dL}{dW} = \lambda W = \sum_i a_i y_i X_i^T$

$\frac{dL}{db} = -\sum_i a_i = 0$

$\frac{dL}{de_i} = C - a_i - v_i = 0$

So, after dropping e_i and b the Lagrangian becomes:

$g(a, v) = \frac{\lambda}{2} \|W\|^2 - \sum_i a_i y_i W X_i^T = \frac{\lambda}{2} \|\frac{\sum_i a_i y_i X_i^T}{\lambda}\| - \sum_i a_i y_i W X_i^T = \frac{1}{2} \sum_i \sum_j a_i a_j y_i y_j X_i X_j - \sum_i a_i y_i \sum_j a_j y_j X_i X_j / \lambda = (\frac{1}{2} - \frac{1}{\lambda}) \sum_i \sum_j a_i a_j y_i y_j X_i X_j$

Exercise 9.

Exercise 10. The gradient of the ellipsoid is given by $2Ax + b$. The gradient points outwards of the ellipsoid and forms the steepest angle of 90 degrees with it. Hence, any vector orthogonal to ellipsoid is a multiple of its norm (gradient). Since $z - z_0$ is orthogonal to the ellipsoid then it is proportional to $2Az_0 + b$.

$(https : //math.stackexchange.com/questions/599488/why - gradient - vector - is - perpendicular - to - the - plane)$

Exercise 11. $f(x, y, z) = x^2 - y^2 - 2xy + z^2$

$$\nabla f = \begin{pmatrix} 2x - 2y \\ -2y - 2x \\ 2z \end{pmatrix}$$

1) Set $(x, y, z) = (x, 1, 0)$. Then, $x^2 \leq 1, x \in [-1, 1]$. Make the derivative of the objective function w.r.t. x be equal to 0. Then, $2x = 2, x = 1$. Since, $x = 1$ is a feasible point we can proceed.

2) Set $(x, y, z) = (1, y, 0)$. Then, $y^2 \leq 1, y \in [-1, 1]$. Make the derivative of the objective function w.r.t. y be equal to 0. Then, $-2y = 1, y = -1$. Since, $y = -1$ is a feasible point we can proceed.

3) Set $(x, y, z) = (1, -1, z)$. Then, $z^2 \leq 1, z \in [-1, 1]$. Make the derivative of the objective function w.r.t. z be equal to 0. Then, $2z = 0, z = 0$. Since, $z = 0$ is a feasible point we got that our objective function subject to its constraints achieves its minimum at $(x, y, z) = (1, -1, 0)$ point which.

Exercise 12. We can show that the supremum of convex functions is convex. Hence, the infimum will be concave.

Let f_i be a family of convex functions on a convex compact set V.

Let $f = \sup f_i$. Then for any $x, y \in V$ and for any $\theta \in [0, 1]$ we have:

$f_i(\theta x + (1 - \theta)y) \leq \theta f_i(x) + (1 - \theta)f_i(y) \leq \theta f(x) + (1 - \theta)f(y)$.

Since, this holds for all f_i then, we can take the sup to get: $f(\theta x+(1-\theta)y) \leq \theta f(x) + (1 - \theta)f(y)$

Exercise 13. In the L_2-regularized linear regression the primal problem is formulated as follows:

minimize $0.5\|y - Dw\|^2$

subject to:

$\|w\|^2 \leq b$

$w \geq 0$

So, $L(w, a, v) = 0.5\|y - Dw\|^2 + a(\|w\|^2 - b) + v^T w$.

$\frac{dL}{dw} = D^T Dw - D^T y + 2aw + v = (D^T D + 2a)w - D^T y + v = 0, w = (D^T D + 2a)^{-1}(D^T y - v)$.

$\frac{dL}{da} = \|w\|^2 - b = 0, \|w\|^2 = +b$

$\frac{dL}{dv} = w = 0$.

Hence, $g(a) = 0.5\|y - D[(D^T D + 2a)^{-1}(D^T y - v)]\|^2$ is the dual objective function.

Exercise 14. The primal problem is: minimize: $0.5\|Ax - b\|^2$

subject to:

$\|x\| \leq r$.

We can introduce an equality constraint to have the primal problem be:

minimize: $0.5\|Ax - b\|^2 + \lambda\|y\|$

subject to:

$y = x$.

Then, $L(x, y, a) = 0.5\|Ax - b\|^2 + \lambda\|y\| + a^T(x - y)$.
$\frac{dL}{dx} = A^T(Ax - b) + a = 0, x = (A^TA)^{-1}(A^Tb - a)$.
Using the conjugate function of the l_1-norm we have:
$\inf_y \lambda\|y\|_1 - a^Ty = -\sup_y a^Ty - \lambda\|y\|_1 = 0$ if $\|a\|_\infty \leq \lambda$ and $-\infty$ otherwise.
So, $g(a) = \frac{1}{2}\|A(A^TA)^{-1}(A^Tb - a) - b\|^2 + a^T(A^TA)^{-1}(A^Tb - a) = \frac{1}{2}(A^Tb - a)^T(A^TA)^{-T}A^TA(A^TA)^{-1}(A^Tb - a) - (A^Tb - a)^T(A^TA)^{-T}A^Tb + b^Tb - a^T(A^TA)^{-1}A^Tb - a^T(A^TA)^{-1}a = \frac{1}{2}(A^Tb - a)^T(A^TA)^{-1}(A^Tb - a) - (A^Tb - a)^T(A^TA)^{-T}A^Tb + a^T(A^TA)^{-1}A^Tb - a^T(A^TA)^{-1}a + \text{const} = \frac{1}{2}b^TA(A^TA)^{-1}A^Tb - b^TA(A^TA)^{-1}a + \frac{1}{2}a^T(A^TA)^{-1}a - b^TA(A^TA)^{-1}A^Tb + a^T(A^TA)^{-1}A^Tb + a^T(A^TA)^{-1}A^Tb - a^T(A^TA)^{-1}a + \text{const} = a^T(A^TA)^{-1}A^Tb - \frac{1}{2}a^T(A^TA)^{-1}a$.
So, the dual problem is:
minimize: $a^T(A^TA)^{-1}A^Tb - \frac{1}{2}a^T(A^TA)^{-1}a$
subject to:
$-\lambda \leq a_i \leq \lambda$ for $i = 1, ..., n$.

Exercise 15. The primal problem is:
minimize: $\frac{1}{2}\|Ax - b\|^2$
subject to:
$-r \leq x_i \leq r$ for $i = 1, ...n$.
Hence, to get the primal gradient-descent we should follow the special case of box constraints procedure as described in section (6.2.2.1).
1) The gradient-descent step: $A^T(Ax - b) = 0, x_i = x_{i-1} - a[A^T(Ax_{i-1} - b)]$.
2) Find the components in the interval bounds: If $x_i \in [-r, r]$ return x_i. If $x_i \geq r$ return r. Otherwise return $-r_i$.

Exercise 16. https://math.stackexchange.com/questions/4286639/best-subset-selection

Exercise 17. The duality gap at the k-th iteration can be obtained by computing the difference between the dual and primal optimal solutions at the k-th step.
By (6.10) we know that the dual and primal variables are related by $W = \sum_{i=1}^n a_i y_i X_i^T$. So, to get the duality gap one must compute:
$J(W^k) - g(a_k)$.

Exercise 18. To satisfy John von Neumann's strong duality conditions we should have that the function is convex in its minimization and concave in its maximization variables.
$(i), f(x, y) = x^2 + 3xy - y^4$. Here f is a quadratic function in x with a positive coefficient. Hence, it is convex. Furthermore, $\frac{df}{dy^2} = -12y^2 \leq 0$. Hence, is a concave function in y. Therefore, it satisfies the John von Neumann's conditions.
(ii) Do not satisfy. Since, $f(x, y)$ is not concave in y.
$(iii - iv)$ $\frac{df}{dx^2} = \frac{-d\cos(y-x)}{dx} = \sin(y - x)$ is convex in $0 \leq x \leq y \leq \pi/2$.
$\frac{df}{dy^2} = \frac{d\cos(y-x)}{dy} = -\sin(y - x)$ is concave in $0 \leq x \leq y \leq \pi/2$.

Exercise 19. The primal problem is equivalent to:
minimize $x^2 + y^2$ subject to $-x - y \leq -1$.
The Lagrangian is: $L(x, y, a) = x^2 + y^2 - a(x + y - 1)$. Then,
$\frac{dL}{dx} = 2x - a = 0, x = a/2$ a $\frac{dL}{dy} = 2y - a = 0, y = a/2$

$\frac{dL}{da} = x + y - 1 = 0$

So, the dual problem is: maximize $a^2/2$ subject to $a = 1$. Hence, we get the primal optimal to be $(x, y) = (1/2, 1/2)$. Right?

Now if we start at $(1, 0)$. Let's solve for x.

plug in $y = 0$ to get minimize x^2 subject to $x \geq 1$. So, $x = 1$.

Then plug in $x = 1$ to solve for y: minimize $y^2 + 1$ subject to $y \geq 0$. We will get $y = 0$.

Hence, basically we are stuck at the starting point $(1, 0)$.

Exercise 20. Take in the constraint $x + y \leq 1$

7 Chapter 7:

7.1 Problems

7.2 Exercises

Exercise 1. By SVD we have $D = U\Delta V^T$.

- $D^T D = V\Delta U^T U\Delta V^T = V\Delta^2 V^T$
- $DD^T = U\Delta V^T V\Delta U^T = U\Delta^2 U^T$

$(\lambda I + D^T D)^{-1} D^T = (\lambda I + V\Delta^2 V^T)^{-1} D^T = [V(\lambda I + \Delta^2)V^T]^{-1} D^T = V(\lambda I + \Delta^2)^{-1} V^T V\Delta U^T = V(\lambda I + \Delta^2)^{-1}\Delta U^T$

At the same time we have:

$D^T(\lambda I + DD^T)^{-1} = D^T(\lambda I + U\Delta^2 U^T)^{-1} = D^T[U(\lambda I + \Delta^2)U^T]^{-1} = D^T U(\lambda I + \Delta^2)^{-1} U^T = V\Delta(\lambda I + \Delta^2)^{-1} U^T$

Now we need to show that:

$$V(\lambda I + \Delta^2)^{-1}\Delta U^T = V\Delta(\lambda I + \Delta^2)^{-1} U^T$$

Multiply from left by V^T and from right by U to get: $(\lambda I + \Delta^2)^{-1}\Delta = \Delta(\lambda I + \Delta^2)^{-1}$.

Now sine both $(\lambda I + \Delta^2)^{-1}$ and Δ are diagonal matrices their product is commutative.

Exercise 2. $Z(\Lambda I + D^T D)^{-1} D^T y = ZD^T(\lambda I + DD^T)^{-1} y$

Where ZD^T is the similarity between Z and training points and DD^T is the similarity between and training points.

Exercise 3. For the truncation error to be the same by the Corollary (7.2.3), Lemma (7.2.6) and Lemma (7.2.7) it is sufficient to have:

$\|D_k\|_F^2 = \sum_{r=1}^{k} \hat{\sigma}_{rr}^2 \|\hat{q}_r \hat{p_r}^T\|_F^2 = \sum_{r=1}^{k} \sigma_{rr}^2 \|q_r p_r^T\|_F^2$

So, as long as the columns of Q, P have unit norm (Lemma (7.2.7)) and at least for one of the matrices $\hat{Q}, P$ the columns are orthogonal (Lemma (7.2.6)) and $\sigma_{rr} = \hat{\sigma}_{rr}$ for the first k-terms we would have the same truncation loss for the $D_k = \hat{Q}\hat{P}^T$

Exercise 4.

Exercise 5. Since $Ax_0 = b$ then for all $v \in V$ we would have $A(x_0 + v) = b$ if $Av = 0$. As it is mentioned in section (7.4.3) we can find the basis of this V by taking $(d - r)$ zero right singular vectors in columns of P.

Exercise 6. Since B is a square, symmetric matrix it has an eigen-decomposition $B = V\Lambda V^T$.

$Bv_i = \lambda_i v_i$

$\begin{pmatrix} 0 & D^T \\ D & 0 \end{pmatrix} \begin{pmatrix} v_{1:d} \\ v_{d+1:n} \end{pmatrix} = \begin{pmatrix} D^T v_{d+1:n} \\ Dv_{1:d} \end{pmatrix} = \lambda_i \begin{pmatrix} v_{1:d} \\ v_{d+1:n} \end{pmatrix}$

$D = Q\Sigma P^T$ where the columns of Q are constructed by the $v^i_{1:d}$ for each ith eigenvector of B. Similarly, the columns of P are constructed by the $v^i_{d+1:n}$ for each ith eigenvector of B.

It remains to show that $v_{1:d}$ is the eigenvector of $D^T D$ with eigenvalue λ_i and the same for DD^T and $v_{d+1:n}$. $D^T Dv_{1:d} = D^T(Dv_{1:d}) = \lambda_i(D^T v_{d+1:n}) = \lambda_i v_{1:d}$

Similarly for DD^T.

Exercise 7. $A^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/4 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^T$

Exercise 8.

Exercise 9. $D' = D + N$

$rank(D') = rank(D + N) \geq rank(D) - rank(N) = r - rank(N)$

Alos, $rank(D') = rank(D + N) \leq rank(D) + rank(N) \leq r + rank(N)$

$D + N = USV^T$

$Q\Sigma P^T + M\Lambda H^T = USV^T$

Exercise 10. Note that by the definition: $J = \|D - UV^T\|_F^2 = \sum_{k,j}(D_{ki} - \sum_j U_{ij}V^T_{jk})^2$

a) We want to find the best matrices U, V^T such that D the above objective is minimized. Hence, $J = \sum_{k,j}(D_{kj} - \sum_i U_{ki}V^T_{ij})^2$

$\frac{dJ}{dV} = -2\sum_k U_{ki}(D_{kj} - \sum_i U_{ki}V^T_{ij}) = 0$

$U^T(D - UV^T) = 0, U^T D = U^T UV^T, D^T U = VU^T U$

The same way we have: $\frac{dJ}{dU} = -2(D - UV^T)V = 0, DV = UV^T V$

b) From above we have: $\frac{dJ}{dU} = (D - UV^T)V = EV$.

$\frac{dJ}{dV} = (D - VU^T)^T U = E^T U$.

Hence, the steepest descent updates are: $U = U + aEV$

$V = V + aE^T U$

c) No, there are not such constraints in the optimization problem.

Exercise 11. $J = \|D - UV^T\|_F^2 + a(\|U\|_F^2 + \|V\|_F^2)$

$\frac{dJ}{dU} = -2EV + 2U$

$\frac{dJ}{dV} = -2E^T U + 2V$

Exercise 12. a) Let A be an $d \times d$ matrix.

$|det(A)| = |det(USV^T)| = |det(U)det(S)det(V^T) = det(S)| = |\prod_{i=1}^n \sigma_i|$

Since $det(U) = det(V^T) =^+_- 1$

b) $\|A^{-1}\|_F^2 = \|(USV^T)^{-1}\|_F^2 = \|VS^{-1}U^T\|_F^2 = tr(US^{-1}V^T VS^{-1}U^T) = tr(U(S^{-1})^2 U^T) = tr((S^{-1})^2 U^T U) = tr((S^{-1})^2) = \sum_{i=1}^d (.ac1\sigma_i)^2$

Exercise 13. If $A = A^T$ then $AA^T = AA = A^T A$.

Let $AA^T = A^T A$ and $A = USV^T$ then $AA^T = USV^T VSU^T = US^2 U^T$

$A^T A = VSU^T USV^T = VS^2 V^T$

Now we have, $US^2 U^T = VS^2 V^T$

Using the fact that, $AV = US$ and $U^T A = SV^T$ we have:

$(AV)SU^T = VS(U^T A) = A(VSU^T) = (VSU^T)A, AA^T = A^T A$

Exercise 14. Not unique since there is a repeated singular value i.e. an eigenvalue of $D^T D/DD^T$ such that the dimension of its eigenspace is bigger than one.

Let $v_1^T = (0, 1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2}), v_2^T = (0, 1/\sqrt{2}, 1/\sqrt{2}, -1/\sqrt{2})$.

Hence, $D^T D v_1 = 1 v_1, D^T D v_2 = 1 v_2$.

We can write instead of v_1, v_2 any linear combination of them.

Exercise 15. For any diagonal matrix we have that $A^T A = AA^T$. Moreover, the singular values are equal to the diagonal elements of A since the eigenvalues of $A^T A$ are the squares of the diagonal elements of A. The eigenvector corresponding to the eigenvalue λ_i is e_i if $\lambda_i > 0$ and otherwise $-e_i$.

For orthogonal matrices we have $A^T A = AA^T = I$. Hence, any vector is an eigenvector for them. Moreover, by definition singular values as a square roots of the eigenvalues of $A^T A = I$ are all 1. One possible choise is $U = A, S = I, V^T = I$.

Exercise 16.

Exercise 17. We need to show that US^2U^T and VS^2V^T are similar. Recall that matrices A, B are similar if there is some matrix P such that $A = PBP^{-1}$.

In this case we have: $US^2U^T = (UV^T)VS^2V^T(VU^T)$. So, $P = UV^T$.

If the matrices are rectangular then they are not similar since P should be a square matrix to be invertible.

Exercise 18. $\|DP\|_F^2 = tr[(DP)^T DP] = tr[P^T D^T DP]$

$\|DPP^T\|_F^2 = tr[(DPP^T)^T(DPP^T)] = tr[PP^T D^T DPP^T] = tr[(P^T D^T DP)P^T P] = tr[P^T D^T DP]$

Where we used the fact that $tr(AB) = tr(BA)$ and the orthogonality of P.

Exercise 19. By the definition of the SVD the columns of P are the eigenvectors of $D_1^T D$ for the D_1 matrix and the eigenvectors of $D_2^T D_2$ for the D_2 matrix. Since, they are the same then $P_1^T = P_2^T = P^T$. The singular values are also the same since they are square roots of $D_1^T D_1 = D_2^T D_2$.

$D_2 = Q_2 \Sigma P^T = Q_2 Q_1^T (Q_1 \Sigma P^T) = Q_{12} D_1$.

Exercise 20. $A = ab^T$. For the outer product matrix every column of it is a multiple of the first column.

Hence, the only non-zero eigenvector/eigenvalue pair is: $Av = \lambda v, ab^T v = (b^T v)a = \lambda v$

Hence, $v = a, \lambda = b^T a$.

$AA^T = ab^T ba^T = (b^T bs)aa^T$.

If we use the normalized a, b vectors then $b^T b = \|b\|^2 = 1$ and the eigenvalues/eigenvectors of aa^T can be found as above.

Exercise 21.

Exercise 22. Let $A = USV^T, AA^T = US^2U^T, A^T A = VS^2V^T$.

Also, note that $(AA^T)^n = US^{2n}U^T$.

a) $A^T(AA^T)^5 = VSU^T US^{2n} U^T = VS^{2n+1}U^T$. Hence, the ith singular value of this matrix is equal to $1/(2n+1)$ power of the ith singular value of the original matrix.

b) $A^T(AA^T)^5A = VS^{2n+1}U^TUSV^T = VS^{2n+2}V^T$.Hence, the ith singular value of this matrix is equal to $1/(2n+2)$ power of the ith singular value of the original matrix.

c) $A(A^TA)^{-2}A^T = A[(A^TA)^{-1}]^2A^T = USV^T(VS^{-2}V^T)^2VSU^T = USV^TVS^{-4}V^TVSU^T = USS^{-2}SU^T = US^{-2}U^T$. Hence, the singular values are equal to -2 power of the original.

d) $A(A^TA)^{-1}A^T = USV^T(VS^{-2}V^T)VSU^T = U\begin{pmatrix}\text{diag}(\sigma_i)_{d\times d}\\ 0_{n-d\times d}\end{pmatrix}^T\begin{pmatrix}\text{diag}(\sigma_i^{-2})_{d\times d}\\ 0_{n-d\times d}\end{pmatrix}\begin{pmatrix}\text{diag}(\sigma_i)_{d\times d}\\ 0_{n-d\times d}\end{pmatrix}U^T = U\begin{pmatrix}I_{d\times d} & 0\\ 0 & 0_{n-d\times n-d}\end{pmatrix}U^T$. https://math.stackexchange.com/questions/3673366/find-svd-of-a-at-a-1-atps **Exercise 23.**

Exercise 24. $x^* = (A^TA+\lambda I_d)^{-1}A^Tb = (VSV^T+\lambda I_d)^{-1}VSU^Tb = V(\lambda I + S^2)^{-1}SU^T$

$\|x^*\| = US(\lambda I+S^2)^{-1}V^TV(\lambda I+S^2)^{-1}SU^T = US(\lambda I+S^2)^{-2}(US)^T = \sum_i \frac{u_iu_i^T\sigma_i^2}{(\lambda+\sigma_i^2)^2}$.

Keeping everything constant as $\lambda\to\infty$ we have $\|x^*\|\to 0$

Exercise 25. ?? $b^TQSP^T(PSQ^TQSPT+\lambda I)^{-2}PSQ^Tb = b^TQSP^T(PS^2P^T+\lambda I)^{-2}PSQ^Tb = b^TQSP^T(P(S^2+\lambda I)P^T)^{-2}PSQ^Tb = b^TQS(S^2+\lambda I)^{-2}SQ^Tb$

Exercise 26. $A^+ = (A^TA)^{-1}A^T = VS^{-2}V^TVSU^T = VS^{-1}U^T$

a) $AA^+A = (USV^T)(VS^{-1}U^T)(USV^T) = USV^T = A$.

b) $A^+AA^+ = (VS^{-1}U^T)(USV^T)(VS^{-1}U^T) = VS^{-1}U^T = A^+$

c) $AA^+ = (USV^T)(VS^{-1}U^T) = I$ hence, is symmetric.

d) $(AA^+)^2 = AA^+AA^+ = (AA^+A)A^+ = AA^+$. Hence, it is idempotent.