CS6210 - Homework/Assignment-4

Arnab Das(u1014840)

November 7, 2016

1: Chapter-6, question-4

(a) The techniques discussed in this chapter are for polynomial data fitting and not exponential data fitting, hence cannot be applied directly to $u(t)$.

(b) Given approximation of the form

$$u(t) = \gamma_1 \exp(\gamma_2 t)$$

and provided data points as $(t_1, z_1), (t_2, z_2), \ldots, (t_m, z_m)$, where $z_i > 0, i = 1, 2, \ldots, m$ and $m > 0$.
Considering instead the following approximation:

$$v(t) = lnu(t) = (ln\gamma_1) + \gamma_2 t$$

such that the data points become $(t_1, b_1), (t_2, b_2), \ldots, (t_m, b_m)$ where $b_i = lnz_i$ and

$$v(t) = x_1 + x_2 t$$

such that $x_1 = ln\gamma_1$ and $x_2 = \gamma_2$ and $v(t) = lnu(t)$
Then for solving $A^T Ax = A^T b$, we define the following matrices:
$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \ldots & \ldots \\ 1 & t_m \end{bmatrix}$ leading to $B = A^T A = \begin{bmatrix} \sum_{i=1}^{m} 1 = m & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} {t_i}^2 \end{bmatrix}$ and $b = \begin{bmatrix} t_1 \\ t_2 \\ \ldots \\ t_m \end{bmatrix}$ leading to $A^T b = \begin{bmatrix} \sum_{i=1}^{m} b_i \\ \sum_{i=1}^{m} t_i b_i \end{bmatrix}$ Thus solving

$$B = A^T Ax = A^T b$$

we get the following two equations:

$$x_1 + x_2 = 1 \tag{1}$$

and

$$3x_1 + 5x_2 = 4.9 \tag{2}$$

Solving (i) and (ii) , we get: $x_1 = ln\gamma_1 = 0.5$ and $x_2 = \gamma_2 = 0.95$
Thus, we have ,

$$v(t) = lnu(t) = ln(0.5) + 0.95t$$
$$=> u(t) = 0.5 \times exp(0.95t)$$

(Answer).

2: Chapter-6, question-5

(a) For tall and skinny matrices, A, of the system of equations , $Ax = b$, the number of rows is much larger than the number of columns. Let the number of rows be m and the number of columns n, then in such cases generally $m >> n$. These systems are overdetermined and b is generally not in the range space of A. Thus applying LU does not gives a unique solution to Ax=b. Instead, a better way to approach such problems is to minimize the residual $||b - Ax||$, such that within the tolerance of the residual we have the best solution for x. When transforming to the normal equation, $A^T Ax = A^T b$, here one can use LU decomposition since we have $n \times n$ matrix, however cholesky decomposition is a better choice here since the matrix $A^T A$ is symmetric positive definite. In case of the QR decomposition, which has an upper triangular part in R, however the Q matrix allows us to extract the upper-triangular system of equation whose solution leads to a solution of the least square problem. Here, the orthonormal behaviour of Q is used to transform into

equivalent set of equations such that the **norms are not affected**. Finally, SVD is used more in cases where A is rank deficient or nearly rank deficient, in which cases LU cannot be used. Thus, in all the cases LU directly doesn't fits the scenario for application except that LU requires to be slightly modified so that for every column it zeroes out, the vector of its remaining elements is orthonormal to the previous columns. With this introduction of orthonormality, it can be used in QR decomposition.

(b) In the normal equation: the way condition number came to be $K(B) = K^2(A)$. This was because of the folliwing derivation:
In normal equation we were solving:

$$A^T Ax = A^T b$$

and

$$(A^T A)^{-1} A^T = V S^{-2} V^T V \sum^T U^T = V(S^{-1}0)U^T$$

where A = $U \sum V^T$, V and U are orthogonal matrices and $\sum$ is a diagonal matrix with singular values of A along the diagonal, and hence $A^T A = V \sum U^T U \sum V^T = V \sum^2 V^T$
Thus $||(A^T A)^{-1} A^T|| = ||(S^{-1}0)||$ and $||((A^T A)^{-1} A^T)^{-1}|| = ||V(S0)U^T||$
Thus the condition number becomes:

$$K(B = A^T A) = \frac{\lambda_1}{\lambda_n} = \frac{\sigma_1^2}{\sigma_n^2} = K^2(A)$$

Now, for QR, we write $x = (A^T A)^{-1} A^T b$, and we have represented A=QR by a QR decomposition where Q is orthonormal and of same dimension as A. The following translation results in the final form:

$$x = (A^T A)^{-1} A^T b = (R^T Q^T QR)^{-1} R^T Q^T b = (R^T R)^{-1} R^T Q^T b) = R^{-1} Q^T b$$

Hence, we get the following relation,

$$x = R^{-1} Q^T b$$

Notice, that multiplication by orthogonal matrices do not affect the norms, thus,

$$||R^{-1} Q^T|| = ||R^{-1}|| = ||A^{-1}||$$

and also,

$$||(Q^T)^{-1} R|| = ||R|| = ||A||$$

Thus the condition number in this case comes to be $||A||||A^{-1}|| = K(A)$.
The main saving comes due to usage of the orthonormal decomposition for the transformations, since the transformations using q only transofrms them in space without affecting the norm values, while for the previous normal equation, the singular values were getting multiplied during creation $A^T A$, introducing change in norm values during the transformations.

3: Chapter-8, question-8

(a) Given a rank deficient matrix, we analyse here its effect on the Gra,-Schmidt process. Consider the span of three vectors $v_1, v_2 and v_3$, such that v_3 is linearly dependent on v_1 and v_2. We can write v_3 as a linear combination of v_1 and v_2, as:

$$\mathbf{v_3 = av_1 + bv_2}$$

Let $\mathbf{u_i}$ denote the orthonormal unit vectors we generate along the Gram-Scmidt process. Then, for the first vector, $\mathbf{v_1}$, we have

$$\mathbf{u_1} = \frac{\mathbf{v_1}}{|\mathbf{v_1}|}$$

And the orthogonal transformation of $\mathbf{v_2}$ that is orthogonal to $\mathbf{v_1}$ is

$$\mathbf{y_2} = \mathbf{v_2} - (\mathbf{v_2}.\mathbf{u_1})\mathbf{u_1}$$

Then, the next orthonormal vector, $\mathbf{u_2}$ will be along $\mathbf{y_2}$, such that $\mathbf{u_2} = \frac{\mathbf{y_2}}{|\mathbf{y_2}|}$.
Now the orthogonal transformation of $\mathbf{v_3}$, say $\mathbf{y_3}$, that is orthogonal to the span$(\mathbf{u_1}, \mathbf{u_2})$, will be,

$$\mathbf{y_3} = \mathbf{v_3} - \Big((\mathbf{v_3}.\mathbf{u_1})\mathbf{u_1} + (\mathbf{v_3}.\mathbf{u_2})\mathbf{u_2}\Big)$$

$$\mathbf{y_3} = \mathbf{v_3} - \Big(((\mathbf{av_1} + \mathbf{bv_2}).\mathbf{u_1}).\mathbf{u_1} + ((\mathbf{av_1} + \mathbf{bv_2}).\mathbf{u_2})\mathbf{u_2}\Big)$$

$$\mathbf{y_3} = \mathbf{v_3} - \Big(\mathbf{av_1} + (\mathbf{bv_2}.\mathbf{u_1})\mathbf{u_1} + (\mathbf{av_1}.\mathbf{u_2})\mathbf{u_2} + (\mathbf{bv_2}.\mathbf{u_2})\mathbf{u_2}\Big)$$

Here:
$(b\mathbf{v_2}.\mathbf{u_1})\mathbf{u_1} = \mathbf{b}(\mathbf{v_2} - \mathbf{y_2})$
$(a\mathbf{v_1}.\mathbf{u_2})\mathbf{u_2} = \mathbf{0}$
$(b\mathbf{v_2}.\mathbf{u_2})\mathbf{u_2} = \mathbf{by_2}$

Replacing them and the value of $\mathbf{v_3}$ in the above equation we get:

$$\mathbf{y_3} = \mathbf{av_1} + \mathbf{bv_2} - (\mathbf{av_1} + \mathbf{bv_2} - \mathbf{by_2} + \mathbf{0} + \mathbf{by_2}) = \mathbf{0}$$

Thus $\mathbf{y_3}$, comes out to be zero, when we encounter the linearly dependent vector. (Answer).

(b) The classical pseudoCode of Gram-Scmidt is as below:

Classical-GS

for k=1:n
  w = a_k
  for j=1:k-1
    $r_{jk} = q_j^T w$
  end
  for j=1:k-1
    w = w - $r_{jk} q_j$
  end
  $r_{kk} = ||w||$
  $q_k = w/r_{kk}$
end

Modified-GS

for k=1:n
  w = a_k
  for j=1:k-1
    $r_{jk} = q_j^T w$
    $w = w - r_{jk} q_j$
  end
  $r_{kk} = ||w||$
  $q_k = w/r_{kk}$
end

Suppose at the k'th iteration the orthonormal q's already calculated are $Q_{k-1} = [q_1, q_2, \ldots, q_{k-1}]$. In the **classical** case, we first calculate the projections, so suppose the calculated values of the projections in the

k'th iteration are

$$[r_{ik}, r_{2k}, \ldots, r_{k-1,k}]$$

where, $r_{jk} = q_j^T w$, where w in the k'th iteration is initialised to a_k, the vector whose orthogonal transformation is done in the k'th step. Then, these projections along the corresponding orthonormal directions are subtracted from w, in the second inner loop, resulting in:

$$w = w - r_{1k}\mathbf{q_1} - \mathbf{r_{2k}q_2} - \cdots - \mathbf{r_{k-1,k}q_{k-1}} \tag{3}$$

In the modified version, instead of precomputing all the r_{jk}'s at a time, we compute r_{jk} and substract from w, thus always *orthogonalizing* against the currently computed version. Suppose, we are in the k'th iteration, and j=1 results in the following computation

$$w_1 = w - r_{1k}q_1 = w - (q_1^T w)q_1$$

Then for j=2 in the k'th iteration:

$$w_2 = w_1 - r_{2k}q_2 = w - (q_1^T w)q_1 - q_2^T(w - (q_1^T w)q_1)q_2$$
$$w_2 = w_1 - r_{2k}q_2 = w - (q_1^T w)q_1 - (q_2^T.w)q_2 - 0$$
$$w_2 = w_1 - r_{2k}q_2 = w - r_{1k}\mathbf{q_1} - \mathbf{r_{2k}q_2} - \mathbf{0}$$

And similarly for increasing j's(j limits to k) it holds true since $q_i and q_j$ are orthonromal for $i \neq j$. Thus , in exact arithmatic the modified and classical version are numerically the same. (Proved).

(c)

4: Chapter-7, Question-9

For the iterative scheme, we have $x_{k+1} = x_k + \alpha_k p_k$, where, p_k is the search direction and α_k is the step size. This includes the basic statioanry methods as well of the form, $x_{k+1} = x_k + M^{-1}r_k$.
(a) Now, consider the given iterative scheme:

$$x_{k+1} = x_k + \alpha(b - Ax_k)$$

Since, $r_k = b - Ax_k$ is the residual at the k'th step, and in gradient descent , the search direction is in the reverse direction of the residual, Hence, $p_k = r_k$. Hence, for a fixed α we get:

$$M^{-1} = \alpha I$$

$$M = (\alpha I)^{-1}$$

(answer).
The iteration matrix(T) is defined as

$$T = I - M^{-1}A = I - \alpha IA = I - \alpha A$$

(answer).

(b) (i)Given that A is symmetric positive definite and its eigen values follows the inequality:

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0 \tag{4}$$

Scaling equation(4) by α, we get the following

$$\alpha\lambda_1 > \alpha\lambda_2 > \cdots > \alpha\lambda_n > 0$$

$$-\alpha\lambda_1 < -\alpha\lambda_2 < \cdots < -\alpha\lambda_n < 0$$

$$1 - \alpha\lambda_1 < 1 - \alpha\lambda_2 < \cdots < 1 - \alpha\lambda_n < 1 \tag{5}$$

Since, $T = I - \alpha A$, then for an eigen value λ_i of A, their will be a corresponding eigen value of T as $(1 - \alpha\lambda_i)$. Thus, equation-5 gives the eigen values of T and the order of their relative magnitudes.

The theorem for Statioanry method converegence says, that if the spectral radius of the iteration matrix is less than 1, then the system converges. The spectral radius of a matrix is defined as :
$\rho(B) = \max|\lambda_i| : \lambda_i$ are the eigen values of B
Equation(5) shows that the spectral radius of T will be $|1 - \alpha\lambda_n|$ and for convergence it should be less than 1. Hence,

$$-1 < 1 - \alpha\lambda_n < 1$$

$$0 < \alpha < \frac{2}{\lambda_n} \tag{6}$$

(condition on α for convergence).

(ii) The condition number of a symmetric positive definite matrix,A, is given by:

$$K(A) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{\lambda_1}{\lambda_n}$$

$$\lambda_n = \frac{\lambda_1}{K(A)}$$

and substituting this in equation(6)

$$0 < \alpha < \frac{2K(A)}{\lambda_1} \tag{7}$$

(Condition on α in terms of the condition number)

(iii) For the statioanry method, at the i'th ieration and for iteration matrix T, the following error relation holds:

$$||e_i|| \leq ||T||^i e_0$$

Also, for the steepest descent at the i'th iteration we have the following relation,

$$||e_i|| \leq \left(\frac{k-1}{k+1}\right)^i e_0$$

where k is the condition number. Since steepest descent is the one that converegs at the best speed for the class of gradient descent, we can equate these two relations to get:

$$||T|| = \frac{k-1}{k+1}$$

$$(1 - \alpha\lambda_n) = \frac{k-1}{k+1}$$

$$\alpha\lambda_n = \frac{2}{k+1} = \frac{2\lambda_n}{\lambda_1 + \lambda_n}$$
$$\alpha = \frac{2}{\lambda_1 + \lambda_n}$$

(Best value for α for maximizing speed) .

(c) "If A is strictly diagonally dominant and $\alpha = 1$, then the iterative scheme converegs to the solution for any initial guess."
Since, $\alpha = 1$, our M is basically the identity matrix. Note that, for any matrix equation, Ax=b, we can scale each row of A and corresponding element in B by the corresponding diagonal element of that row such that the system of equation remains unalterted. By performing such an operation we can ensure that all the diagonal elements are 1, but the diagonal dominance of the original matrix still remains. Since , our M is the identity matrix, and the diagonal matrix of A is also the identity matrix, this means in our problem we chose the splitting of M to be the diagonal matrix, which means we are performing the Jacobi iteration. Thus it just remains to prove that Jacobi iterations converge if the matrix is diagonally dominant.
Proof:
By defintition of diagonal dominance, we have:

$$|a_{ii} > \sum_{j \neq i} |a_{ij}|$$
$$\sum_{j \neq i} \frac{a_{ij}}{a_{ii}} < 1$$

For Convergence, we are required to show that $||M^{-1}N|| < 1$. Here we pick the infinity norm:

$$||G||_\infty = ||M^{-1}N|| = ||D^{-1}(L+U)||_\infty = max_{1 \leq i \leq m} \sum_{j \neq i} \frac{|a_{ij}|}{a_{ii}} < 1$$

Hence Jacobi converges, and by consequence of that, if a Matrix is stricly diagonally dominant and $\alpha = 1$, then the iterative scheme converges to the solution for any initial guess. (Proved).

5: Chapter-7, question-12

Given a linear system , Ax=b, where A is symmetric. Suppose M-N is a splitting of A, where M is symmetric positive definite and N is symmetric.
Since, M is symmetric Positive Definite, hence we can write the condition number of A as,

$$K(A) = ||M||||M^{-1}|| \frac{\lambda_{max}(M)}{\lambda_{min}(M)}$$

or, from here we get:

$$||M^{-1}|| = \frac{\lambda_{max}(M)}{\lambda_{min}(M) \times ||M||} \tag{8}$$

Now, for an iterative scheme with a splitting of A=M-N, we have:

$$x_{k+1} = M^{-1}Nx_k + M^{-1}b$$

This converegs if the norm of $M^{-1}N$ is less than 1.

$$||M^{-1}N|| \leq ||M^{-1}||||N|| = \frac{\lambda_{max}(M)}{\lambda_{min}(M) \times ||M||} ||N||$$

Now the norm of a symmetric positive definite matrix is the magnitude of its highest eigen-vector. Hence, $||M|| = \lambda_{max}$, and hence they cancel each other. Since, it is given that $\lambda_{min}(M) > \rho(N)$, and for a symmetric matrix N, $\rho(N) \leq ||N||$. Thus, we get the following:

$$||M^{-1}N|| \leq \frac{||N||}{\lambda_{min}(M)} \leq \frac{\rho(N)}{\lambda_{min}(M)} < 1$$

Hence, the system converegs if $\rho(N) < \lambda_{min}(M)$ (Proved).

6: GMRES —-

7: Chapter-8, question-7

(a) Given a column stochastic matrix,P, of size $n \times n$ whose all entries are non-negative and each column sum to 1.

$$A(\alpha) = \alpha P + (1 - \alpha)E$$

For this matrix,A, the entry at its i'th row and j'th column will be

$$a_{ij} = \alpha p_{ij} + \frac{1 - \alpha}{n}$$

Then the sum of the elements of its j'th column will be:

$$\sum_{i=1}^{n} a_{ij} = \sum_{i=1}^{n} (\alpha p_{ij} + \frac{1 - \alpha}{n})$$

$$\sum_{i=1}^{n} a_{ij} = \alpha \sum_{i=1}^{n} p_{ij} + \frac{1 - \alpha}{n} \sum_{i=1}^{n} 1$$

Since P is also stochastic, hence $\sum_{i=1}^{n} p_{ij} = 1$, thus we get:

$$\sum_{i=1}^{n} a_{ij} = \alpha + \frac{1 - \alpha}{n} \times n = 1$$

Thus, the sum of elements of a column in A is also 1. Hence, A is also stochastic. (Proved).

(b) A column stochastic matrix, or a Markov matrix,'A', has an eigen value of 1 and all others are less than 1. So, $\lambda = 1$, is the dominant eigen value. If A is column stochastic, then A^T is row-stochastic, thus the sum of the elements of each row of A^T is 1. So:

$$A^T \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

Hence, A^T has an eigen value of 1. Since the determinants of $det(A - \lambda_n I)$ and $det(A^T - \lambda_n I)$ are the same, hence A and A^T has the same eigen value. Thus, A has an eigen value of 1.
Now, suppose A has an eigen vector , say v, whose eigen-value λ is greater than 1, that is, $|\lambda| > 1$. This implies that $A^n v = \lambda^n v$ has exponentially growing length for large $n -> \infty$. This further implies that there is a large coefficient $[A^n]_{i,j}$ which is larger than 1, since A is non-negative. Since, matrix multiplication of two stochastic matrices results in stochastic matrix, hence as A is stochastic, so A^n is also stochastic which indicates that all its entries has to be ≤ 1, providing a contradiction. So, all eigen values other than 1, has to be less than 1. Hence,

the dominant eigen value is 1(Proved).
Let the dominant eigen-vector be v, and its corresponding eigen value is 1.
Hence,

$$A(\alpha)v = v = \alpha Pv + (1-\alpha)Ev$$
$$(I - \alpha P)v = (1-\alpha)Ev$$

Now, Ev = $\begin{bmatrix} 1/n & 1/n & \dots & 1/n \\ 1/n & 1/n & \dots & 1/n \\ \dots & \dots & \dots & \dots \\ 1/n & 1/n & \dots & 1/n \end{bmatrix}$ **v =** $\frac{1}{n}\begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$**, as the eigen-vectors are also stochastic vectors.**
Substituing them in the original equation, we get:

$$(I - \alpha P)v = (1-\alpha)\frac{1}{n}\begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

v = $\frac{1-\alpha}{n}(I - \alpha P)^{-1}\begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$ **This is the dominant eigen-vector of** $A(\alpha)$**.**

(c) Suppose the second dominant eigen value of A is λ_2 **and its corresponding eigen-vector is x. Then:**

$$\alpha Px + (1-\alpha)Ex = \lambda_2 x \tag{9}$$

We will come back to equation(9) after proving the following lemma.
lemma-1: If x_i **is the eigen vector of A with eigen value** λ_i **and** y_j **is the eigen vector of** A^T **with eigen value** λ_j**, then if** $\lambda_i \neq \lambda_j$**, then** $x_i^T y_j = 0$
proof: We can write the following :

$$y_j^T A = \lambda_j y_j^T$$
$$Ax_i = \lambda_i x_i$$

Transposing and then Multiplying the first by y_j **and the second by** x_i^T**, we get**

$$x_i^T A^T y_j = \lambda_i x_i^T y_j$$
$$x_i^T A^T y_j = \lambda_j x_i^T y_j$$

Subtracting the above leads to the fact that is $\lambda_i \neq \lambda_j$**, then:**

$$x_i^T y_j = 0$$

(Proved)

Coming back to our original equation(9), notice that E's columns are vectors $e^T = \frac{1}{n}[1, 1, \dots, 1]$**. This all 1 vector is an eigen-vector of** A^T **and corresponds to eigen-value of 1. Thus, applying the lemma we just proved, we then get** $Ex_2 = 0$**. Now the equation(9) becomes,**

$$\alpha Px_2 = \lambda_2 x_2$$
$$Px_2 = \frac{\lambda_2}{\alpha}x_2$$

Since, P is also stochastic and the x_2 **is also an eigen vector of P and 1 is the dominant eigen vector, hence:**

$$\frac{\lambda_2}{\alpha} \leq 1$$

$$\lambda_2 \leq \alpha$$

Thus, the second largest wigenvalue of $A(\alpha)$ is upper bounded by α.

(d) (i) $A(\alpha)$ contains to essential components , one of them $P(\alpha)$ is large and sparse, while E is large and sense. Since P is very sparse, it can have a compresses row or compressed column representation and we perform matvec operations on this compresses structure over only the non-zero entries. Additionally, we do not require any storage for the last row entry of the matrix A. This is because, Since A is stochastic, the column sums of A are 1, hence the last element of the matvec will be all the previous elements of the result subtracted from the column sum of the vector. So, at every step of the sparse matrix vector product, we keep accumulating the sum, until we reach the last row, where-in we completely abandon the computation and subtract this sum from the column sum of the vector, and that gives us the last element.
The next saving we do for A, is by utilizing the structure of E. Note that although E is large and dense, it is just a rank-1 matrix. Hence , multiplying a vector, $v = [v_1, v_2, \ldots, v_n]$ to E, results in the following:

$$\begin{bmatrix} 1/n & 1/n & \ldots & 1/n \\ 1/n & 1/n & \ldots & 1/n \\ \ldots & \ldots & \ldots & \ldots \\ 1/n & 1/n & \ldots & 1/n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \ldots \\ v_n \end{bmatrix} = \begin{bmatrix} \frac{v_1 + v_2 + \cdots + v_n}{n} \\ \frac{v_1 + v_2 + \cdots + v_n}{n} \\ \ldots \\ \frac{v_1 + v_2 + \cdots + v_n}{n} \end{bmatrix} = \sum_{i=1}^{n} v_i . \frac{1}{n} \begin{bmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{bmatrix}$$

Thus every multiplication with E, results in a scalar value that scales the all 1 column space, and this value is essentially the sum of the elements of the vector. Thus, instead of computing the dense matrix vector multiplication of $O(n^2)$, we only require summing the vector elements which is $O(n)$.

(ii) If the initial guess v_0 satisfies $||v_0||_1 = 1$, and supposing v_0 is non-negative , that means v_0 is a column stochastic vector. Since A is a stochastic matrix, if we show that the matrix vector product of a stochastic matrix and a stochastic vector results in a stochastic vector, then we can conclude that the result of each iteration of the power-method will result in a stochastic vector whose l1-norm will be 1 and hence normalization will not be required.
Proof: Let A be an $n \times n$ stochastic matrix and v be a $n \times 1$ stochastic vector.
Then, $\sum_{i=1}^{n} a_{ij} = 1$, for all j=1,2,... and $\sum_{j=1}^{n} v_j = 1$, then
Av = $\begin{bmatrix} \sum_{j=1}^{n} a_{1j}.v_j \\ \sum_{j=1}^{n} a_{2j}.v_j \\ \ldots \\ \sum_{j=1}^{n} a_{2j}.v_j \end{bmatrix}$ **Then, $||Av|| =$**

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}.v_j = \sum_{j=1}^{n} (\sum_{i=1}^{n} a_{ij}).v_j = \sum_{j=1}^{n} v_j = 1$$

Thus, the result of the multipication of a stochastic vector by a stochastic matrix results in a stochastic vector whose l1-norm is 1. (Proved).

8: chapter-8,question-10

Given the least squares problem

$$min_x ||b - Ax||_2$$

where A is ill conditioned. Considering the regularization approach that replaces the the normal equations by the modified, better conditioned system.

$$(A^TA + \gamma I)x_\gamma = A^Tb \tag{10}$$

(a)

$$K_2(A^TA + \gamma I) = \frac{\lambda_{max} + \gamma}{\lambda_{min} + \gamma}$$
$$= \frac{\lambda_{max}(1 + \frac{\gamma}{\lambda_{max}})}{\lambda_{min}(1 + \frac{\gamma}{\lambda_{min}})}$$

Now, since, $\lambda_{max} > \lambda_{min}$**, hence** $\frac{1}{\lambda_{max}} < \frac{1}{\lambda_{min}}$**, and using this in the previous equation , we get**

$$K_2(A^TA + \gamma I) = \frac{\lambda_{max}(1 + \frac{\gamma}{\lambda_{max}})}{\lambda_{min}(1 + \frac{\gamma}{\lambda_{min}})} \leq \frac{\lambda_{max}}{\lambda_{min}} = K_2^2(A)$$

(Proved).

(b) We will reformulate the equations for x_y **as a least square problem. Equation-10 can be written as** $Cx_\gamma = D$**, and to translate it to a least square problem, we would like to minimize the residual, that is** $||D - Cx_\gamma||$**. We ca summarize the following data for relations between C,D,A and b.**
We can write , $A = U\Sigma V^T$ **and correspondingly** $A^T = V\Sigma U^T$**, where U,V are orthogonal matrices and** Σ **is a diagonal matrix of singular values. Then** $A^TA = V\Sigma^2V^T$ **and** $AA^T = U\Sigma^2U^T$**. We have defined** $C = (A^TA + \gamma I)$**, which can be further broken into:**

$$C = A^TA + \gamma I = V\Sigma^2V^T + \gamma I = V(\Sigma^2 + \gamma I)V^T$$

and $D = A^Tb = (V\Sigma U^T)b$

Now, relacing C and D into the minimization form, we get

$$||D - Cx_\gamma|| = ||A^Tb - (A^TA + \gamma I)x_\gamma|| = ||(V\Sigma U^T)b - V(\Sigma^2 + \gamma I)V^Tx_\gamma||$$

Since V is orthogonal, we can safely multiply by V^T **without changing the norms. Hence,**

$$||V^T(V\Sigma U^T)b - V^TV(\Sigma^2 + \gamma I)V^Tx_\gamma|| = ||(\Sigma U^T)b - (\Sigma^2 + \gamma I)V^Tx_\gamma||$$

This can be written in the form:

$$||z - \Sigma' y||$$

where, $\mathbf{z} = U^Tb$**,** $\mathbf{y} = V^Tx_\gamma$**, and** $\Sigma' = \Sigma^2 + \gamma I$**.If the rank of the matrix is** $r < n$**, set** $y_i = 0$**, for** $i = r+1, r+2, \ldots, n$ **and set the other** y_i **for** $0 \leq i \leq r$**, as:**

$$y_i = \frac{z_i}{((\sigma_i)')} = \frac{\sigma_i.{u_i}^Tb}{(\sigma_i)^2 + \gamma}$$

Then compute $\mathbf{x}_\gamma = \mathbf{Vy}$**.** x_γ **in terms of the columns** u_i **of U and** v_i **of V can be written as :**

$$x_\gamma = \sum_{i=1}^{r} \frac{\sigma_i {u_i}^Tb}{{\sigma_i}^2 + \gamma} = V\Sigma'' U^T$$

where,

$$\Sigma'' = 0, \sigma_i = 0$$
$$\Sigma'' = \frac{\sigma_i}{{\sigma_i}^2 + \gamma}, \sigma_i \neq 0$$

(c)

Then,

$$||x_\gamma|| = ||V\Sigma'' U^T|| = ||\Sigma''||$$

Since V and U are orthogonal

Now:

$$\frac{\sigma_i}{{\sigma_i}^2 + \gamma} = \frac{1}{\sigma_i + \frac{\gamma}{\sigma_i}} \leq \frac{1}{\sigma_i} \tag{11}$$

Since , for the original equation, the final x is written as $x = V\Sigma^* U^T$, where the Σ^* is defined as

$$\Sigma^* = 0, \sigma_i = 0$$
$$\Sigma^* = \frac{1}{\sigma_i}, \sigma_i \neq 0$$

Similarly,

$$||x|| = ||E^*||$$

Then , due to equation(11), we have $||E''|| \leq ||E^*||$. Hence, we get:

$$||x_\gamma|| \leq ||x|| \tag{12}$$

(Proved)

(d) For the original equation we have

$$A^T Ax = A^T b$$
$$Bx = A^T b$$
$$x = B^{-1} A^T b$$

where $B = A^T A$ For the regularized equation , we have

$$(A^T A + \gamma I)x_\gamma = A^T b$$
$$B^{-1}(B + \gamma I)x_\gamma = (B^{-1})A^T b$$
$$x_\gamma + B^{-1}\gamma x_\gamma = x$$
$$x - x_\gamma = B^{-1}\gamma x_\gamma \tag{13}$$

Taking the norms we get:

$$||x - x_\gamma|| \leq ||B^{-1}||||\gamma||||x_\gamma|| \leq ||B^{-1}||||\gamma||||x||$$

Now, $||B^{-1}|| = ||(A^T A)^{-1}|| = \frac{1}{\sigma_{min}^2}$, hence:

$$\frac{||x - x_\gamma||}{||x||} \leq \frac{\gamma}{\sigma_{min}^2}$$

(answer).

To guarantee that the relative error is bounded by ϵ, we can write:

$$\frac{\gamma}{\sigma_{min}^2} \leq \epsilon$$

$$\gamma \leq \epsilon \times \sigma_{min}^2$$

(e) The matlab code for this part can be found as Prob8.m. Comparing the result for $\gamma = 0$ with that of example-8.8, the svd computation differs from matlab's backslash(using QR) by 0.0002 for $||x||$ since we get $||x|| = 0.9473$ while book's answer is 0.9471. Note here, that for non-zero small γ values in order of $10^{-6}, 10^{-12}$ we get the similar result as reported in example-8.8, which points towards improved accuracy with slightly perturbed γ. The residual norm comes to be the same for $\gamma = 0$ and that of the book's example.

(f) For ill-conditioned problems, where the condition number is very large, indicating a large ratio of the max and min singular values, it is difficult to tune the truncation of the svd. Instead, using the regularization approach, as we derived in (d) of this question, there is a nice way to bound the relative error, using the γ as a knob for tuning and hence provides better control on the relatove error.