

# Mathematical Techniques for Computer Science Applications

CSCI-GA.1180-001

New York University, Fall 2016

Instructor: Margaret Wright, mhw@cs.nyu.edu

Grader: Abhishek Nilesh Shah, ans556@nyu.edu

Homework Assignment 4

Assigned Thursday 27 October 2016; due 11:59pm, Friday 4 November 2016

*Whenever calculations are needed to solve a problem, those calculations must be submitted as part of the homework assignment.*

*Homework must be submitted electronically, by 11:59pm on the due date. Unless express permission has been given in advance by the instructor for a late homework submission, a 30% percent penalty will be deducted for each late day (or part of a late day).*

*Homework grades are based on the correctness, quality, and clarity of your explanations and proofs.*

**Exercise 4.1.** Consider the linear least-squares problem of minimizing  $\|b - Ax\|_2^2$ , where  $A$  is a nonzero  $m \times n$  matrix with  $\text{rank}(A) = n$ , and  $b$  is nonzero.

- If the optimal residual  $b - Ax$  is zero and the optimal  $x$  is nonzero, what does this imply about the representation of  $b$  as a linear combination of vectors in the range space of  $A$  and the null space of  $A^T$ ? Explain your answer.
- If the optimal solution  $x$  is zero and the optimal residual is nonzero, what does this imply about the representation of  $b$  as a linear combination of vectors in the range space of  $A$  and the null space of  $A^T$ ? Explain your answer.
- Let  $b_1$  and  $b_2$  be nonzero  $m$ -vectors such that  $b_1 \neq b_2$ . Let  $x_1$  denote a solution of the linear least-squares problem involving  $A$  and  $b_1$ , with a similar meaning for  $x_2$  and the least-squares problem involving  $A$  and  $b_2$ . If  $x_1 = x_2$ , what does this imply about the relationship between  $b_1$  and  $b_2$ ? Explain.

**Exercise 4.2.** Let  $v$  be any nonzero  $n$ -vector. The associated Householder matrix is defined as

$$H(v) = I - \frac{2vv^T}{\|v\|_2^2}.$$

- Show that, for any nonzero  $v$ , its associated Householder matrix is orthogonal.
- Show that, for any scalar  $\alpha \neq 0$ ,  $H(\alpha v) = H(v)$ .
- For any two nonzero  $n$ -vectors  $a$  and  $b$  with  $a \neq b$  and  $\|a\|_2 = \|b\|_2$ , assume that there is an  $n$ -vector  $u$  such that

$$\|u\|_2 = 1 \quad \text{and} \quad Ha = (I - 2uu^T)a = b, \quad \text{where } H = H(u).$$

Give an expression for  $u$  in terms of  $a$  and  $b$ .

- Let

$$a = \begin{pmatrix} \frac{9}{2} \\ -1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} \frac{7}{2} \\ -3 \end{pmatrix},$$

so that  $\|a\|_2^2 = \|b\|_2^2 = 85/4$ . Give the explicit vector  $u$  defined in part (c), including an explanation of how you computed it.

- Write down the explicit  $2 \times 2$  Householder matrix associated with the vector  $u$  from part (d), and confirm numerically that  $Ha = b$ .

**Exercise 4.3.** Let  $A$  be an  $m \times n$  matrix with  $m > n$  and  $\text{rank}(A) = n$ . Assume that the singular value decomposition of  $A$  is  $USV^T$ , where  $U$  is  $m \times m$  and orthogonal,  $V$  is  $n \times n$  and orthogonal, and  $S$  is an  $m \times n$  “diagonal” matrix of the form

$$S = \begin{pmatrix} S_n \\ 0 \end{pmatrix},$$

where  $S_n$  is a nonsingular  $n \times n$  diagonal matrix  $S_n = \text{diag}(\sigma_1, \dots, \sigma_n)$ , with  $\sigma_1 \geq \dots \geq \sigma_n > 0$ .

- (1) Using the pseudoinverse of  $A$  and the fact that any  $m$ -vector  $b$  can be expressed as a linear combination of the columns of  $U$ , mathematically express the solution  $x$  of the least-squares problem  $\min \|b - Ax\|_2^2$  as a *linear combination of the columns of  $V$* , i.e., specify the coefficients in the linear combination.
- (2) In Matlab calculations for this problem, use `format long e`. Let

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 + 10^{-6} \\ 1 & 1 + 10^{-6} \end{pmatrix}, \quad b_1 = \begin{pmatrix} 1 \\ 2.22474 \\ -0.22474 \end{pmatrix} \quad \text{and} \quad b_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix},$$

where  $\|b_1\|_2 = 2.44948487$  and  $\|b_2\|_2 = 2.44948974$  (both rounded to 9 figures), so that  $\|b_1 - b_2\|_2$  is about  $4.87 \times 10^{-6}$ .

- (a) Give  $U$ ,  $S$ , and  $V$  from the computed SVD of  $A$ . Express  $b_1$  and  $b_2$  numerically as linear combinations of the columns of  $U$ , i.e., give the computed values of the coefficients in the linear combinations.
- (b) Give the computed matrix  $A^\dagger$ , the pseudoinverse of  $A$ , obtained from the matrices  $U$ ,  $V$ , and  $S$ . (Do not use the Matlab command `pinv` except to check your result.)
- (c) Let  $x_1$  be the least-squares solution of  $\min \|b_1 - Ax\|_2^2$ . Give  $x_1$ ,  $\|x_1\|_2$ , and the ratio  $\|x_1\|/\|b_1\|$ . Express  $x_1$  as a linear combination of the columns of  $V$ , i.e., give the coefficients in the linear combination, both in mathematical form and as calculated numbers.
- (d) Let  $x_2$  be the least-squares solution of  $\min \|b_2 - Ax\|_2^2$ . Give  $x_2$ ,  $\|x_2\|_2$ , and the ratio  $\|x_2\|/\|b_2\|$ . Express  $x_2$  as a linear combination of the columns of  $V$ , i.e., give the coefficients in the linear combination, both in mathematical form and as calculated numbers.
- (e) Comment on the difference in norm between  $x_1$  and  $x_2$  relative to the difference in norm between  $b_1$  and  $b_2$ . Explain this difference using your results from part (a).

**Exercise 4.4.** Let  $A$  be an  $m \times n$  matrix with  $m < n$  such that  $\text{rank}(A) = m$ . Assume that  $A$  can be written as

$$A = \begin{pmatrix} B & S \end{pmatrix}, \quad \text{where } B \text{ is nonsingular.}$$

- (a) Given  $A$  in the form above, let  $Z$  be defined as

$$Z = \begin{pmatrix} -B^{-1}S \\ I_{n-m} \end{pmatrix}, \quad \text{where } I_{n-m} \text{ is the } (n-m) \times (n-m) \text{ identity.}$$

Show that (i)  $AZ = 0$  and that (ii) the columns of  $Z$  are linearly independent. (Which means that the columns of  $Z$  form a basis for the null space of  $A$ .)

- (b) When  $Z$  is defined as given above, describe how to form the matrix-products  $Zv$  and  $Z^Tq$  using only  $S$  and the LU factorization of  $B$  (so that the explicit matrix  $B^{-1}$  is not needed).

**Exercise 4.5.** This problem involves a variation on the (extremely silly) model discussed in class, developed by someone from another planet based on looking at pictures of smiling people in *Vogue* magazine. The underlying premise is that the happiest people seem to be tall, thin, and blonde. Assume that the visitor from another planet believes that happiness for a given individual is given by

$$h \approx x_1 \sqrt{t-4} + x_2 \left( \frac{100}{w} \right) + x_3 \exp(-c), \quad (1)$$

where  $t$  is “tallness” (measured in feet),  $w$  is weight (measured in pounds),  $c$  is hair color (where the measured value of  $c$  ranges from 0 to 1, with larger values meaning blonder hair), and  $h$  is happiness. The three coefficients  $x_1$ ,  $x_2$ , and  $x_3$  are unknown.

Suppose that you are given the following observed data for 6 people.

$t_i$	5.2	6.0	5.9	5.6	6.2	5.7
$w_i$	240	162.3	130.8	150.1	95.9	141.2
$c_i$	0.13	0.83	1.0	0.24	0.31	0.47

Assume that you wish to find the optimal coefficients  $x = (x_1, x_2, x_3)^T$  that produce the best fit of the model (1) to a given set of data by minimizing  $\|b - Ax\|_2^2$  for an appropriate matrix  $A$  and right-hand side vector  $b$ .

If using Matlab for this problem, use `format short e` (and the equivalent in other systems).

- Give the *mathematical* form of the entries in row  $i$  of  $A$ , expressed in terms of  $t_i$ ,  $w_i$ , and  $c_i$ .
- Give the specific matrix  $A$  and the value of  $\text{cond}_2(A)$ .
- Assume that the observed values of happiness for the six individuals are

$$h = (9, 8, 7, 10, 12, 9).$$

- Give the vector  $b$  that will appear in the least-squares problem, corresponding to the data above, and give  $\|b\|_2$ .
  - Solve the least-squares problem numerically, and state how you performed this calculation. Matlab will do this with the command `x = A\b`, or you may prefer to form and solve the normal equations explicitly.
  - Give the optimal  $x$ , the vector  $Ax$ , the optimal residual vector  $r = b - Ax$ ,  $\|r\|_2$ , and the ratio  $\|r\|_2/\|b\|_2$ .
  - Based on the numerical results, comment on whether the model seems to you to be effective at predicting the happiness of individuals in this group. Be explicit about how you measure effectiveness.
  - Which individual is predicted to be the happiest? Why (based on his/her individual properties) might this be the case?
  - Which individual is predicted to be the least happy? Why (based on his/her individual properties) might this be the case?
  - Which individual’s observed happiness is closest to the model’s prediction? Why (based on the individual’s properties) might this be the case?
  - Which individual’s observed happiness is farthest from the model’s prediction? Why (based on the individual’s properties) might this be the case?
- (d) Assume that the happiness data were recorded incorrectly, and that the observed happiness values for the same six individuals should be

$$h = (4.5, 6.5, 10, 5.5, 11, 6).$$

- (i) Solve the least-squares problem numerically using the same  $A$  as in part (b), but with the second set of happiness data. Give the vector  $b$  for the second set of happiness data, the optimal  $x$ , the vector  $Ax$ , the optimal residual vector  $r = b - Ax$ ,  $\|r\|_2$ , and the ratio  $\|r\|_2/\|b\|_2$ .
- (ii) Is the model an effective predictor of happiness with this set of data? Explain your opinion using the numerical results.
- (iii) Does anything about the optimal coefficients in this case strike you as strange? If so, please explain and comment on possible causes, making an explicit comparison to the results in part (c).