

You are encouraged to collaborate on the homework and ask for assistance. You are required to write your own solutions, list your collaborators, acknowledge all sources of help, and cite all external references. If you worked alone, write “I did not collaborate.” Failure to follow these guidelines will be considered a breach of academic honesty regulations.

Submit your solution by Wednesday September 24 in class or electronically via the link on the course webpage. Late submissions won’t be accepted.

Question 1

Linear equation solvers can sometimes be used to solve constraint systems that are not exactly linear like:

$$\begin{aligned}x + y &= 2 \text{ or } 6 \\x + 2y &= 1 \text{ or } 7 \\x + 3y &= 0 \text{ or } 2 \\2x + y &= 1 \text{ or } 5 \\3x + y &= 7 \text{ or } 8.\end{aligned}\tag{1}$$

- (a) Express (1) as a *linear* system of equations in the five “variables” x^2 , y^2 , xy , x , and y .
(**Hint:** Write each constraint “ $e = a$ or b ” as $(e - a)(e - b) = 0$ and expand.)
- (b) Treating these variables as independent, solve the system. You may want to use a computer for this. What are x and y ? Verify that they satisfy all constraints in (1).
- (c) Solve (1) using the following alternative method. Take two constraints in (1). For all four possible pairs of values of the right-hand side, find x and y . Among these four, keep the one that is consistent with the other three constraints. Explain your steps.
- (d) Suppose you have a system with 50 unknowns x_1 to x_{50} , and 5000 constraints of the type $a_1x_1 + \dots + a_{50}x_{50} = b$ or c . Which of the two methods do you think is preferable for solving such a system? Justify your answer.

Question 2

In this question you will investigate Gradient Descent on underdetermined linear systems.

- (a) Write down the sum of squares loss for the equation $x + y = 1$ and calculate its gradient.
- (b) Suppose you run gradient descent with rate ρ on part (a). How do the values of $x + y - 1$ and $x - y$ change in each iteration? What is the maximum rate that guarantees convergence?
- (c) To which target (x^*, y^*) does gradient descent in part (b) converge to under initialization $x = 0, y = 0$? How about $x = 1, y = -1$?
- (d) For each of the initializations in part (b), calculate the distance between the state (x_t, y_t) of Gradient Descent at step t and (x^*, y^*) as a function of ρ . At which step does the distance dip below 0.01 when $\rho = 0.1$?
- (e) (**Extra credit**) Prove that, in general, the target $\mathbf{x}^*$ of gradient descent is affected by the choice of initialization if and only if the columns of the input matrix A are linearly dependent.

Question 3

The condition number κ of a linear system is a measure of proximity to linear dependence. In Lecture 2 we argued that it controls the convergence rate of Gradient Descent. You will investigate it in this question.

- (a) The condition number of a PSD matrix S is defined as the ratio between its largest and its smallest eigenvalues. Find the condition number of the matrix

$$S = \begin{bmatrix} 1 & 1 \\ 1 & 1.01 \end{bmatrix}.$$

You may use any method you like, but you must explain how you arrived at your answer.

- (b) The condition number of a general matrix A is defined as the square root of the condition number of the PSD matrix $A^T A$: $\kappa(A) = \sqrt{\kappa(A^T A)}$. Prove that when S is PSD this is consistent with the definition in part (a), i.e., $\kappa(S) = \sqrt{\kappa(S^T S)}$.
- (c) Use part (a) to calculate the condition number of

$$A = \begin{bmatrix} 1 & 0.9 \\ 1 & 1.1 \end{bmatrix}.$$

- (d) Prove that the condition number of a square matrix A is finite if and only if its rows are linearly independent. Use this equivalence to explain qualitatively why the answer in part (c) is so large. You may use the fact that if the rows of A are linearly dependent then so are its columns.

Question 4

Find your personalized hidden parity instance here: <https://andrejb.net/csi4103/hw/25H01.html>

The instance consists of 20 equations in 12 unknowns. $+$ and $-$ stand for the numbers $+1$ and -1 , respectively. You need to find a subset of the columns that multiplies to the right-hand side.

Write down your solution in the form of indices of the relevant columns (in increasing order). For instance, the solution to the example in Section 7 of Lecture Notes 1 is: 2, 4.

Explain clearly how you arrived at your solution. Undocumented computer code will not be entertained as a satisfactory explanation.