

Indian Institute of Technology Madras
Department of Data Science and Artificial Intelligence
DA5000: Mathematical Foundations of Data Science
Tutorial II

Problem

1. Consider the system of equations in 7 variables:

$$\begin{aligned}Q_{\text{in}} &= Q_1, \\Q_1 + Q_{32} &= Q_2, \\Q_2 + Q_{43} &= Q_{32} + Q_3, \\Q_3 &= Q_{\text{out}} + Q_{43}.\end{aligned}\tag{1}$$

Express the system of equations in its matrix representation.

- (a) Obtain the Row reduced echelon form of the matrix.
 - (b) Find the row space, column space, null space and rank of the matrix?
 - (c) Find the basis vectors in the \mathbb{R}^7 space.
 - (d) Are the basis vectors orthogonal to each other? If not, provide an orthonormal basis.
2. Consider a full rank matrix A of size $n \times n$ and b of size $n \times 1$. You need to solve for x in the system $Ax = b$ using the following methods,
- (a) Using the inverse explicitly, $x = A^{-1}b$.
 - (b) Using Gauss Elimination on the augmented matrix, $[A \mid b]$.

Compare the two methods in terms of computational cost/time complexity. Which method is generally preferred, and why?

3. A synthetic data is provided along with this tutorial. It has 10000 observations with 1000 features and one target value for each observation. It can be represented as $y = X\beta$; targets, $y : (10000 \times 1)$, features, $X : (10000 \times 1000)$ and β is the unknown coefficient vector.
- (a) Find the rank of X and $(X \mid y)$, and comment on existence of a solution for β .
 - (b) Find the best possible β that could model this data?