Univ. of Washington

Lecture : Matrix Inverses and Systems of Equations

Winter 2021

Lecturer: Dan Calderone

Systems of Equations

Matrices are used to represent and solve systems of linear equations. Suppose we $A \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ that satisfy.

$$y = Ax \tag{1}$$

Note that this equation is slightly more complicated than it first appears. Depending on the shape of A it may have a unique solution, no solution, or a whole subspace of solutions.

0.1 Unique Solution

The simplest case is that A is square, ie. $x, y \in \mathbb{R}^n$ and the columns are linearly independent. This means there is a unique linear commbination of the columns that reaches every individual point y in the co-domain. We can compute this exact linear combination by doing *Gaussian elimination* also known as *row reduction*. Each step of Gaussian elimination, each *elementary row operation* can be represented by left-multiplication of Equation (1) by a specific type of matrix called *elementary matrices*. These elementary matrices come in three types: row-multiplying, row-swapping, and row-adding demonstrated below

$$\underbrace{\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \alpha & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{multiplying a row by } \alpha}, \quad \underbrace{\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & \alpha & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{adding a row times } \alpha}, \quad \underbrace{\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{swapping rows}} \tag{2}$$

When we perform Gaussian elimination on Equation (1) to transform A into the identity, we left-multiply by the appropriate set of elementary matrices $\{E_1, \ldots, E_k\}$

$$\underbrace{(E_k \cdots E_1)}_{A_l^{-1}} y = \underbrace{(E_k \cdots E_1) A}_{I} x \tag{3}$$

These elementary matrices multiplied together are called the *left-inverse* $A_l^{-1} = (E_k \cdots E_1)$, ie. the matrix that transforms A into the identity by left-multiplying. Note that we could have performed a similar procedure to solve the equation $y^T = x^T A$ except we would multiply on the right by *elementary column matrices*. This procedure would construct the *right inverse* of A, denoted A_r^{-1}. $y^T A_r^{-1} = x^T A A_r^{-1} = x^T$. Assuming A is square and invertible, these two left and right inverses are the same and we simply denote them as $A^{-1} = A_l^{-1} = A_r^{-1}$. This can be seen from

$$A_l^{-1} \cdot A = I \tag{4}$$
$$A_l^{-1} \cdot A \cdot A_r^{-1} = I \cdot A_r^{-1} \tag{5}$$
$$A_l^{-1} = A_r^{-1} \tag{6}$$

0.2 No solution (Least Squares)

If $m > n$, ie. A is "tall", then it is unlikely that there is any solution at all. The columns of A span a subspace of the co-domain called the range of A. There will only be a solution for x if y happens to lie in this subspace. If the columns of A are linearly independent, then A will still have a left-inverse. This is based on the fact that the linear independence of the columns of implies that the matrix $A^T A$ will be invertible. This in turn implies that we can construct a left-inverse as $A_l^{-1} = (A^T A)^{-1} A^T$. Supposing that y is actually in the range of A, ie. there does exist an x solving (1), we can find this x using this left-inverse.

$$\begin{aligned} \text{Assume } y \text{ in range of } A... \qquad & y = Ax \\ & (A^T A)^{-1} A^T y = (A^T A)^{-1} A^T \cdot Ax = x \end{aligned} \tag{7}$$

Now suppose y is not in the range of A. We can still try to find an x that makes Ax as close to y as possible, ie. we can try to minimize

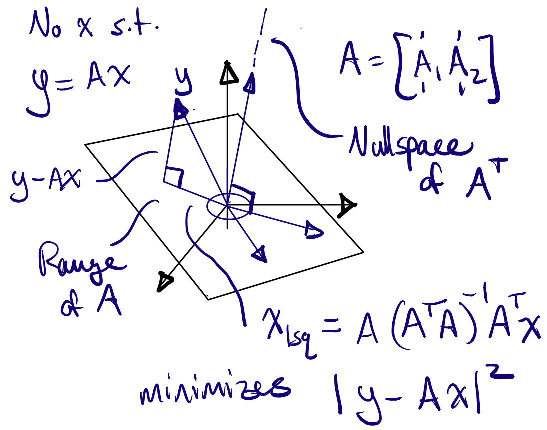
$$|y - Ax|^2 = (y - Ax)^T (y - Ax) = \sum_i (y_i - A_{i:} x)^2 \tag{8}$$

x that minimizes this quantity is called the *least squares solution*, x_{lsq}. It turns out that we can use the left-inverse given above to compute the least-squares solution

$$x_{\text{lsq}} = (A^T A)^{-1} A^T y \tag{9}$$

Note that $A x_{\text{lsq}} = A(A^T A)^{-1} A^T y$ which is the projection of y onto the range of A. We can derive the least squares solution by computing the derivative of (8) and set it equal to 0.

$$\frac{\partial}{\partial x}\Big(y^T y - y^T A x - x^T A y + x^T A^T A x\Big) = -2 y^T A + 2 x^T A^T A = 0 \tag{10}$$
$$\Rightarrow \qquad x = (A^T A)^{-1} A^T y \tag{11}$$



0.3 Subspace/Continuum of Solutions

Suppose $n > m$, ie. A is "fat", and there are more than m linearly independent columns. In . this case, we have more columns than we need to span the space. If we pick any m linear independent columns, we can compute a solution. Suppose the first m columns of A are linearly independent, $A = [\bar{A} \ \cdots]$ where $\bar{A} \in \mathbb{R}^{m \times m}$. We can then compute one solution as $x^1 = [\bar{A}^{-1}y \ \mathbf{0}]^T$ where $\mathbf{0}$ is the appropriate size vector of zeros. The same procedure with different sets of columns produces up to $n - m + 1$ linearly independent solutions which we can organize as the columns of $X = [x^1 \ \cdots \ x^{n-m+1}]$. Note that $A(x^i - x^j) = 0$, ie. $x^i - x^j$ is in the nullspace of A. A basis for the nullspace of A can be computed as the columns of XW where the matrix $W \in \mathbb{R}^{(n-m+1)\times(n-m)}$ is given by $W = [\mathbf{1} - I]^T$ where $\mathbf{1}$ is a vector of ones of the appropriate size. (Note that W computes differences between the columns of X. A different W that computes column differences could be used.) Any solution of (1) has the form

$$x = x^0 + x_{\text{NS}} = x^0 + XWz$$

for some $z \in \mathbb{R}^{n-m}$, ie. any solution consists of some specific solution x^0 plus some component in the nullspace of A. We can compute a specific solution using the method above (selecting m linearly independent columns). However, assuming the rows of A are linearly independent and if we want a specific solution x^0 that is orthogonal to the nullspace of A, then we can select x as a linear combination of the rows of A. Assume x^0 has the form $x^0 = A^T w$ with $w \in \mathbb{R}^m$. Plugging into (1), gives

$$y = AA^T w \qquad \Rightarrow \qquad w = (AA^T)^{-1} y \qquad \Rightarrow \qquad x^0 = A^T(AA^T)^{-1}y \tag{12}$$

Note that x^0 is y times a right-inverse of A. Note also that x^0 is orthogonal to the nullspace of A since $x_{NS}^T A^T(AA^T)^{-1} = 0$. Note also that x^0 computed in this way is the solution with the *minimum 2-norm*. To see this, note that adding some component from the nullspace only increases the square of the 2-norm.

$$|x^0 + x_{\text{NS}}|^2 = (x^0 + x_{\text{NS}})^T (x^0 + x_{\text{NS}}) \tag{13}$$
$$= (x^0)^T x^0 + 2x_{\text{NS}}^T x^0 + x_{\text{NS}}^T x_{\text{NS}} \tag{14}$$
$$= (x^0)^T x^0 + x_{\text{NS}}^T x_{\text{NS}} = |x^0|^2 + |x_{\text{NS}}|^2 \geq= |x^0|^2 \tag{15}$$

Inverse Properties

Properties of inverses:

$P, Q \in \mathbb{C}^{n \times n}$ invertible, and $k \in \mathbb{C}$.

- $(P^{-1})^{-1} = P$
- $(kP)^{-1} = \frac{1}{k} P^{-1}$
- $(PQ)^{-1} = Q^{-1} P^{-1}$
- $\det(P^{-1}) = \frac{1}{\det(P)}$
- $P^{-1} = \frac{1}{\det(P)} \text{Adj}(P)$

Equivalent Inverse Properties:

- P is invertible, ie. P^{-1} exists.
- P^T is invertible
- P can be row reduced to the identity (via Gaussian Elimination (GE))
- P can be column reduced to the identity (via GE).
- P is a product of elementary matrices.
- P (square) is full row rank.
- P (square) is full column rank.
- Columns of P (square) are linearly independent, ie. $Px = 0 \;\Rightarrow\; x = 0$.
- Rows of P (square) are linearly independent, ie. $y^T P = 0 \;\Rightarrow\; y^T = 0$. Rows of P (square) are linearly independent.
- $y = Px$ has a unique solution for each y.
- P has a trivial nullspace. $\mathcal{N}(P) = \{0\}$

- 0 is not an eigenvalue of P.
- $\det(P) \neq 0$.
- There exists Q such that $PQ = QP = I$ ($P^{-1} = Q$).
- P has a left and a right inverse.

Inverse Formulas

- 2×2 inverse

$$P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad P^{-1} = \frac{1}{\det(P)}\text{Adj}(P) = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$
$$= \frac{1}{\det(P)}\Big[\text{Tr}(P)I - P\Big]$$

- 3×3 inverse

$$P^{-1} = \frac{1}{\det(P)}\text{Adj}(P)$$
$$= \frac{1}{\det(P)}\left[\tfrac{1}{2}\Big(\text{Tr}(P)^2 - \text{Tr}(P^2)\Big)I - P\text{Tr}(P) + P^2\right]$$

- **Block Matrix Inversion**

$$P^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

assuming D^{-1} and $(A - BD^{-1}C)^{-1}$ exist or A^{-1} and $(D - CA^{-1}B)^{-1}$ exist.

Proof:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \left(\begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}\right)^{-1}$$
$$= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix}\begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \left(\begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}\right)^{-1}$$
$$= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$$

- **Woodbury Matrix Identity**

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

where $A \in \mathbb{C}^{n\times n}$, $U \in \mathbb{C}^{n\times k}$, $C \in \mathbb{C}^{k\times k}$, and $V \in \mathbb{C}^{k\times n}$. This formula is particularly useful when $n > k$ (U is tall and V is fat). In particular, if U is a column vector, V is a row vector, and C is a scalar, then this equation is called the *Sherman-Morrison Formula*.

- **Neumann Series**

$$A^{-1} = \sum_{n=0}^{\infty}(I - A)^n, \qquad \text{if } \lim_{n\to\infty}(I - A)^n = 0$$

- **Derivative of Inverse**

For $P(t)$

$$\frac{\partial P^{-1}}{\partial t} = -P^{-1}\frac{\partial P}{\partial t}P^{-1}$$