Math Camp II

Basic Linear Algebra

Yiqing Xu

MIT

Aug 26, 2014

Systems of Linear Equations

Definition

A linear equation in the variables $x_1, \ldots x_n$ is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where b and the coefficients $a_1, \ldots a_n$ are real or complex numbers. The subscript n may be any positive integer.

Definition

A system of linear equations (or a linear system) is a collection of one or more linear equations involving the same variables—say, $x_1, \ldots, x_n$.

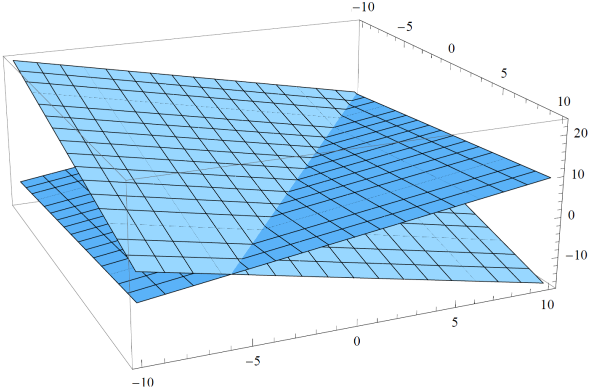
Properties of systems of linear equations

Systems of linear equations fall into one of three camps

- A single solution (the two lines intersect at one point)
- No solution (the two lines are exactly parallel but are shifted by some constant)
- An infinite number of solutions

Graphical View of the Solutions

$$x + y + z = 3$$
$$x + 2y - 3z = 2$$
$$x + y + z = 5$$



Definition Properties of Vector

An ordered list of number in **R**:

$$\mathbf{v} = \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

Properties

- Vector Addition

$$\begin{pmatrix} u_1 + v_1 & u_2 + v_2 & \cdots & u_k + v_n \end{pmatrix}$$

- Scalar Multiplication

$$c\mathbf{v} = \begin{pmatrix} cv_1 & cv_2 & \dots & cv_n \end{pmatrix}$$

- Vector Inner Product

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

- Vector Norm

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1 v_1 + v_2 v_2 + \cdots + v_n v_n}$$

Functions of Vectors

C	P	G	I	V
3	2	4	5	6
6	7	3	3	1
5	6	4	1	5
3	5	5	5	6

$$\mathbf{C} = \beta_1\mathbf{P} + \beta_2\mathbf{G} + \beta_3\mathbf{I} + \beta_4\mathbf{V}$$

Functions of Vectors

Definition (Linear combinations)

The vector $\mathbf{u}$ is a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ if

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

Definition (Linear independence)

A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ is linearly independent if the only solution to the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = 0$$

is $c_1 = c_2 = \cdots = c_k = 0$. If another solution exists, the set of vectors is linearly dependent.

Systems of Linear Equations

We might have a system of m equations with n unknowns

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & \vdots & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

A **solution** to a linear system of m equations in n unknowns is a set of n numbers $x_1, x_2, \cdots, x_n$ that satisfy each of the m equations.

Can we rewrite it some way?

Elementary row (column) operations

Definition

Elementary row (column) operations on an $m \times n$ matrix A includes the following.

1. Interchange rows (columns) r and s of A.
2. Multiply row (column) r of A by a nonzero scalar $k \neq 0$.
3. Add k times row (column) r of A to row (column) s of A where $r \neq s$.

An $m \times n$ matrix A is said to be row (column) equivalent to an $m \times n$ matrix B if B can be obtained by applying a finite sequence of elementary row (column) operations to A.

Elementary equation operations

Elementary equation operations transform the equations of a linear system while maintaining an **equivalent** linear system — equivalent in that the same values of x_j solve the original and transformed systems, i.e., transformed systems look different but have the same properties. These operations are:

1. Interchanging two equations,
2. Multiplying two sides of an equation by a constant, and
3. Adding equations to each other

Vector Space

Definition

A real vector space is a set of “vectors” together with rules for vector addition and for multiplication by real numbers

- It’s a set of vectors
- There are rules

Subspace

Definition

A subspace of a vector space is a set of vectors (including $\mathbf{0}$) that satisfies two requirements: If V and W are vectors in the subspace and c is any scalar, then

1. $V + W$ is in the subspace
2. cV is in the subspace

- All linear combinations stay in the subspace
- Subspace must go through the origin

Questions

- Is a identity matrix a subspace by itself?
- Is the union of a line and a plane (the line is not on the plane) a subspace?
- Is the intersection of two subspaces a subspace?

Column Space

Suppose A is a $m \times n$ matrix.

Definition

The column space consists of all linear combinations of the columns. The combinations are all possible vectors Ax. They fill the column space of A, denoted by $C(A)$.

- The system $Ax = b$ is solvable iff b is in $C(A)$.
- $C(A)$ is a subspace of $\mathbb{R}^m$ (why $C(A)$ is a subspace?)

Spanning

Definition

Subspace SS is a span of S if it contains all combinations of vectors in S

- V is a vector space, e.g., $\mathbb{R}^3$
- S is a set of vectors in V (probably not a subspace)
- SS is the span of S, a subspace of V

Null Space

Suppose A is a $m \times n$ matrix.

Definition

The null space of A, denoted by $N(A)$, consists of all solutions to $Ax = 0$.

- $N(A)$ is subspace of $\mathbb{R}^n$. Why?
- Remember, $C(A)$ is a subspace of $\mathbb{R}^m$
- Any null space must include the origin

Reduced Row Echelon Form

Definition

A rectangular matrix is in echelon form (or row echelon form) if it has the following three properties:

1. All nonzero rows are above any rows of all zeros
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

   If a matrix in echelon form satisfies the following additional conditions, then it is in reduced echelon form (or reduced row echelon form)

4. The leading entry in each nonzero row is 1.
5. Each leading 1 is the only nonzero entry in its column

Pivots

Definition

A pivot position in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A pivot column is a column of A that contains a pivot position

$$\mathbf{A} = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Special Solutions

- Every free column of an echelon matrix leads to a special solution. The free variable equals 1 and the other free variables equal 0. Back substitution solves $Ax = 0$
- The complete solution to $Ax = 0$ is a combination of the special solutions
- If $n > m$ then A has at least one column without pivots, giving a special solution. So there are nonzero vectors x in $N(A)$

Rank

$$\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

- The rank r of A is the number of pivots
- The r pivot columns are not combinations of earlier columns
- The $n - r$ free columns are combinations of the pivot columns

Augmented Matrices

Instead of writing out all of the x terms of each matrix, we can write the **augmented matrix** which is just the matrix representation of $A\mathbf{x} = b$ where $\mathbf{x}$ is a vector $(x_1, x_2, \cdots, x_n)'$. Often times we'll just work with the augmented matrix in RREF, $[\mathbf{R}\ d]$, e.g.,

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

One Particular Solution

1. Choose the free variable(s) to be zero (e.g., $x_3 = 0$)
2. Nonzero equations give the values of the pivot variables (e.g., $x_1 = -2,\ x_2 = 4$)
3. The particular solution in the following case is $(-2, 4, 0)$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The Complete Solution of $Ax = b$

The Complete Solution of $Ax = b$ is

- One particular solution
- Add the null space
- $x_p + x_n$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Rank and the Complete Solution of $Ax = b$

- $A\mathbf{x} = b$ is solvable iff the last $m - r$ equations reduce to $0 = 0$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- Full column rank $r = n$ means no free variables: one solution or none

$$\left[\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{array}\right] \qquad \left[\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{array}\right]$$

- Full row rank $r = m$ means one solution if $m = n$ or infinitely many if $m < n$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 1 & 3 \end{array}\right] \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

Full Column Rank

Every matrix A with full column rank ($r = n$) has the following properties

- All columns of A are pivot columns
- No free variables or special solutions
- $N(A)$ contains only the zero vector $\mathbf{x} = 0$
- $A\mathbf{x} = b$ has zero or only one solution

A few variables, many many equations

$$[\mathbf{R}\ d] = \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{array}\right] = \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{array}\right]$$

Full Row Rank

Every matrix A with full column rank ($r = m$) has the following properties

- All row have pivots
- $A\mathbf{x} = b$ has a solution for every right side b
- The column space is the whole space of $\mathbb{R}^m$
- There are $n - r = n - m$ special solutions in $N(A)$

A few equations, many many variables

$$[\mathbf{R}\ d] = \left[\begin{array}{ccc|c} 1 & 0 & 3 & 2 \\ 0 & 1 & -2 & 1 \end{array}\right]$$

Four Cases

Ranks	RREF	A	$A\mathbf{x} = b$
$r = m = n$	$[I]$	Square and invertible	1 solution
$r = m < n$	$[I \ \ F]$	Short and wide	∞ solutions
$r = n < m$	$\begin{bmatrix} I \\ 0 \end{bmatrix}$	Tall and thin	0 or 1 solution
$r < m, r < n$	$\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$	Not full rank	0 or ∞ solutions

Linear Independence and Spanning, Again

Definition

The columns of a matrix A are linearly independent when the only solutions to $A\mathbf{x} = 0$ is $\mathbf{x} = 0$. No other combination $\mathbf{x}$ of the columns gives the zero vector.

Definition

A set of vectors spans a space if their linear combinations fill the space

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Basis

Definition

A basis for a vector space is a sequence of vectors with two properties:

- The basis vectors are linearly independent
- They span the space

- The columns of every invertible n by n matrix give a basis for $\mathbb{R}^n$, why?
- Can you list one basis for the space of all 2 by 2 matrix?

Dimensions

Definition

The dimension of a space is the number of vectors in very basis

- The dimension of the whole n by n matrix space is n^2
- The dimension of the subspace of upper triangular matrices is $(n^2 + n)/2$
- The dimension of the subspace of diagonal matrices is n
- The dimension of the subspace of symmetric matrices is $(n^2 + n)/2$

- Dimension of column space + dimension of null space = dimension of $\mathbb{R}^n$, why?

In Summary

1. Independent vectors: no extra vectors
2. Spanning a space: enough vectors to produce the rest
3. Basis for a space: not too many, not too few
4. Dimension of a space: the number of vectors in a basis

Matrix

A matrix is an array of mn real numbers arranged in m rows by n columns.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Vectors are special cases of matrices; a column vector of length k is a $k \times 1$ matrix, while a row vector of the same length is a $1 \times k$ matrix.

You can also think of larger matrices as being made up of a collection of row or column vectors. For example,

$$\mathbf{A} = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \end{pmatrix}$$

Matrix Operations

Let A and B be $m \times n$ matrices.

1. (equality) $A = B$ if $a_{ij} = b_{ij}$.
2. (addition) $C = A + B$ if $c_{ij} = a_{ij} + b_{ij}$ and C is an $m \times n$ matrix.
3. (scalar multiplication) Given $k \in \mathbf{R}$, $C = kA$ if $c_{ij} = ka_{ij}$ where C is an $m \times n$ matrix.
4. (product) Let C be an $n \times l$ matrix. $D = AC$ if $d_{ij} = \sum_{k=1}^{n} a_{ik} c_{kj}$ and D is an $m \times l$ matrix.
5. (transpose) $C = A^\top$ if $c_{ij} = a_{ji}$ and C is an $n \times m$ matrix.

Matrix Multiplication

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} aA + bC & aB + bD \\ cA + dC & cB + dD \\ eA + fC & eB + fD \end{pmatrix}$$

The number of columns of the first matrix must equal the number of rows of the second matrix. If so they are **conformable for multiplication**. The sizes of the matrices (including the resulting product) must be

$$(m \times k)(k \times n) = (m \times n)$$

Matrix Algebra Laws

1. Associative
   $(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$; $(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})$
2. Commutative
   $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$
3. Distributive
   $\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$; $(\mathbf{A}+\mathbf{B})\mathbf{C}=\mathbf{AC}+\mathbf{BC}$

Note Commutative law for multiplication does not hold – the order of multiplication matters:

$$\mathbf{AB}\neq\mathbf{BA}$$

Rules of Matrix Transpose

1. $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
2. $(\mathbf{A}^T)^T = \mathbf{A}$
3. $(s\mathbf{A})^T = s\mathbf{A}^T$
4. $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

Square Matrix

Square matrices have the same number of rows and columns; a $k \times k$ square matrix is referred to as a matrix of order k.

Definition

Let A be an $m \times n$ matrix.

1. A is called a square matrix if $n = m$.
2. A is called symmetric if $A^\top = A$.
3. A square matrix A is called a diagonal matrix if $a_{ij} = 0$ for $i \neq j$. A is called upper triangular if $a_{ij} = 0$ for $i > j$ and called lower triangular if $a_{ij} = 0$ for $i < j$.
4. A diagonal matrix A is called an identity matrix if $a_{ij} = 1$ for $i = j$ and is denoted by I_n.

Examples of the Square Matrix

1. **Symmetric Matrix**: A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^T$; this implies that $a_{ij} = a_{ji}$ for all i and j.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = \mathbf{A}^T, \qquad \mathbf{B} = \begin{pmatrix} 4 & 2 & -1 \\ 2 & 1 & 3 \\ -1 & 3 & 1 \end{pmatrix} = \mathbf{B}'$$

2. **Diagonal Matrix**: A matrix $\mathbf{A}$ is diagonal if all of its non-diagonal entries are zero; formally, if $a_{ij} = 0$ for all $i \neq j$

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Examples of the Square Matrix

1. **Triangular Matrix**: A matrix is triangular one of two cases. If all entries below the diagonal are zero ($a_{ij} = 0$ for all $i > j$), it is **upper triangular**. Conversely, if all entries above the diagonal are zero ($a_{ij} = 0$ for all $i < j$), it is **lower triangular**.

$$\mathbf{A}_{LT} = \begin{pmatrix} 1 & 0 & 0 \\ 4 & 2 & 0 \\ -3 & 2 & 5 \end{pmatrix}, \qquad \mathbf{A}_{UT} = \begin{pmatrix} 1 & 7 & -4 \\ 0 & 3 & 9 \\ 0 & 0 & -3 \end{pmatrix}$$

2. **Identity Matrix**: The $n \times n$ identity matrix $\mathbf{I}_n$ is the matrix whose diagonal elements are 1 and all off-diagonal elements are 0. Examples:

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Trace

Definition

If A is an $n \times n$ matrix, then the trace of A denoted by $\mathrm{tr(A)}$ is defined as the sum of all the main diagonal elements of A. That is, $\mathrm{tr(A)} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{a_{ii}}$.

Properties of the trace operator:

If $\mathbf{A}$ and $\mathbf{B}$ are square matrices of order k, then

1. $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
2. $\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}(\mathbf{A})$
3. $\mathrm{tr}(s\mathbf{A}) = s\,\mathrm{tr}(\mathbf{A})$
4. $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$

Matrix Inverse

Definition

Inverse Matrix: An $n \times n$ matrix $\mathbf{A}$ is **invertible** if there exists an $n \times n$ matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

$\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$. If there is no such $\mathbf{A}^{-1}$, then $\mathbf{A}$ is noninvertible (another term for this is "singular").

Matrix Inverse

Theorem (Uniqueness of Inverse)

The inverse of a matrix, if it exists, is unique. We denote the unique inverse of A by A^{-1}.

Theorem (Properties of Inverse)

Let A and B be nonsingular $n \times n$ matrices.

1. *AB is nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$.*
2. *A^{-1} is nonsingular and $(A^{-1})^{-1} = A$.*
3. *$(A^{\top})^{-1} = (A^{-1})^{\top}$.*

Shortcut for 2×2

It turns out that there is a shortcut for 2x2 matrices that are invertible:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ then } A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Solving Linear Systems

Matrix representation of a linear system

$$\mathbf{Ax} = \mathbf{b}$$

If $\mathbf{A}$ is an $n \times n$ matrix,then $\mathbf{Ax} = \mathbf{b}$ is a system of n equations in n unknowns. If $\mathbf{A}$ is invertible then $\mathbf{A}^{-1}$ exists. To solve this system, we can left multiply each side by $\mathbf{A}^{-1}$.

$$\begin{aligned} \mathbf{A}^{-1}(\mathbf{Ax}) &= \mathbf{A}^{-\mathbf{1}}\mathbf{b} \\ (\mathbf{A}^{-1}\mathbf{A})\mathbf{x} &= \mathbf{A}^{-\mathbf{1}}\mathbf{b} \\ \mathbf{I}_n\mathbf{x} &= \mathbf{A}^{-\mathbf{1}}\mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-\mathbf{1}}\mathbf{b} \end{aligned}$$

Hence, given $\mathbf{A}$ and $\mathbf{b}$ and given that $\mathbf{A}$ is nonsingular, then $\mathbf{x} = \mathbf{A}^{-\mathbf{1}}\mathbf{b}$ is a unique solution to this system.

Determinants

If the determinant is non-zero then the matrix is invertible.

- For a 1×1 matrix $\mathbf{A} = a$ the determinant is simply a. We want the determinant to equal zero when the inverse does not exist. Since the inverse of a, $1/a$, does not exist when $a = 0$.

$$|a| = a$$

- For a 2×2 matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, the determinant is defined as $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \det A = a_{11}a_{22} - a_{12}a_{21} == a_{11}|a_{22}| - a_{12}|a_{21}|$

Properties of determinants

1. $|\mathbf{A}| = |\mathbf{A}^T|$
2. Interchanging rows changes the sign of the determinant
3. If two rows of **A** are exactly the same, then $|\mathbf{A}| = 0$
4. If a row of **A** has all 0 then $|\mathbf{A}| = 0$
5. If $c \neq 0$ is some constant then $|\mathbf{A}| = |c\mathbf{A}|$
6. A square matrix is nonsingular iff its determinant is $\neq 0$
7. $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$

Bringing It Together

Theorem

Let A be an $n \times n$ matrix. Then the following statements are equivalent to each other.

- *A is an invertible matrix.*
- *A ir row equivalent to the $n \times n$ identity matrix.*
- *The equation $Ax = 0$ has only the trivial solution.*
- *The equation $Ax = b$ has at least one solution for each b in $\mathbb{R}^n$.*
- *There is an $n \times n$ matrix C such that AC=I.*
- *There is an $n \times n$ matrix D such that DA=I.*
- *A^T is an invertible matrix.*
- *rank(A)=n.*
- *det*$(A) \neq 0$

Geometrics of Ordinary Least Square

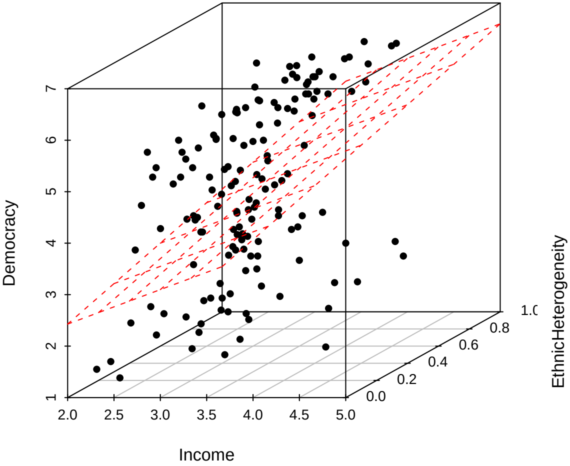
- One way is to think about the OLS is to find (or approximate) solutions to $A\mathbf{x} = \mathbf{b}$, or $X\mathbf{b} = \mathbf{y}$
- X is a n by k matrix
- When $n > k$, a solution does not exist
- Two ways to think about it
    1. Find a vector $\hat{\mathbf{b}}$ such that $X\hat{\mathbf{b}}$ is closest to $\mathbf{y}$
    2. Project $\mathbf{y} \in \mathbb{R}^n$ onto subspace $C(X)$, a k-dimensional subspace in $\mathbb{R}^n$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}$$

First Interpretation

Find a vector $\hat{\mathbf{b}}$ such that $X\hat{\mathbf{b}}$ is closest to $\mathbf{y}$

- $X = [\mathbf{1} \text{ income ethnic}]$
- $X\hat{\mathbf{b}}$ is on a plane (why?)



Projection

If is not super intuitive because humans do not have intuitions for high-dimensional spaces

- We denote the projection of a vector as $\mathbf{p} = P\mathbf{b}$, P is called the projection matrix
- $\mathbf{b} - \mathbf{p}$ is orthogonal to $\mathbf{p}$

Definition

Vectors V and W are orthogonal if $V'W = 0$

Projection onto a Subspace

- **Problem**: Given $x_1, x_2, \cdots, x_k$ are $n \times 1$ vectors (k variables), find the combination $\hat{b}_1 x_1 + \hat{b}_1 x_2 + \cdots + \hat{b}_k x_k$ closest to a given vector $\mathbf{y}$
- We know that the error vector $\mathbf{y} - X\hat{\mathbf{b}}$ is orthogonal to $C(X)$

$$
\begin{aligned}
x_1'(\mathbf{y} - X\hat{\mathbf{b}}) &= 0 \\
&\vdots \\
x_k'(\mathbf{y} - X\hat{\mathbf{b}}) &= 0 \\
&\text{or} \\
X'(\mathbf{y} - X\hat{\mathbf{b}}) &= 0
\end{aligned}
$$

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}$$

$$P = X(X'X)^{-1}X'$$

"Residual Maker" and OLS Mechanics

- The residual maker:

$$M = I - P = I - X(X'X)^{-1}X'$$

- M is both symmetric and idempotent $M^2 = M$ (why?)
- $MX = 0$ (why?)
- $\hat{\mathbf{y}}$ and residual $\mathbf{e}$ are orthogonal (why?)
- $\mathbf{y} = P\mathbf{y} + M\mathbf{y}$
- $\mathbf{y}'\mathbf{y} = \mathbf{y}'P'P\mathbf{y} + \mathbf{y}'M'M\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}$
- $\mathbf{e}'\mathbf{e} = \mathbf{e}'\mathbf{y} = \mathbf{y}'\mathbf{e}$ (why?)
- $\mathbf{y}'P'P\mathbf{y} = \hat{\mathbf{b}}'X'X\hat{\mathbf{b}} = \hat{\mathbf{b}}'X'\mathbf{y} = \mathbf{y}'X\hat{\mathbf{b}}$ (why?)