Linear Algebra Concepts

Ken Kreutz-Delgado
(Nuno Vasconcelos)

ECE 175A – Winter 2012 – UCSD
Vector spaces

• **Definition:** a vector space is a set $\mathcal{H}$ where
  
  – addition and scalar multiplication are defined and satisfy:

  1) $x + (x' + x'') = (x + x') + x''$
  2) $x + x' = x' + x \in \mathcal{H}$
  3) $0 \in \mathcal{H}, 0 + x = x$
  4) $-x \in \mathcal{H}, -x + x = 0$
  5) $\lambda x \in \mathcal{H}$
  6) $1x = x$
  7) $\lambda(\lambda' x) = (\lambda \lambda')x$
  8) $\lambda(x + x') = \lambda x + \lambda x'$
  9) $(\lambda + \lambda')x = \lambda x + \lambda' x$

  ($\lambda = \text{scalar}; \ x, x', x'' \in \mathcal{H}$)

• the canonical example is $\mathbb{R}^d$ with standard
  vector addition and scalar multiplication

• Definition: a vector space is a set $\mathcal{H}$ where
  – addition and scalar multiplication are defined and satisfy:

  1) $x + (x' + x'') = (x + x') + x''$
  2) $x + x' = x' + x \in \mathcal{H}$
  3) $0 \in \mathcal{H}, 0 + x = x$
  4) $-x \in \mathcal{H}, -x + x = 0$
  5) $\lambda x \in \mathcal{H}$
  6) $1x = x$
  7) $\lambda(\lambda' x) = (\lambda \lambda')x$
  8) $\lambda(x + x') = \lambda x + \lambda x'$
  9) $(\lambda + \lambda')x = \lambda x + \lambda' x$

  ($\lambda = \text{scalar}; \ x, x', x'' \in \mathcal{H}$)

• the canonical example is $\mathbb{R}^d$ with standard
  vector addition and scalar multiplication

• Definition: a vector space is a set $\mathcal{H}$ where
  – addition and scalar multiplication are defined and satisfy:

  1) $x + (x' + x'') = (x + x') + x''$
  2) $x + x' = x' + x \in \mathcal{H}$
  3) $0 \in \mathcal{H}, 0 + x = x$
  4) $-x \in \mathcal{H}, -x + x = 0$
  5) $\lambda x \in \mathcal{H}$
  6) $1x = x$
  7) $\lambda(\lambda' x) = (\lambda \lambda')x$
  8) $\lambda(x + x') = \lambda x + \lambda x'$
  9) $(\lambda + \lambda')x = \lambda x + \lambda' x$

  ($\lambda = \text{scalar}; \ x, x', x'' \in \mathcal{H}$)

• the canonical example is $\mathbb{R}^d$ with standard
  vector addition and scalar multiplication
Vector spaces

• But there are much more interesting examples
• E.g., the \textit{space of functions} \( f: \mathcal{X} \rightarrow \mathbb{R} \) with

\[ (f + g)(x) = f(x) + g(x) \quad \text{and} \quad (\lambda f)(x) = \lambda f(x) \]

• \( \mathbb{R}^d \) is a vector space of \textit{finite} dimension, e.g.
  – \( f = (f_1, \ldots, f_d)^T \)
• When \( d \) goes to infinity we have a function
  – \( f = f(t) \)
• The space of \textit{all} functions is an \textit{infinite} dimensional vector space
Data Vector Spaces

• In this course we will talk a lot about “data” and “features”
• Data/features will always be represented in a vector space:
  – an example is really just a point (“datapoint”) on such a space
  – from above we know how to perform basic operations on datapoints
  – this is nice, because datapoints can be quite abstract
  – e.g. images:
    ▪ an image is a function on the image plane
    ▪ it assigns a color \( f(x,y) \) to each each image location \( (x,y) \)
    ▪ the space \( \Psi \) of images is a vector space (note: assumes that images can be negative)
    ▪ this image is a point in \( \Psi \)
Images

- Because of this \textit{we can manipulate images by manipulating their equivalent vector representations}.
- E.g., Suppose one wants to \textit{“morph”} \(a(x,y)\) into \(b(x,y)\):
  - One way to do this is via the path along the line from \(a\) to \(b\).
    \[
    c(\alpha) = a + \alpha (b-a) = (1-\alpha) a + \alpha b
    \]
    - for \(\alpha = 0\) we have \(a(x,y)\)
    - for \(\alpha = 1\) we have \(b(x,y)\)
    - for \(\alpha\) in \((0,1)\) we have a point on the line between \(a(x,y)\) and \(b(x,y)\)
- To morph an image we can simply apply this rule to the image vector representations!
Images

- When we make
  \[ c(x,y) = (1-\alpha) \, a(x,y) + \alpha \, b(x,y) \]
  we get “image morphing”:

- The point is that this is possible because we exploit the structure of a vector space.
Images

- Images are usually approximated as points in $\mathbb{R}^d$
  - **Sample** (*discretize*) an image on a finite grid to get an array of pixels $a(x,y) \rightarrow a(i,j)$
  - Images are always stored like this on digital computers
  - We can now stack all the rows (or columns) into a vector. E.g. a $3 \times 3$ image can be converted into a $9 \times 1$ vector as follows:

  ![Image Conversion Example]

  - In general $n \times m$ image vector is transformed into a $nm \times 1$ vector
  - Note that *this is yet another vector space*

- The point is that there are generally multiple different, *but isomorphic*, vector spaces in which the data can be represented
Another common type of data is **text**.

Documents are represented by **word counts**:

- associate a counter with each word
- slide a window through the text
- whenever the word occurs increment its counter

This is the way search engines represent web pages.
Text

- E.g. word counts for three documents in a certain corpus (only 12 words shown for clarity)

- Note that:
  - Each document is a $d = 12$ dimensional vector
  - If I add two word count vectors (documents), I get a new word count vector (document)
  - If I multiply a word count vector (document) by a scalar, I get a word count vector
  - Note: once again we assume word counts could be negative (to make this happen we can simply subtract the average value)

- This means:
  - We are once again in a vector space (positive subset of $\mathbb{R}^d$)
  - A document is a point in this space
Bilinear forms

- One reason to use *inner product vector spaces* is that they allow us to measure distances between data points.
- We will see that this is **crucial for classification**.
- The main tool for this is the *inner product* ("dot-product").
- We can define the dot-product using the notion of a *bilinear form* (assuming a *real* vector space).

**Definition:** a *bilinear form* on a real vector space $\mathcal{H}$ is a bilinear mapping

$$Q: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$$

$$(x,x') \rightarrow Q(x,x')$$

"Bi-linear" means that $\forall x,x',x'' \in \mathcal{H}$

i) $Q[(\lambda x + \lambda' x'),x''] = \lambda Q(x,x'') + \lambda' Q(x',x'')$

ii) $Q[x'',(\lambda x + \lambda' x')] = \lambda Q(x'',x) + \lambda' Q(x'',x')$
Inner Products

- **Definition**: an *inner product* on a *real* vector space $\mathcal{H}$ is a *bilinear form*

  $$<\cdot,\cdot>: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$$
  $$(x,x') \rightarrow <x,x'>$$

  such that

  i) $<x,x> \geq 0$, $\forall \ x \in \mathcal{H}$
  
  ii) $<x,x> = 0$ if and only if $x = 0$
  
  iii) $<x,y> = <y,x>$ for all $x$ and $y$

- The *positive-definiteness* conditions i) and ii) make the inner product a natural measure of similarity
- This becomes more precise with introduction of a *norm*
Inner Products and Norms

• Any inner product *induces* a **norm** via the assignment

\[ \|x\|^2 = \langle x, x \rangle \]

• **By definition, any norm must** obey the following **properties**
  – Positive-definiteness: \( \|x\| \geq 0, \; \& \; \|x\| = 0 \iff x = 0 \)
  – Homogeneity: \( \|\lambda x\| = |\lambda| \|x\| \)
  – Triangle Inequality: \( \|x + y\| \leq \|x\| + \|y\| \)

• A norm defines a corresponding **metric**

\[ d(x,y) = \|x-y\| \]

which is a **measure of the distance** between \( x \) and \( y \)

• Always remember that the induced norm **changes** with a different choice of inner product!
Inner Product

• Back to our examples:
  – In $\mathbb{R}^d$ the **standard** (or **unweighted**) inner product is
    \[ \langle x, y \rangle = x^T y = \sum_{i=1}^{d} x_i y_i \]
    \[ \|x\| = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{d} x_i^2} \]
  – Which leads to the **standard (unweighted) Euclidean norm** in $\mathbb{R}^d$
  – The distance between two vectors is the **standard (unweighted) Euclidean distance** in $\mathbb{R}^d$
    \[ d(x, y) = \|x - y\| = \sqrt{(x - y)^T (x - y)} = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2} \]
Inner Products and Norms

• Note, e.g., that this immediately gives a measure of similarity between web pages
  – compute word count vector \( x_i \) from page \( i \), for all \( i \)
  – distance between page \( i \) and page \( j \) can be simply defined as:

\[
d(x_i, x_j) = \| x_i - x_j \| = \sqrt{(x_i - x_j)^T (x_i - x_j)}
\]

  – This allows us to find, in the web, the most similar page \( i \) to any given page \( j \), at least with respect to this simple metric.

• In fact, this is very close to the measure of similarity used by most search engines!

• What about norms on function spaces, as used to represent, e.g., images and other continuous valued signals?
Inner Products on Function Spaces

- Recall that the space of functions is an infinite dimensional vector space
  - The standard (unweighted) inner product is the natural extension of that in $\mathbb{R}^d$ (just replace summations by integrals)
    \[
    \langle f(x), g(x) \rangle = \int f(x)g(x)dx
    \]
  - The norm is related to the “energy” of the function
    \[
    \| f(x) \|^2 = \int f^2(x)dx
    \]
  - The distance between functions is related to the energy of the difference between them
    \[
    d(f(x), g(x)) = \| f(x) - g(x) \|^2 = \int [f(x) - g(x)]^2 dx
    \]
Basis Vectors

• We know how to measure distances in a vector space

• Another interesting property is that we can usually fully characterize a vector space by one of its *bases*

• A *set* of vectors \( x_1, \ldots, x_k \) is a *basis* of a vector space \( \mathcal{H} \) if and only if (iff)
  – they are *linearly independent*

\[
\sum_i c_i x_i = 0 \iff c_i = 0, \forall i
\]

  – and they *span* \( \mathcal{H} \): i.e., for any \( v \) in \( \mathcal{H} \), \( v \) can be written as

\[
v = \sum_i c_i x_i
\]

• These *two conditions* mean that *any* \( v \in \mathcal{H} \) can be *uniquely* represented in this form.
**Basis**

- Note that
  - By making the canonical representations for the vectors $x_i$ the columns of a matrix $X$, these two conditions can be compactly written as
  - Condition 1. The vectors $x_i$ are **linear independent**:
    \[ Xc = 0 \iff c = 0 \]
  - Condition 2. The vectors $x_i$ span $H$
    \[ \forall v \neq 0, \exists c \neq 0 \text{ such that } v = Xc \]

- Also, all bases of $H$ have the **same** number of vectors, which is called the **dimension** of $H$
  - *This is valid for any vector space!*
Basis

• example
  – A basis of the vector space of images of faces
  – The figure only show the first 16 basis vectors but there actually more
  – These vectors are orthonormal
**Orthogonality**

- Two vectors are **orthogonal** iff their inner product is zero
  - e.g. \[ \int_0^{2\pi} \sin(ax)\cos(ax)\,dx = \left. \frac{\sin^2 ax}{2a} \right|_0^{2\pi} = 0 \]

  in the space of functions defined on \([0,2\pi]\), \(\cos(ax)\) and \(\sin(ax)\) are orthogonal

- **Two subspaces** \(V\) and \(W\) are orthogonal, \(V \perp W\), if **every** vector in \(V\) is orthogonal to **every** vector in \(W\)

- a **set** of vectors \(x_1, \ldots, x_k\) is called
  - **orthogonal** if **all pairs** of vectors are orthogonal.
  - **orthonormal** if all of the orthogonal vectors also have unit norm.

\[
\langle x_i, x_j \rangle = \begin{cases} 
0, & \text{if } i \neq j \\
1, & \text{if } i = j 
\end{cases}
\]
Matrix

- an \textbf{m x n matrix represents} a linear operator that maps a vector from the \textit{domain} \( \mathcal{X} = \mathbb{R}^n \) to a vector in the \textit{codomain} \( \mathcal{Y} = \mathbb{R}^m \)

- E.g. the equation \( y = Ax \) sends \( x \) in \( \mathbb{R}^n \) to \( y \) in \( \mathbb{R}^m \) according to

\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_m
\end{bmatrix} =
\begin{bmatrix}
a_{11} & a_{1n} \\
\vdots & \ddots & \vdots \\
a_{m1} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
\vdots \\
x_n
\end{bmatrix}
\]

\( \mathcal{X} \quad \xrightarrow{A} \quad \mathcal{Y} \)
Matrix-Vector Multiplication I

• Consider \( y = Ax \), i.e. \( y_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m \)
• This is equivalent to

\[
\begin{pmatrix}
\vdots \\
y_i \\
\vdots
\end{pmatrix} =
\begin{pmatrix}
a_{i1} & \cdots & a_{in}
\end{pmatrix}
\begin{pmatrix}
x_1 \\
\vdots \\
x_n
\end{pmatrix}
= \sum_{j=1}^{n} a_{ij} x_j
= \begin{pmatrix}
\vdots \\
(\,-a_i\,-) x \\
\vdots
\end{pmatrix}
\quad (m \text{ rows})
\]

• where “(\,- a_i\,-)” means the \( i^{th} \) row of \( A \). Hence
  – the \( i^{th} \) component of \( y \) is the inner product of (\,- a_i\,-) and \( x \).
  – The \( m \) components of \( y \) are obtained by “projecting” \( x \) onto (i.e., taking the inner product with) the \( m \) rows of \( A \) in the domain space \( \mathcal{X} = \mathbb{R}^n \)
Matrix-Vector Multiplication II

- But there is more. Let \( y = Ax \), i.e. \( y_i = \sum_{j=1}^{n} a_{ij}x_j \), now be written as

\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_m
\end{bmatrix} = \sum_{j=1}^{n} a_{ij}x_j
= \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\
\vdots \\
a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}
= \begin{bmatrix} a_1 \end{bmatrix}x_1 + \cdots + \begin{bmatrix} a_n \end{bmatrix}x_n
\]

where \( a_i \) with “|” above and below means the \( i^{th} \) column of \( A \).

- The component \( x_i \) weights the \( i^{th} \) column of \( A \) in codomain \( \mathcal{Y} = \mathbb{R}^m \) (= column space = space spanned by the \( n \) columns of \( A \)).
- I.e, \( y \) is a linear combination of the \( n \) columns of \( A \) in the codomain \( \mathcal{Y} = \mathbb{R}^m \)
Matrix-Vector Multiplication I & II

• Thus there are **two alternative** (dual) pictures of $y = Ax$:
  – “Coordinates of $y$” = “$x$ ‘projected’ onto **row space of $A$’” (The $\mathcal{X} = \mathbb{R}^n$ viewpoint)

    ![Diagram of matrix-vector multiplication](image)

    - Domain $\mathcal{X} = \mathbb{R}^n$
    - Codomain $\mathcal{Y} = \mathbb{R}^m$

  – “Components of $x$” = “‘coordinates’ of $y$ along columns of $A$” (\(\mathcal{Y} = \mathbb{R}^m\) viewpoint)

- $\mathbf{y} = \begin{bmatrix} 1 \mid a_1 \mid x_1 + \cdots + \mid a_n \mid x_n \end{bmatrix}$
- $\mathbf{y} = \begin{bmatrix} \vdots \end{bmatrix}$
- $\mathbf{y} = \begin{bmatrix} a_i \end{bmatrix} \mathbf{x} \begin{bmatrix} -a_1 - \cdots - a_m \end{bmatrix}$ (m rows)
Block Matrix Multiplication

- the matrix multiplication formula

\[ C = AB \iff c_{ij} = \sum_k a_{ik} b_{kj} \]

also applies to “block matrices” when these are defined to be **conformal**.

- for example, if \(A,B,C,D,E,F,G,H\) are conformal matrices,

\[
\begin{bmatrix}
A & B \\
C & D
\end{bmatrix}
\begin{bmatrix}
E & F \\
G & H
\end{bmatrix}
= \begin{bmatrix}
AE+BG & AF+BH \\
CE+DG & CF+DH
\end{bmatrix}
\]

- To be **conformal** means that the sizes of the matrices \(A,B,C,D,E,F,G,H\) have to be such that the intermediate operations make sense!
Matrix-Vector Multiplication I & II

- This makes it easy to derive the two alternative pictures

- The *row space picture (or viewpoint):*

\[
\begin{bmatrix}
\vdots \\
y_i \\
\vdots 
\end{bmatrix} = \begin{bmatrix}
a_{in} & \cdots & a_{in} \\
\vdots & \ddots & \vdots \\
\vdots & & \vdots 
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_n 
\end{bmatrix} = \begin{bmatrix}
(\mathbf{- a_i -}) \mathbf{1}_{\times n} \\
\vdots \\
\vdots 
\end{bmatrix} \begin{bmatrix}
x_{nxl} 
\end{bmatrix} = \begin{bmatrix}
(\mathbf{- a_i -}) \mathbf{x} 
\end{bmatrix}
\]

Scalar multiplication between the *row blocks* \((-a_i\mathbf{-})\) and \(\mathbf{x}\)

- The *column space picture (or viewpoint):*

\[
\begin{bmatrix}
\vdots \\
y_i \\
\vdots 
\end{bmatrix} = \begin{bmatrix}
a_{in} & \cdots & a_{in} \\
\vdots & \ddots & \vdots \\
\vdots & & \vdots 
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_n 
\end{bmatrix} = \begin{bmatrix}
\mathbf{a_1} & \cdots & \mathbf{a_n} \\
\vdots & \ddots & \vdots \\
\vdots & \cdots & \vdots 
\end{bmatrix} \begin{bmatrix}
(x_1)_{1\times l} \\
(x_n)_{1\times l} 
\end{bmatrix} = \sum_i \begin{bmatrix}
\mathbf{a_i} 
\end{bmatrix} \mathbf{x_i}
\]

Inner products between blocks given by the (scalar) blocks \(x_i\) and the *column blocks* of \(A\).
Square $n \times n$ matrices

- in this case $m = n$ and the row and column subspaces are both equal to (copies of) $R^n$
Orthogonal matrices

• A matrix is called **orthogonal** if it is square and has **orthonormal** columns.

• Important properties:
  
  1) The **inverse** of an orthogonal matrix is its **transpose**
     this can be easily shown with the block matrix trick. (Also see later.)

     \[
     A^T A = \begin{pmatrix}
     -a_i^T & \\
     \vdots & \\
     -a_n^T & \\
     \end{pmatrix}
     \begin{pmatrix}
     \vdots & \\
     a_j & \\
     \vdots & \\
     \end{pmatrix}
     = \begin{pmatrix}
     1 & 0 & \cdots & 0 \\
     0 & 1 & \cdots & 0 \\
     \vdots & \vdots & \ddots & \vdots \\
     0 & 0 & \cdots & 1 \\
     \end{pmatrix}
     \]

  2) A **proper** (\(\det(A) = 1\)) orthogonal matrix is a rotation matrix
     this follows from the fact that it is **unitary**, i.e., does not change the norms (“sizes”) of the vectors on which it operates,

     \[
     \|Ax\|^2 = (Ax)^T (Ax) = x^T A^T A x = x^T x = \|x\|^2,
     \]

     **AND** does **NOT** induce a **reflection**.
Rotation matrices

• The combination of
  1. “operator” interpretation
  2. “block matrix trick”

  is useful in many situations

• Example:
  – “What is the matrix $R$ that rotates the plane $\mathbb{R}^2$ by $\theta$ degrees?”
Rotation matrices

- The **key** is to consider how the matrix operates on the vectors $\mathbf{e}_i$ of the **canonical basis**
  
  - note that $\mathbf{R}$ sends $\mathbf{e}_1$ to $\mathbf{e}'_1$

$$
\begin{bmatrix}
    r_{11} & r_{12} \\
    r_{21} & r_{22}
  \end{bmatrix}
\begin{bmatrix}
    1 \\
    0
  \end{bmatrix}
$$

- using the **column space picture**

$$
\begin{bmatrix}
    r_{11} \\
    r_{21}
  \end{bmatrix}
\times 1 + \begin{bmatrix}
    r_{12} \\
    r_{22}
  \end{bmatrix}
\times 0 = \begin{bmatrix}
    r_{11} \\
    r_{21}
  \end{bmatrix}
$$

- from which we have the first column of the matrix

$$
\mathbf{R} = 
\begin{bmatrix}
    e'_1 & r_{12} \\
    r_{21} & r_{22}
  \end{bmatrix}
= 
\begin{bmatrix}
    \cos \theta & r_{12} \\
    \sin \theta & r_{22}
  \end{bmatrix}
$$
Rotation Matrices

- and we do the same for $e_2$
  - $R$ sends $e_2$ to $e'_2$
    
    \[
    e'_2 = \begin{bmatrix} \bar{r}_{11} & \bar{r}_{12} \\ \bar{r}_{21} & \bar{r}_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \bar{r}_{11} \\ \bar{r}_{21} \end{bmatrix} \times 0 + \begin{bmatrix} \bar{r}_{12} \\ \bar{r}_{22} \end{bmatrix} \times 1 = \begin{bmatrix} \bar{r}_{12} \\ \bar{r}_{22} \end{bmatrix}
    \]
  - from which
    
    \[
    R = \begin{bmatrix} e'_1 & e'_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}
    \]
  - check
    
    \[
    R^T R = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = I
    \]
### Projections

- **What if A is not orthogonal?**
  - Consider $y = A^T x$ and $x' = Ay$ (Note that $x' \in$ column space)
    - $x' = x$ for all $x$ if and only if $AA^T = I$!
    - this means that A has to be **orthogonal** to have $x' = x$
- **What happens when this is not the case?** Then take ECE 174!!
  - E.g., if $(AA^T)^2 = AA^T$, then $AA^T$ is **idempotent** (and also **obviously symmetric**) so we get an **orthogonal projection** of $x$ onto the column space of A

\[ A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ then } y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \]

\[ x' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \]
Null Space of a Matrix

• What happens to the part that is lost?
• For the previous example this part belongs to the “null space” of $A^T$

$$N(A^T) = \left\{ x \mid A^T x = 0 \right\}$$

– In the example, this is comprised of all vectors of the type $\begin{bmatrix} 0 \\ 0 \\ \alpha \end{bmatrix}$ since

$$A^T x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

**FACT:** $N(A)$ is always orthogonal to the row space of $A$:

– $x$ is in the null space iff it is orthogonal to all rows of $A$
– For the previous example this means that $N(A^T)$ is orthogonal to the column space of $A$
Orthogonal Matrices – Cont.

• An orthogonal matrix has linearly independent columns and therefore must have an inverse.

• Note that \( A^T A = I \) (proven earlier) and the existence of an inverse \( AA^{-1} = I \) implies

\[
A^{-1} = I A^{-1} = A^T AA^{-1} = A^T I = A^T .
\]

Thus

\[
A^T A = AA^T = I
\]

• This means that
  – \( A \) has **orthonormal columns and rows**
  – **Each** of these two sets of vectors span **all** of \( \mathbb{R}^n \)
  – There is **no** extra room for an orthogonal subspace in the rowspace
  – The null space of \( A^T \) has to be empty
  – The square matrix \( A \) has **full rank**
The **Four Fundamental Subspaces**

- These exist for any matrix:
  - **Column Space**: space spanned by the columns
  - **Row Space**: space spanned by the rows
  - **Nullspace**: space of vectors orthogonal to all rows (also known as the orthogonal complement of the row space)
  - **Left Nullspace**: space of vectors orthogonal to all columns (also known as the orthogonal complement of the column space)

Assume Domain of \( A = \) Codomain of \( A \). Then:

- **Special Case I**: Square Symmetric Matrices \((A = A^T)\):
  - Column Space is equal to the Row Space
  - Nullspace is equal to the Left Nullspace, and is therefore orthogonal to the Column Space

- **Special Case II**: \( n \times n \) Orthogonal Matrices \((A^TA = AA^T = I)\)
  - Column Space = Row Space = \( \mathbb{R}^n \)
  - Nullspace = Left Nullspace = \( \{0\} \) = the **Trivial Subspace**
END