Princeton University MAT 202 Spring 2008

Willie W. Wong

Mar 3 - 7

Mar 3, 2008

- Let's begin by recalling our notation from last week: given a basis $\mathcal{B} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_P\}$ for a subspace W of $\mathbb{R}^M$, we have the following interchangeable notions

$$\boldsymbol{v} = b_1\boldsymbol{w}_1 + b_2\boldsymbol{w}_2 + \cdots + b_P\boldsymbol{w}_P \iff [\boldsymbol{v}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_P \end{bmatrix}$$

- Now, suppose we have another basis $\mathcal{C} = \{\boldsymbol{u}_1, \ldots \boldsymbol{u}_Q\}$ of U in $\mathbb{R}^M$, such that each one of the $\boldsymbol{w}$s is a linear combination of the $\boldsymbol{u}$s (notice that $Q \geq P$ since $\boldsymbol{w}$s are linearly independent in U). Then we can also decompose the vector $\boldsymbol{v}$ in the $\mathcal{C}$ basis

$$\boldsymbol{v} = c_1\boldsymbol{u}_1 + \cdots + c_Q\boldsymbol{u}_Q \iff [\boldsymbol{v}]_{\mathcal{C}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_Q \end{bmatrix}$$

- Now we ask, what is the relationship between the numbers $b_1 \ldots b_P$ and $c_1 \ldots c_Q$? Let us decompose the $\mathcal{B}$-basis vectors in terms of those of $\mathcal{C}$: we write

$$\boldsymbol{w}_j = s_{1j}\boldsymbol{u}_1 + s_{2j}\boldsymbol{u}_2 + \cdots s_{Qj}\boldsymbol{u}_Q$$

  In our new notation, this is

$$[\boldsymbol{w}_j]_{\mathcal{C}} = \begin{bmatrix} s_{1j} \\ s_{2j} \\ \vdots \\ s_{Qj} \end{bmatrix}$$

And we plug this into the expression above for $\boldsymbol{v}$:

$$\begin{aligned}\boldsymbol{v} &= b_1\boldsymbol{w}_1 + b_2\boldsymbol{w}_2 + \cdots + b_P\boldsymbol{w}_p \\ &= b_1(s_{11}\boldsymbol{u}_1 + s_{21}\boldsymbol{u}_2 + \cdots) + b_2(s_{12}\boldsymbol{u}_1 + s_{22}\boldsymbol{u}_2 + \cdots) + b_3(s_{13}\boldsymbol{u}_1 + s_{23}\boldsymbol{u}_2 + \cdots) + \cdots \\ &= (s_{11}b_1 + s_{12}b_2 + \cdots)\boldsymbol{u}_1 + (s_{21}b_1 + s_{22}b_2 + \cdots)\boldsymbol{u}_2 + (s_{31}b_1 + s_{32}b_2 + \cdots)\boldsymbol{u}_3 + \cdots \\ &= c_1\boldsymbol{u}_1 + \cdots c_Q\boldsymbol{u}_Q\end{aligned}$$

And so we have that

$$\begin{aligned}c_1 &= s_{11}b_1 + s_{12}b_2 + s_{13}b_3 + \cdots \\ c_2 &= s_{21}b_1 + s_{22}b_2 + s_{23}b_3 + \cdots \\ &\vdots \\ c_Q &= s_{Q1}b_1 + s_{Q2}b_s + s_{Q3}b_3 + \cdots\end{aligned}$$

This we see means that *the column vector of the $\mathcal{C}$-coördinates of $\boldsymbol{v}$ is equal to the $\mathcal{B}$-coördinates of $\boldsymbol{v}$ multiplied by some matrix $\mathcal{S}$.* In particular, the column vectors of $\mathcal{S}$ are precisely the column vectors corresponding to the $\mathcal{C}$-coördinates of the $\mathcal{B}$-basis vectors. In mathematical notation, we write

$$[\boldsymbol{v}]_{\mathcal{C}} = \mathcal{S}[\boldsymbol{v}]_{\mathcal{B}}$$

where

$$\mathcal{S} = \begin{bmatrix}[\boldsymbol{w}_1]_{\mathcal{C}} & [\boldsymbol{w}_2]_{\mathcal{C}} & \ldots & [\boldsymbol{w}_P]_{\mathcal{C}}\end{bmatrix}$$

- A very important special case of the above general rule is when $\mathcal{C}$ is the standard basis for $\mathbb{R}^M$. Then $\mathcal{S}$ is an $M \times P$ matrix formed by the Cartesian coördinates of the basis vectors of $\mathcal{B}$. Even more special is the case when $M = P$: then $\mathcal{S}$ is the invertible $P \times P$ matrix whose column vectors are just the basis vectors of $\mathcal{B}$ written in standard coördinates.

- Example: given two bases $\mathcal{B} = \{\boldsymbol{w}_1, \boldsymbol{w}_2\}$ and $\mathcal{C} = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ of $\mathbb{R}^2$. Suppose that we know

$$\boldsymbol{u}_1 = 2\boldsymbol{w}_1 - \boldsymbol{w}_2\ , \quad \boldsymbol{u}_2 = \boldsymbol{w}_1 + 3\boldsymbol{w}_2$$

Given a vector $\boldsymbol{v}$ such that

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{bmatrix}4\\5\end{bmatrix}$$

find $[\boldsymbol{v}]_{\mathcal{C}}$.

  - First we need to find the $\mathcal{S}$ matrix such that the columns of $\mathcal{S}$ is the $\mathcal{C}$-coördinates of the vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. So we need to solve the give linear system for $\boldsymbol{w}$s. We can do this in 2 different ways.
  - First, we can directly solve it:

$$\left\{\begin{array}{rl} 2\boldsymbol{w}_1 - \boldsymbol{w}_2 & = \boldsymbol{u}_1 \\ \boldsymbol{w}_1 + 3\boldsymbol{2}_2 & = \boldsymbol{u}_2\end{array}\right. \Rightarrow \left\{\begin{array}{rl} 2\boldsymbol{w}_1 - \boldsymbol{w}_2 & = \boldsymbol{u}_1 \\ 7\boldsymbol{w}_1 + 0\boldsymbol{w} & = 3\boldsymbol{u}_1 + \boldsymbol{u}_2\end{array}\right. \Rightarrow \left\{\begin{array}{rl} -7\boldsymbol{w}_2 & = 1\boldsymbol{u}_1 - 2\boldsymbol{u}_2 \\ 7\boldsymbol{w}_1 & = 3\boldsymbol{u}_1 + \boldsymbol{u}_2\end{array}\right.$$

So

$$\boldsymbol{w}_1 = \frac{3}{7}\boldsymbol{u}_1 + \frac{1}{7}\boldsymbol{u}_2\,, \quad \boldsymbol{w}_2 = -\frac{1}{7}\boldsymbol{u}_1 + \frac{2}{7}\boldsymbol{u}_2$$

and

$$\mathcal{S} = \begin{bmatrix}[\boldsymbol{w}_1]_\mathcal{C} & [\boldsymbol{w}_2]_\mathcal{C}\end{bmatrix} = \begin{bmatrix}3/7 & -1/7 \\ 1/7 & 2/7\end{bmatrix}$$

- The second method is to start with our given relation between the basis vectors and write them all in $\mathcal{C}$-coördinates: notice that $[\boldsymbol{u}_1]_\mathcal{C} = \begin{bmatrix}1 \\ 0\end{bmatrix}$ and $[\boldsymbol{u}_2]_\mathcal{C} = \begin{bmatrix}0 \\ 1\end{bmatrix}$. Then we have

$$\begin{bmatrix}1 \\ 0\end{bmatrix} = 2[\boldsymbol{w}_1]_\mathcal{C} - [\boldsymbol{w}_2]_\mathcal{C} = \begin{bmatrix}[\boldsymbol{w}_1]_\mathcal{C} & [\boldsymbol{w}_2]_\mathcal{C}\end{bmatrix}\begin{bmatrix}2 \\ -1\end{bmatrix} = \mathcal{S}\begin{bmatrix}2 \\ -1\end{bmatrix}$$

and similarly, we have

$$\begin{bmatrix}0 \\ 1\end{bmatrix} = \mathcal{S}\begin{bmatrix}1 \\ 3\end{bmatrix}$$

Putting the two together, we have

$$I_2 = \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix} = \mathcal{S}\begin{bmatrix}2 & 1 \\ -1 & 3\end{bmatrix}$$

Notice that the matrix on the far right, let's call it $\mathcal{T}$, is nothing but the matrix of $\mathcal{B}$-coördinates of $\mathcal{C}$-basis vectors. From the relation we have

$$\mathcal{S} = \mathcal{T}^{-1}$$

which we can calculate as

$$\mathcal{S} = \begin{bmatrix}3/7 & -1/7 \\ 1/7 & 2/7\end{bmatrix}$$

- This conclusion is hardly surprising, since we have that, by definition

$$[\boldsymbol{v}]_\mathcal{B} = \mathcal{T}[\boldsymbol{v}]_\mathcal{C}$$

it follows that $\mathcal{S} = \mathcal{T}^{-1}$.

- To finish the problem, we just need to multiply

$$[\boldsymbol{v}]_\mathcal{C} = \mathcal{S}[\boldsymbol{v}]_\mathcal{B} = \begin{bmatrix}3/7 & -1/7 \\ 1/7 & 2/7\end{bmatrix}\begin{bmatrix}4 \\ 5\end{bmatrix} = \begin{bmatrix}(3\times 4 - 1\times 5)/7 \\ (1\times 4 + 2\times 5)/7\end{bmatrix} = \begin{bmatrix}1 \\ 2\end{bmatrix}$$

- Now, we can not only just write vectors in coördinates: we can also write linear transformations as matrices in a coördinate system. Now, I am going to ask you to dig back in memory to something we did in the second week of class, about how we construct the matrix corresponding to a linear transformation: first we decompose the vector $\boldsymbol{x}$ in standard basis

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + \cdots x_M\boldsymbol{e}_M$$

and notice that using linearity of the transformation:

$$T(\boldsymbol{x}) = x_1 T(\boldsymbol{e}_1) + \dots + x_M T(\boldsymbol{e}_M)$$

we concluded that the matrix of the linear transformation is

$$A = \begin{bmatrix} T(\boldsymbol{e}_1) & \dots & T(\boldsymbol{e}_M) \end{bmatrix}$$

- With our machinery of arbitrary basis vectors now, we can easily generalize the above construction to a basis $\mathcal{B}$. Let $T : \mathbb{R}^M \to \mathbb{R}^M$ be a linear transformation and let $\mathcal{B} = \{\boldsymbol{w}_1 \dots \boldsymbol{w}_M\}$ be a basis for $\mathbb{R}^M$. Very naturally, we can define the $\mathcal{B}$-matrix of T, which we will, by abuse of notation, write $[T]_\mathcal{B}$ as

  $$[T]_\mathcal{B} = \begin{bmatrix} [T(\boldsymbol{w}_1)]_\mathcal{B} & [T(\boldsymbol{w}_2)]_\mathcal{B} & \dots & [T(\boldsymbol{w}_M)]_\mathcal{B} \end{bmatrix}$$

  This matrix satisfies
  $$[T(\boldsymbol{v})]_\mathcal{B} = [T]_\mathcal{B}[\boldsymbol{v}]_\mathcal{B}$$

- In fact, this machinery can be made even more general: let $T : \mathbb{R}^M \to \mathbb{R}^N$ be a linear transformation. Let $\mathcal{B} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_M\}$ be a basis of $\mathbb{R}^M$. Let $\mathcal{C} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_N\}$ be a basis of $\mathbb{R}^N$, then we can define the $\mathcal{B} \to \mathcal{C}$-matrix of T, which we write

  $$[T]_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix} [T(\boldsymbol{w}_1)]_\mathcal{C} & [T(\boldsymbol{w}_2)]_\mathcal{C} & \dots & [T(\boldsymbol{w}_M)]_\mathcal{C} \end{bmatrix}$$

  The matrix satisfies
  $$[T(\boldsymbol{v})]_\mathcal{C} = [T]_{\mathcal{B}\to\mathcal{C}}[\boldsymbol{v}]_\mathcal{B}$$
  this definition is just for your amusement... it probably will not appear on your exams.

- Example: consider the operation “rotate by 45 degrees counter clockwise and scale up by a factor of $\sqrt{2}$”. In standard coördinates, its matrix is

  $$A = [T]_{\text{standard basis}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

  Let $\mathcal{B}$ be the basis given by, in standard coördinates

  $$\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \boldsymbol{w}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

  what is the $\mathcal{B}$-matrix of T?

  - First we need to find $T(\boldsymbol{w}_1)$ and $T(\boldsymbol{w}_2)$. This we can do in standard coördinates by applying the matrix A. We get

    $$T(\boldsymbol{w}_1) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \quad T(\boldsymbol{w}_2) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

- Now we need to find the matrix $\mathcal{S}$ such that

$$[\boldsymbol{v}]_{\mathcal{B}} = \mathcal{S}[\boldsymbol{v}]_{\text{standard}}$$

  but we are already given the matrix $\mathcal{T}$ which give

$$[\boldsymbol{v}]_{\text{standard}} = \mathcal{T}[\boldsymbol{v}]_{\mathcal{B}}$$

  where

$$\mathcal{T} = \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix}$$

  (we read this off by the defining relation of the standard coördinate representation of the vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$.) Using the previous example, we have $\mathcal{S} = \mathcal{T}^{-1}$. So

$$\mathcal{S} = \frac{1}{7}\begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix}$$

  So

$$[T(\boldsymbol{w}_1)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad , \quad T(\boldsymbol{w}_2) = \begin{bmatrix} -5/7 \\ 8/7 \end{bmatrix}$$

- Therefore

$$[T]_{\mathcal{B}} = \begin{bmatrix} 0 & -5/7 \\ 2 & 8/7 \end{bmatrix}$$

- Now, suppose we have two bases $\mathcal{B}$ and $\mathcal{C}$, and the transition matrix $\mathcal{S}$ which gives

$$[\boldsymbol{v}]_{\mathcal{C}} = \mathcal{S}[\boldsymbol{v}]_{\mathcal{B}}$$

  Now, given a linear transformation T, what can we say about the matrices $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{C}}$? We use the definition

$$[T(\boldsymbol{v})]_{\mathcal{B}} = [T]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}$$

  So

$$[T(\boldsymbol{v})]_{\mathcal{C}} = \mathcal{S}[T(\boldsymbol{v})]_{\mathcal{B}} = \mathcal{S}[T]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}$$

  But we can also substitute in

$$[\boldsymbol{v}]_{\mathcal{B}} = \mathcal{S}^{-1}[\boldsymbol{v}]_{\mathcal{C}}$$

  So

$$[T(\boldsymbol{v})]_{\mathcal{C}} = \mathcal{S}[T]_{\mathcal{B}}\mathcal{S}^{-1}[\boldsymbol{v}]_{\mathcal{C}}$$

  which means that

$$[T]_{\mathcal{C}} = \mathcal{S}[T]_{\mathcal{B}}\mathcal{S}^{-1}$$

- Using this, the previous example becomes much easier: you can check that

$$\begin{bmatrix} 0 & -5/7 \\ 2 & 8/7 \end{bmatrix} = \begin{bmatrix} 1/7 & -2/7 \\ 3/7 & 1/7 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix}$$

- A terminology: when we have two square matrics A and B and an invertible matrix $\mathcal{S}$ such that

  $$A = \mathcal{S}B\mathcal{S}^{-1}$$

  or

  $$A\mathcal{S} = \mathcal{S}B$$

  we say that the two matrices are *similar*.

- Similarity is an equivalence relation:
  - It is *reflexive*: any matrix A is similar to itself (by the identity matrix)
  - It is *symmetric*: if A is similar to B, B is similar to A (by setting $\mathcal{T} = \mathcal{S}^{-1}$, we have that $A = \mathcal{S}B\mathcal{S}^{-1}$ implies $B = \mathcal{T}A\mathcal{T}^{-1}$)
  - It is *transitive*: if A is similar to B, and B is similar to C, then A is similar to C.

- One cool thing about similar matrices: suppose A and B are similar, that is $A = \mathcal{S}B\mathcal{S}^{-1}$. Then a power of A, i.e. $A^k = \underbrace{A \cdot A \cdot A \cdots A}_{k \text{ times}}$ can be written as

  $$A^k = \mathcal{S}B^k\mathcal{S}^{-1}$$

MAR 5, 2008

- Time to move-on to something completely different. Today we talk about
  - Orthogonality
  - Orthonormal bases
  - Some calculus (in)equalities

- The basic idea for orthogonality we have already introduced in the second week: that the dot product, in some sense, measures how aligned/parallel two vectors are. Remember we have already defined the length of a vector

  $$|\boldsymbol{v}| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$$

  and that a vector with length 1 is called a unit vector. And remember that we say that two non-zero vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ are perpendicular (or orthogonal) if the dot-product $\boldsymbol{v} \cdot \boldsymbol{w} = 0$.

- An important fact: dot-products are not preserved under change of coördinates (in general). When in doubt, when taking dot products, first calculate the coördinates of the vector relative to the standard basis, and then take the dot product thereof.

- Remember from last Friday we we started talking about linear coördinates: one of the advantages for the standard or Cartesian coördinates that I mentioned is the fact that all of the basis vectors are mutually perpendicular[1]. We want to find "nice" coördinates like this (which, incidentally, also allows the taking of dot products in its coördinates).

- A definition: the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P$ in $\mathbb{R}^M$ are called *orthonormal* if they are all unit vectors and they are all mutually perpendicular, i.e.

$$\boldsymbol{u}_i \cdot \boldsymbol{u}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

- Example: the standard basis of $\mathbb{R}^M$ is orthonormal.

- Fix an angle θ, the vectors $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \begin{bmatrix} \sin\theta \\ -\cos\theta \end{bmatrix}$ are orthonormal.

- Notice, by construction: orthonormal vectors are linearly independent. (Suppose they are not, then there exists a linear relation among them

$$c_1 \boldsymbol{u}_1 + \cdots + c_P \boldsymbol{u}_P = 0$$

  we can take its dot product against $\boldsymbol{u}_i$. On the one hand, 0 dotted against anything is 0. On the other

$$(c_1 \boldsymbol{u}_1 + \cdots + c_P \boldsymbol{u}_P) \cdot \boldsymbol{u}_i = c_i$$

  therefore $c_i = 0$ for all i.) And so, if you have a set of M orthonormal vectors in $\mathbb{R}^M$, they form a basis.

- With the help of orthonormal vectors, we can generalize the concept of a projection. Remember what we did to construct a projection onto a line L: we take the unit vector $\boldsymbol{u}$ that spans L, and define

$$proj_{\,L}(\boldsymbol{x}) = (\boldsymbol{x} \cdot \boldsymbol{u})\boldsymbol{u}$$

- What are the important properties of the projection (onto a line)? The most important property is that we can decompose a vector $\boldsymbol{x}$ into two parts

$$\boldsymbol{x} = proj_{\,L}(\boldsymbol{x}) + (\boldsymbol{x} - proj_{\,L}(\boldsymbol{x}))$$

  the first part is parallel to L and the second part perpendicular to L. We can generalize this to higher dimensions: Let V be a subspace of $\mathbb{R}^M$, with $\dim(V) = P$. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P$ be an orthonormal basis of V (meaning that the vectors form a basis of V, and, on top of that, they are orthonormal). Then we define the projection

$$proj_{\,V}(\boldsymbol{x}) = (\boldsymbol{x}) \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{x}) \cdot \boldsymbol{u}_2)\boldsymbol{u}_2 + \cdots + (\boldsymbol{x}) \cdot \boldsymbol{u}_P)\boldsymbol{u}_P$$

[1] In a sense, this is a tautology, since the dot-product is only defined after a coördinate system is chosen, and by choosing to work with the standard basis, the vectors are by definition mutually perpendicular and by definition unit.

and check that

$$\boldsymbol{x} = proj_{\,V}(\boldsymbol{x}) + (\boldsymbol{x} - proj_{\,V}(\boldsymbol{x}))$$

with the first part parallel to V and the second part perpendicular to V. The parallel property is easy: a vector is parallel to a subspace if it lies in the subspace. Since V is spanned by $\boldsymbol{u}_1 \dots \boldsymbol{u}_P$, and $proj_V(\boldsymbol{x})$ is a linear combination of them, the projected $proj_V(\boldsymbol{x}) \in V$. To check the orthogonality of the second part, we take the dot product of it against one of the basis vectors $\boldsymbol{u}_i$

$$\boldsymbol{u}_i \cdot (\boldsymbol{x} - proj_{\,V}(\boldsymbol{x})) = \boldsymbol{u}_i \cdot \boldsymbol{x} - \boldsymbol{u}_i \cdot ((\boldsymbol{x}) \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{x}) \cdot \boldsymbol{u}_2)\boldsymbol{u}_2 + \cdots + (\boldsymbol{x}) \cdot \boldsymbol{u}_P)\boldsymbol{u}_P)$$

By orthonormality of the $\boldsymbol{u}$s, we have that only the $\boldsymbol{u}_i$ term from inside the right-most parenthesis survive the dot product, and becomes

$$\boldsymbol{u}_i \cdot (\boldsymbol{x} - proj_{\,V}(\boldsymbol{x})) = \boldsymbol{u}_i \cdot \boldsymbol{x} - (\boldsymbol{u}_i \cdot \boldsymbol{x})(\boldsymbol{u}_i \cdot \boldsymbol{u}_i) = 0$$

Since this is true for arbitrary $\boldsymbol{u}_i$, and the $\boldsymbol{u}$s are a basis for V, we have that $(\boldsymbol{x} - proj_{\,V}(\boldsymbol{x}))$ is perpendicular to V.

- So we have: given a subspace V in $\mathbb{R}^M$, we can always write an arbitrary vector $\boldsymbol{x} \in \mathbb{R}^M$ as

  $$\boldsymbol{x} = \boldsymbol{x}^{\parallel} + \boldsymbol{x}^{\perp}$$

  where $\boldsymbol{x}^{\parallel} = proj_V(\boldsymbol{x})$ is in V, and $\boldsymbol{x}^{\perp} = \boldsymbol{x} - proj_V(\boldsymbol{x})$ is perpendicular to V.

- The map from $\boldsymbol{x}$ to $proj_V(\boldsymbol{x})$ is a linear transformation, and hence the decomposition into V-parallel and V-orthogonal parts is a unique decomposition.

- Example: let V be the image of

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

  Find the projection of $\begin{bmatrix} 1 \\ 5 \\ 3 \\ 2 \end{bmatrix}$ onto V.

  – First V is spanned by the two column vectors of the matrix. Taking their dot product, we see that they are orthogonal. Therefore, to get an orthonormal basis of V, all we need to do is to *normalize* the vectors.

  – For both of the vectors, we can calculate their lengths to be 2. So the unit vectors corresponding to them are

  $$\boldsymbol{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \boldsymbol{u}_2 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

– To calculate the projection, we use

$$\begin{aligned} proj_V(\boldsymbol{x}) &= (\boldsymbol{u}_1 \cdot \boldsymbol{x})\boldsymbol{u}_1 + (\boldsymbol{u}_2 \cdot \boldsymbol{x})\boldsymbol{u}_2 \\ &= (\frac{1}{2} + \frac{5}{2} - \frac{3}{2} + 1)\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + (\frac{1}{2} - \frac{5}{2} + \frac{3}{2} + 1)\begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ 1 \\ -1 \\ \frac{3}{2} \end{bmatrix} \end{aligned}$$

- Now we take the subspace V to be the entirety of $\mathbb{R}^M$ and suppose we have some orthonormal basis $\boldsymbol{u}_1 \ldots \boldsymbol{u}_M$. Since the projection onto V now becomes just the identity transformation, we have:

$$\boldsymbol{x} = (\boldsymbol{x} \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{x} \cdot \boldsymbol{u}_M)\boldsymbol{u}_M$$

  and we have that, for calculating the coördinates of a vector $\boldsymbol{x}$ in an orthonormal basis, one doesn't need to solve linear systems (i.e., invert the transformation matrix $\mathcal{S}$); one need only to take dot products.

- Notation: given a subspace V in $\mathbb{R}^M$, we write $V^\perp$ for what we call the *orthogonal complement* of V. It is defined as the set of all vectors such that

$$V^\perp := \{\boldsymbol{x} \in \mathbb{R}^M : \boldsymbol{v} \cdot \boldsymbol{x} = 0 \text{ whenever } \boldsymbol{v} \in V\}$$

- It is also equal to the kernel of the transformation $proj_V$.

- Some properties

  – $V^\perp$ is a subspace
  – The only vector that is both in $V^\perp$ and V is the 0 vector.
  – $\dim(V) + \dim(V^\perp) = M$
  – $(V^\perp)^\perp = V$

- Some calculus inequalities:

  – Let's look at two vectors $\boldsymbol{x}, \boldsymbol{y}$.

$$|\boldsymbol{x} + \boldsymbol{y}|^2 = (\boldsymbol{x} + \boldsymbol{y}) \cdot (\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{x} + \boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{y} \cdot \boldsymbol{x} + \boldsymbol{y} \cdot \boldsymbol{y} = |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2 + 2\boldsymbol{x} \cdot \boldsymbol{y}$$

  – So, if (and only if) $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal, we have the *Pythagorean theorem*

$$|\boldsymbol{x}|^2 + |\boldsymbol{y}|^2 = |\boldsymbol{x} + \boldsymbol{y}|^2$$

  – Next, we use the Pythagorean theorem on the projection to subspace V:

$$|\boldsymbol{x}|^2 = |\boldsymbol{x}^\parallel|^2 + |\boldsymbol{x}^\perp|^2 = |\,proj_V(\boldsymbol{x})|^2 + |\boldsymbol{x}^\perp|^2$$

  So

$$|\,proj_V(\boldsymbol{x})| \leq |\boldsymbol{x}|$$

- Next, write $\boldsymbol{y} = |\boldsymbol{y}|\boldsymbol{u}$ where $\boldsymbol{u}$ is the unit vector in the direction of $\boldsymbol{y}$. Then

$$\boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{y}| \times \boldsymbol{x} \cdot \boldsymbol{u}$$

  But

$$|\boldsymbol{x} \cdot \boldsymbol{u}| = |\, proj_{\,\boldsymbol{u}}(\boldsymbol{x})| \leq |\boldsymbol{x}|$$

  So we have the Cauchy Schwarz inequality

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leq |\boldsymbol{x}||\boldsymbol{y}|$$

- Lastly, we want to define the angle between two vectors: we write

$$\cos\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|}$$

  where θ is the angle between the vectors, defined to be between 0 and π. It is well-defined since by Cauchy-Schwarz, the numerator is smaller than the denominator.

- A important application of the Pythagorean theorem: let $\boldsymbol{u}_1 \dots \boldsymbol{u}_P$ be an orthonormal basis of V, a subspace of $\mathbb{R}^M$, and let

$$\boldsymbol{x} = c_1\boldsymbol{u}_1 + \cdots + c_P\boldsymbol{u}_P$$

  we have that

$$|\boldsymbol{x}| = \sqrt{c_1^2 + c_2^2 + \cdot + c_P^2}$$

- A nice application of dot-products and finding angles between vectors is in finding whether there is a correlation between different characteristics of a population. Due to the lack of time, I will not talk about it in class, please read your books on pps 196-198 for details. While this will probably not be on the exam, if you do have any questions, don't hesitate to come to office hours.

MAR 7, 2008

- Today's topic: Gram-Schmidt process and QR factorization.

- The big question: while we learned about how to use orthonormal bases of a subspace on Wednesday, we don't know yet how to come up with an orthonormal basis.

- The Gram-Schmidt process is a process to find an orthonormal basis starting with an arbitrary set of basis vectors.

- The basic idea is that of the orthogonal projection onto subspaces. We start with $\boldsymbol{v}_1, \dots, \boldsymbol{v}_P$, basis vectors for some subspace V. We build up an orthonormal basis like this:

- First we take $\boldsymbol{v}_1$, we find the corresponding unit vector $\boldsymbol{u}_1$. Then $\boldsymbol{u}_1$ is an orthonormal basis for the subspace V_1 spanned by $\boldsymbol{v}_1$.
- Next, we look at the subspace spanned by $\boldsymbol{u}_1$ and $\boldsymbol{v}_2$. Since $\boldsymbol{v}_1$ is a linear combination of $\boldsymbol{u}_1$, this is the same subspace V_2 spanned by $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. We use the orthogonal projection to decompose $\boldsymbol{v}_2$ into two parts

$$\boldsymbol{v}_2 = \boldsymbol{v}_2^{\|} + \boldsymbol{v}_2^{\perp}$$

  where $\boldsymbol{v}_2^{\|}$ is in V_1 and $\boldsymbol{v}_2^{\perp}$ is orthogonal to V_1. We define $\boldsymbol{u}_2$ as the unit vector for $\boldsymbol{v}_2^{\perp}$. We see that since

$$\boldsymbol{v}_2 = (\boldsymbol{v}_2 \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + |\boldsymbol{v}_2^{\perp}|\boldsymbol{u}_2$$

  that $\boldsymbol{v}_2$ is a linear combination of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, and so V_2 is spanned by $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Notice that also, by construction, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are orthonormal.
- We continue with the subspace V_3 spanned by $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$, which is the same as the subspace spanned by $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}_3$. Again, we decompose $\boldsymbol{v}_3$ by orthogonal projection

$$\boldsymbol{v}_3 = \boldsymbol{v}_3^{\|} + \boldsymbol{v}_3^{\perp}$$

  this time the parallel and orthogonal parts are taken relative to the subspace V_2. And we define $\boldsymbol{u}_3$ to be the unit vector in the $\boldsymbol{v}_3^{\perp}$ direction. Again, $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ span V_3 and are orthonormal.
- Rinse and repeat.

- Pictorially, the process is simple. You start with a bunch of vectors. You pick one out, scale it so it has length 1. Then take all the other ones and project onto the orthogonal subspace to the first. Then you iterate until you exhaust all the dimensions.

- The Gram-Schmidt process constructs a new basis from an old basis. We want a way to write down exactly what we did in this change of basis. This is captured in QR factorization.

- To do this, let's go back to our change of basis formula for an ordinary vector. Given a vector $\boldsymbol{x} \in X$, a P-dimensional subspace of $\mathbb{R}^M$, and the bases $\mathcal{B} = \{\boldsymbol{v}_i\}$, $\mathcal{C} = \{\boldsymbol{w}_i\}$ of X, as well as a basis $\mathcal{D} = \{\boldsymbol{u}_i\}$ of $\mathbb{R}^M$, remember that if we define the matrices

$$\mathcal{S} = \big[[\boldsymbol{v}_1]_{\mathcal{D}} \quad \ldots \quad [\boldsymbol{v}_P]_{\mathcal{D}}\big] \ , \quad \mathcal{T} = \big[[\boldsymbol{w}_1]_{\mathcal{D}} \quad \ldots \quad [\boldsymbol{w}_P]_{\mathcal{D}}\big] \ , \quad \mathcal{U} = \big[[\boldsymbol{w}_1]_{\mathcal{B}} \quad \ldots \quad [\boldsymbol{w}_P]_{\mathcal{B}}\big]$$

  then we have

$$[\boldsymbol{x}]_{\mathcal{D}} = \mathcal{S}[\boldsymbol{x}]_{\mathcal{B}} \ , \quad [\boldsymbol{x}]_{\mathcal{D}} = \mathcal{T}[\boldsymbol{x}]_{\mathcal{C}}$$

  and that

$$[\boldsymbol{x}]_{\mathcal{B}} = \mathcal{U}[\boldsymbol{x}]_{\mathcal{C}}$$

  Comparing them, we have

$$\mathcal{T} = \mathcal{S}\mathcal{U}$$

- Now let $\mathcal{D}$ be the standard basis in $\mathbb{R}^M$. $\mathcal{T}$ just becomes the matrix with column vectors equal to the basis vectors of $\mathcal{C}$, and $\mathcal{S}$ is the matrix with column vectors equal to the basis vectors of $\mathcal{B}$. So we have that

$$[\text{The matrix of basis vectors of } \mathcal{C}] = [\text{The matrix of the basis vectors of } \mathcal{B}] \times [\text{Something}]$$

  where that something is equal to

$$[\text{Something}] = [\text{Matrix for changing } \mathcal{C}\text{-coörd. to } \mathcal{B}\text{-coörd.}]$$

- This is called QR factorization. Given a P-dimensional subspace of $\mathbb{R}^M$, and letting $\boldsymbol{v}_1 \dots \boldsymbol{v}_P$ and $\boldsymbol{w}_1 \dots \boldsymbol{w}_P$ be bases of the subspace. If we write

$$\mathcal{M} = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_P \end{bmatrix} , \quad \mathcal{Q} = \begin{bmatrix} \boldsymbol{w}_1 & \dots & \boldsymbol{w}_P \end{bmatrix}$$

  and so $\mathcal{M}$ and $\mathcal{Q}$ are $M \times P$ matrices, we can find a $P \times P$ matrix $\mathcal{R}$ that gives

$$\mathcal{M} = \mathcal{Q}\mathcal{R}$$

- Now, let's bring our attention back to the Gram-Schmidt process. We were given a basis $\boldsymbol{v}_i$, which we apply the process to get an orthonormal basis $\boldsymbol{u}_i$. Then let $\mathcal{M}$ be the matrix of $\boldsymbol{v}_i$ and $\mathcal{Q}$ be the matrix of $\boldsymbol{u}_i$, we have that the matrix $\mathcal{R}$ in the QR factorization is given uniquely by

$$r_{11} = |\boldsymbol{v}_1| , \quad r_{jj} = |\boldsymbol{v}_j^{\perp}| \text{ for } 2 \leq j \leq M$$

  and

$$r_{ij} = \begin{cases} \boldsymbol{u}_i \cdot \boldsymbol{v}_j & \text{for } i < j \\ 0 & \text{for } i > j \end{cases}$$

  This, in particular, means that the matrix $\mathcal{R}$ is upper-triangular.

- Example: Suppose the matrix $\mathcal{M}$ is given by

$$\mathcal{M} = \begin{bmatrix} 1 & 1 & 7 & 1 & 7 \\ -2 & 3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 2 & -2 & 1 & -3 & 1 \\ 0 & -2 & -2 & 2 & -2 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

  find its QR factorization with $\mathcal{Q}$ the matrix of corresponding unit-vectors from the Gram-Schmidt process.

  – First we take $\boldsymbol{v}_1$ and normalize it, we get $r_{11} = |\boldsymbol{v}_1| = 3$ so $\boldsymbol{u}_1 = \boldsymbol{v}_1/3$.
  – Next we calculate $r_{12} = \boldsymbol{u}_1 \cdot \boldsymbol{v}_2 = -3$, so $\boldsymbol{v}_2^{\perp} = \boldsymbol{v}_2 + \boldsymbol{v}_1 = (2, 1, 0, 0, -2, 0)$. So we get $r_{22} = |\boldsymbol{v}_2^{\perp}| = 3$.

- ... iterate the calculation on the board ...
- Final answer:

$$\mathcal{Q} = \begin{bmatrix} 1/3 & 2/3 & 2/3 & 0 & 0 \\ -2/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 2/3 & 0 & -1/3 & 0 & 0 \\ 0 & -2/3 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad , \quad \mathcal{R} = \begin{bmatrix} 3 & -3 & 3 & -3 & 3 \\ 0 & 3 & 6 & 0 & 6 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

- Everything up til now will be on the midterm.

Homework for this week

5.1: 5, 12, 14, 16, 28, 29

5.2: 4, 14, 17, 28, 34, 39