Review for the Final Exam

Selected Topics

Orthogonal Sets

Definition

A set of *nonzero* vectors is **orthogonal** if each pair of vectors is orthogonal. It is **orthonormal** if, in addition, each vector is a unit vector.

Example: $\mathcal{B}_1 = \left\{ \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix} \right\}$ is not orthogonal.

Example: $\mathcal{B}_2 = \left\{ \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\-1\end{pmatrix} \right\}$ is orthogonal but not orthonormal.

Example: $\mathcal{B}_3 = \left\{ \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix}1\\-2\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix} \right\}$ is orthonormal.

To go from an orthogonal set $\{u_1, u_2, \ldots, u_m\}$ to an orthonormal set, replace each u_i with $u_i/\|u_i\|$.

Theorem

An orthogonal set is linearly independent. In particular, it is a basis for its span.

Orthogonal Projection

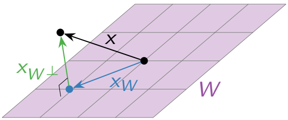
Let W be a subspace of $\mathbf{R}^n$, and let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an *orthogonal* basis for W. The **orthogonal projection** of a vector x onto W is

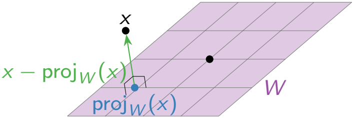
$$\mathrm{proj}_W(x) \overset{\mathrm{def}}{=} \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$

This is the closest vector to x that lies on W. In other words, the difference $x - \mathrm{proj}_W(x)$ is perpendicular to W: it is in $W^\perp$. Notation:

$$\boxed{x_W = \mathrm{proj}_W(x) \qquad x_{W^\perp} = x - \mathrm{proj}_W(x).}$$

So x_W is in W, $x_{W^\perp}$ is in $W^\perp$, and $x = x_W + x_{W^\perp}$.





Special case: If x is in W, then $x = \text{proj}_W(x)$, so

$$x = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$

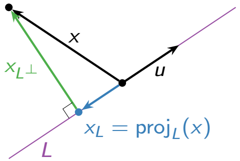
In other words, the $\mathcal{B}$-coordinates of x are

$$\left(\frac{x \cdot u_1}{u_1 \cdot u_1},\ \frac{x \cdot u_2}{u_1 \cdot u_2},\ \ldots,\ \frac{x \cdot u_m}{u_1 \cdot u_m} \right),$$

where $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$, an orthogonal basis for W.

Special case: If $W = L$ is a line, then $L = \text{Span}\{u\}$ for some nonzero vector u, and

$$\text{proj}_L(x) = \frac{x \cdot u}{u \cdot u}\, u$$



Let W be a subspace of $\mathbf{R}^n$.

Theorem
The orthogonal projection proj_W is a *linear* transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$. Its range is W.

If A is the matrix for proj_W, then $A^2 = A$ because projecting twice is the same as projecting once: $\mathrm{proj}_W \circ \mathrm{proj}_W = \mathrm{proj}_W$.

Theorem
The only eigenvalues of A are 1 and 0.

Why?
$$Av = \lambda v \implies A^2 v = A(Av) = A(\lambda v) = \lambda(Av) = \lambda^2 v.$$

So if λ is an eigenvalue of A, then λ^2 is an eigenvalue of A^2. But $A^2 = A$, so $\lambda^2 = \lambda$, and hence $\lambda = 0$ or 1.

The 1-eigenspace of A is W, and the 0-eigenspace is $W^\perp$.

The Gram–Schmidt Process

Let $\{v_1, v_2, \ldots, v_m\}$ be a basis for a subspace W of $\mathbf{R}^n$. Define:

1. $u_1 = v_1$
2. $u_2 = v_2 - \mathrm{proj}_{\mathrm{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1$
3. $u_3 = v_3 - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2\}}(v_3) \qquad = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$\vdots$

m. $u_m = v_m - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2, \ldots, u_{m-1}\}}(v_m) \ = v_m - \displaystyle\sum_{i=1}^{m-1} \frac{v_m \cdot u_i}{u_i \cdot u_i}\, u_i$

Then $\{u_1, u_2, \ldots, u_m\}$ is an *orthogonal* basis for the same subspace W.

In fact, for each i,

$$\mathrm{Span}\{u_1, u_2, \ldots, u_i\} = \mathrm{Span}\{v_1, v_2, \ldots, v_i\}.$$

Note if v_i is in $\mathrm{Span}\{v_1, v_2, \ldots, v_{i-1}\} = \mathrm{Span}\{u_1, u_2, \ldots, u_{i-1}\}$, then $v_i = \mathrm{proj}_{\mathrm{Span}\{u_1, u_2, \ldots, u_{i-1}\}}(v_i)$, so $u_i = 0$. So this also detects linear dependence.

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Step 1: Let $v_1, v_2, \ldots, v_m$ be the columns of A. Run Gram–Schmidt on $\{v_1, v_2, \ldots, v_m\}$ to get an orthogonal basis $\{u_1, u_2, \ldots, u_m\}$, and solve for each v_i in terms of $u_1, u_2, \ldots, u_i$.

Step 2: Put the resulting equations in matrix form to get $A = \widehat{Q}\widehat{R}$ where

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix} \qquad \widehat{Q} = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{pmatrix}$$

and $\widehat{R}$ contains the coefficients from $v_i =$ (linear combination of $u_1, u_2, \ldots, u_{i-1}$) in the columns.

Step 3: Scale each column of $\widehat{Q}$ by its length to get a matrix with orthonormal columns, and scale each row of $\widehat{R}$ by the opposite factor to get Q and R, respectively.

Find the QR factorization of $A = \begin{pmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & -2 \\ 1 & -1 & 0 \end{pmatrix}$.

Step 1: Let v_1, v_2, v_3 be the columns. Run Gram–Schmidt and solve for v_1, v_2, v_3 in terms of u_1, u_2, u_3:

$$u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_1 = u_1$$

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = v_2 - \frac{3}{2} u_1 = \begin{pmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{pmatrix} \qquad v_2 = \frac{3}{2} u_1 + u_2$$

$$u_3 = v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2 = v_3 + \frac{4}{5} u_2 = \begin{pmatrix} 2 \\ 0 \\ 0 \\ -2 \end{pmatrix} \qquad v_3 = -\frac{4}{5} u_2 + u_3$$

$$v_1 = 1\, u_1 \qquad v_2 = \frac{3}{2}\, u_1 + 1\, u_2 \qquad v_3 = 0\, u_1 - \frac{4}{5}\, u_2 + 1\, u_3$$

Step 2: write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

$$\widehat{Q} = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix}$$

$$\widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \widehat{Q} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix} \qquad \widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: normalize the columns of $\widehat{Q}$ and the rows of $\widehat{R}$ to get Q and R:

$$Q = \begin{pmatrix} | & | & | \\ u_1/\|u_1\| & u_2/\|u_2\| & u_3/\|u_3\| \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} \end{pmatrix}$$

$$R = \begin{pmatrix} 1 \cdot \|u_1\| & 3/2 \cdot \|u_1\| & 0 \cdot \|u_1\| \\ 0 & 1 \cdot \|u_2\| & -4/5 \cdot \|u_2\| \\ 0 & 0 & 1 \cdot \|u_3\| \end{pmatrix} = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 5 & -4 \\ 0 & 0 & 2\sqrt{2} \end{pmatrix}$$

The final QR decomposition is

$$A = QR \qquad Q = \begin{pmatrix} 1/2 & -1/2 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} \end{pmatrix} \qquad R = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 5 & -4 \\ 0 & 0 & 2\sqrt{2} \end{pmatrix}.$$

Subspaces

Definition

A **subspace** of $\mathbf{R}^n$ is a subset V of $\mathbf{R}^n$ satisfying:

1. The zero vector is in V. "not empty"
2. If u and v are in V, then $u + v$ is also in V. "closed under addition"
3. If u is in V and c is in $\mathbf{R}$, then cu is in V. "closed under $\times$ scalars"

Examples:

- Any $\mathrm{Span}\{v_1, v_2, \ldots, v_m\}$.
- The *column space* of a matrix: $\mathrm{Col}\, A = \mathrm{Span}\{\text{columns of } A\}$.
- The range of a linear transformation (same as above).
- The *null space* of a matrix: $\mathrm{Nul}\, A = \{x \mid Ax = 0\}$.
- The *row space* of a matrix: $\mathrm{Row}\, A = \mathrm{Span}\{\text{rows of } A\}$.
- The λ-eigenspace of a matrix, where λ is an eigenvalue.
- The orthogonal complement $W^\perp$ of a subspace W.
- The zero subspace $\{0\}$.
- All of $\mathbf{R}^n$.

Subspaces and Bases

Definition

Let V be a subspace of $\mathbf{R}^n$. A **basis** of V is a set of vectors $\{v_1, v_2, \ldots, v_m\}$ in $\mathbf{R}^n$ such that:

1. $V = \mathrm{Span}\{v_1, v_2, \ldots, v_m\}$, and
2. $\{v_1, v_2, \ldots, v_m\}$ is linearly independent.

The number of vectors in a basis is the **dimension** of V, and is written $\dim V$.

Every subspace has a basis, so every subspace is a span. But subspaces have many different bases, and some might be better than others. For instance, Gram–Schmidt takes a basis and produces an *orthogonal* basis. Or, diagonalization produces a basis of *eigenvectors* of a matrix.

How do I know if a subset V is a subspace or not?

- Can you write V as one of the examples on the previous slide?
- If not, does it satisfy the three defining properties?

Note on subspaces versus subsets: A **subset** of $\mathbf{R}^n$ is any collection of vectors whatsoever. Like, the unit circle in $\mathbf{R}^2$, or all vectors with whole-number coefficients. A *subspace* is a subset that satisfies three additional properties. Most subsets are not subspaces.

Definition

Two $n \times n$ matrices A and B are **similar** if there is an invertible $n \times n$ matrix P such that

$$A = PBP^{-1}.$$

Important Facts:

1. Similar matrices have the same characteristic polynomial.
2. It follows that similar matrices have the same eigenvalues.
3. If A is similar to B and B is similar to C, then A is similar to C.

Caveats:

1. Matrices with the same characteristic polynomial need not be similar.
2. Similarity has nothing to do with row equivalence.
3. Similar matrices usually do not have the same eigenvectors.

Let $A = PBP^{-1}$, and let $v_1, v_2, \ldots, v_n$ be the columns of P. These form a basis $\mathcal{B}$ for $\mathbf{R}^n$ because P is invertible. *Key relation:* for any vector x in $\mathbf{R}^n$,

$$[Ax]_{\mathcal{B}} = B[x]_{\mathcal{B}}.$$

This says:

A acts on the usual coordinates of x
in the same way that
B acts on the $\mathcal{B}$-coordinates of x.

Example:

$$A = \frac{1}{4}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \qquad P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then $A = PBP^{-1}$. B acts on the usual coordinates by scaling the first coordinate by 2, and the second by $1/2$:

$$B\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ x_2/2 \end{pmatrix}.$$

The unit coordinate vectors are eigenvectors: e_1 has eigenvalue 2, and e_2 has eigenvalue $1/2$.

Similarity

Example

$$A = \frac{1}{4}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \qquad P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad [Ax]_{\mathcal{B}} = B[x]_{\mathcal{B}}.$$

In this case, $\mathcal{B} = \left\{ \binom{1}{1}, \binom{1}{-1} \right\}$. Let $v_1 = \binom{1}{1}$ and $v_2 = \binom{1}{-1}$.

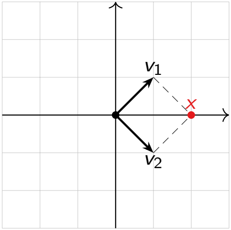
To compute $y = Ax$:

1. Find $[x]_{\mathcal{B}}$.
2. $[y]_{\mathcal{B}} = B[x]_{\mathcal{B}}$.
3. Compute y from $[y]_{\mathcal{B}}$.

Say $x = \binom{2}{0}$.

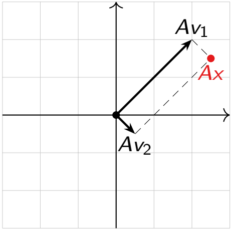
1. $x = v_1 + v_2$ so $[x]_{\mathcal{B}} = \binom{1}{1}$.
2. $[y]_{\mathcal{B}} = B\binom{1}{1} = \binom{2}{1/2}$.
3. $y = 2v_1 + \frac{1}{2}v_2 = \binom{5/2}{3/2}$.

Picture:



A

A scales the v_1-coordinate by 2, and the v_2-coordinate by $\frac{1}{2}$.



Definition

A matrix equation $Ax = b$ is **consistent** if it has a solution, and **inconsistent** otherwise.

If A has columns $v_1, v_2, \ldots, v_n$, then

$$b = Ax = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n.$$

So if $Ax = b$ has a solution, then b is a linear combination of $v_1, v_2, \ldots, v_n$, and conversely. Equivalently, b is in $\mathrm{Span}\{v_1, v_2, \ldots, v_n\} = \mathrm{Col}\, A$.

Important

$Ax = b$ is consistent if and only if b is in $\mathrm{Col}\, A$.

Least-Squares Solutions

Suppose that $Ax = b$ is *in*consistent. Let $\widehat{b} = \mathrm{proj}_{\mathrm{Col}\,A}(b)$ be the closest vector for which $A\widehat{x} = \widehat{b}$ *does* have a solution.

Definition

A solution to $A\widehat{x} = \widehat{b}$ is a **least squares solution** to $Ax = b$. This is the solution $\widehat{x}$ for which $A\widehat{x}$ is *closest* to b (with respect to the usual notion of distance in $\mathbf{R}^n$).

Theorem

The least-squares solutions to $Ax = b$ are the solutions to

$$A^T A\widehat{x} = A^T b.$$

If A has *orthogonal* columns $u_1, u_2, \ldots, u_n$, then the least-squares solution is

$$\widehat{x} = \left(\frac{x \cdot u_1}{u_1 \cdot u_1}, \frac{x \cdot u_2}{u_2 \cdot u_2}, \cdots, \frac{x \cdot u_m}{u_m \cdot u_m} \right)$$

because

$$A\widehat{x} = \widehat{b} = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$