Lecture 16: Cauchy-Schwarz, triangle inequality, orthogonal projection, and Gram-Schmidt orthogonalization (1)

Travis Schedler
Thurs, Nov 4, 2010 (version: Thurs, Nov 4, 4:40 PM)

Goals (2)

• Corrected change-of-basis formula!
• The Cauchy-Schwarz inequality
• Triangle inequality
• Orthogonal projection and least-squares approximations
• Orthonormal bases
• Gram-Schmidt orthogonalization

Corrected change-of-basis formula! (3)

Take the linear transformation $T := L_A \in \mathcal{L}({\text{Mat}}(2, 1, \mathbb{F}))$ of left-multiplication by the matrix

$$A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$ 

Now, in the standard basis $(v_1, v_2)$, $\mathcal{M}(T) = A$. Consider the new basis $v'_1 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v'_2 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then we get

$$Av'_1 = \begin{pmatrix} 3 \\ 7 \end{pmatrix} = 3v'_1 + 4v'_2, \quad Av'_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2v'_1 + 2v'_2.$$

Thus,

$$\mathcal{M}_{(v'_1, v'_2)}(T) = \begin{pmatrix} 3 & 2 \\ 4 & 2 \end{pmatrix}. $$
Change-of-basis continued (4)

Incorrect change-of-basis matrix: \( (v'_1 \ v'_2) = S (v_1 \ v_2) \), i.e., \( S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \). Then

\[
S^{-1} M_{(v_1, v_2)}(T) S = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 4 & 2 \end{pmatrix} = M_{(v'_1, v'_2)}(T).
\]

Correct one is the transpose: \( (v'_1 \ v'_2) = (v_1 \ v_2) S \). Then

\[
S^{-1} M_{(v_1, v_2)}(T) S = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -2 & -4 \\ 3 & 7 \end{pmatrix} \neq M_{(v'_1, v'_2)}(T).
\]

Correct change-of-basis and triangular changes-of-basis (5)

- Correct: Change of basis matrix from \( (v_i) \) to \( (v'_i) \) is: \( S = M_{(v'_i), (v_i)}(I) \) = put on \( i \)-th column the coefficients of \( v'_i \) in terms of \( v_1, \ldots, v_n \). (cf. Homework 5.)
- I.e. \( S = (s_{ij}) \) with \( v'_i = s_{1i}v_1 + \cdots + s_{ni}v_n \).
- If \( T \in L(\text{Mat}(n, 1, F)) \) and \( (v_i) = \) standard basis, then the change-of-basis is \( S = (v'_1 \cdots v'_n) \): put columns together.

**Corollary 1** (similar to Proposition 5.12). The change-of-basis matrix \( S = (s_{ij}) \) from \( (v_i) \) to \( (v'_i) \) is upper-triangular if and only if \( v'_i = s_{1i}v_1 + \cdots + s_{ii}v_i \) for all \( i \), i.e., \( s_{ji} = 0 \) for \( j > i \).

**Corollary 2** (cf. Proposition 5.12). Let \( S \) be an upper-triangular change of basis. Then \( M_{(v_i)}(T) \) is upper-triangular if \( M_{(v'_i)}(T) \) is upper-triangular.

Proof: If \( A \) is upper triangular, so is \( S^{-1}AS \), and vice-versa.

The Cauchy-Schwarz inequality (6)

**Theorem 3** (6.6: Cauchy-Schwarz inequality). \( |\langle u, v \rangle| \leq ||u|| ||v|| \). Moreover, this is an equality iff one of \( u \) and \( v \) is a scalar multiple of the other.

**Proof.**

- It is enough to suppose \( v \neq 0 \). Write \( u = \text{proj}_v(u) + w \).

  - Then \( ||u||^2 = ||\text{proj}_v(u) + w||^2 = ||\text{proj}_v(u)||^2 + ||w||^2 \geq ||\frac{\langle u, v \rangle}{\langle v, v \rangle} v||^2 = \left|\frac{\langle u, v \rangle}{\langle v, v \rangle}\right| \langle v, v \rangle = \frac{|\langle u, v \rangle|^2}{||v||^2} \).

  - Hence, \( ||u|| ||v|| \geq |\langle u, v \rangle|^2 \) as desired.

  - Equality holds if \( w = 0 \), i.e., \( \text{proj}_v(u) = u \), i.e., \( u \) is a multiple of \( v \). \( \square \)
Proof of triangle inequality (7)

Theorem: \(|u + v| \leq |u| + |v|\). Equality holds iff one of \(u, v\) is a positive multiple of the other.

\[
\|u + v\|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \langle v, v \rangle + (\langle u, v \rangle + \langle v, u \rangle)
\]

\[
= \|u\|^2 + \|v\|^2 + 2 \text{Re}\langle u, v \rangle
\]

\[
\leq \|u\|^2 + \|v\|^2 + 2\|u\||v|,
\]

\[
\leq \|u\|^2 + \|v\|^2 + 2|u||v|
\]

\[
= (\|u\| + |v|)^2.
\]

Here, \(\text{Re}\) is the real part: \(\text{Re}(a + bi) := a\) for \(a, b \in \mathbb{R}\). Thus \(2\text{Re}(z) = z + \overline{z}\).

Equality holds in the first inequality iff \(\langle u, v \rangle\) is nonnegative. Equality holds in the second inequality iff one of \(u, v\) is a multiple of the other. For both equalities to hold: say \(u = \lambda v\) and \(\langle u, v \rangle \geq 0\): then \(\lambda(v, v) \geq 0\), i.e., \(\lambda \geq 0\).

Orthogonal complement (8)

Definition 4. Let \(U \subseteq V\). Then, \(U^\perp := \{u' \in V : \langle u', u \rangle = 0, \forall u \in U\}\).

Theorem 5 (Theorem 6.29). \(U^\perp\) is a complement to \(U\), i.e., \(V = U \oplus U^\perp\).

Proof. 

• First, \(u \in U \cap U^\perp\) implies \(\langle u, u \rangle = 0\), i.e., \(u = 0\). So \(U \cap U^\perp = \{0\}\).

• Let \((u_1, \ldots, u_k)\) be a basis of \(U\). Consider the map \(T : V \to \mathbb{F}^k\) given by \(T(v) = ((v, u_1), \ldots, (v, u_k))\).

• Then, \(U^\perp = \text{null} T\).

• Since \(U \cap U^\perp = 0\), \(T|_U\) is injective.

• Since \(\dim U = k = \dim \mathbb{F}^k\), \(T|_U\) is also surjective.

• By rank-nullity, \(\dim V = \dim U + \dim U^\perp\). So \(V = U \oplus U^\perp\). \(\Box\)

Orthogonal projection (9)

Let \(U \subseteq V\). Then \(P_{U,U^\perp}\) is the orthogonal projection.

Proposition 0.1 (Proposition 6.36). Let \(v \in V\). Then, for all \(u \in U\), \(\|v - P_{U,U^\perp}(v)\| \leq \|v - u\|\).

Thus, \(u := P_{U,U^\perp}(v)\) is the best approximation of \(v\) by an element of \(U\). It is also called the least-squares approximation because, when \(V = \mathbb{F}^n\), it minimizes \(\|v - u\| = \sum_{i=1}^{n} |v_i - u_i|^2\).

Proof. 

• Write \(v - u = (v - P_{U,U^\perp}(v)) + (P_{U,U^\perp}(v) - u)\).

• By definition, \(v - P_{U,U^\perp}(v) \in U^\perp\), and \(P_{U,U^\perp}(v) - u \in U\).

• By the Pythagorean theorem, \(\|v - u\|^2 = \|v - P_{U,U^\perp}(v)\|^2 + \|P_{U,U^\perp}(v) - u\|^2\). \(\Box\)

Note: equality holds iff \(u = P_{U,U^\perp}(v)\).
Orthonormal bases (10)

Lemma 6. Suppose that \( v_1, \ldots, v_k \) is a set of nonzero vectors such that \( v_i \perp v_j \) for all \( i \neq j \). Then \( v_1, \ldots, v_k \) are linearly independent.

Proof. • Observe \( \langle v_i, \lambda_1 v_1 + \cdots + \lambda_k v_k \rangle = \lambda_i \langle v_i, v_i \rangle \).
  • If \( \lambda_i \neq 0 \), this is nonzero. Hence \( \lambda_1 v_1 + \cdots + \lambda_k v_k \neq 0 \).
  • Thus, there can be no linear dependence. \qed

Definition 7. An orthonormal basis \( v_1, \ldots, v_n \) of \((V, \langle -,- \rangle)\) is a basis of \( V \) such that \( \langle v_i, v_j \rangle = \delta_{ij} \).

Corollary 8. If \( \dim V = n \), then any orthonormal list of length \( n \) is an orthonormal basis.

Formula for orthogonal projection (11)

Proposition 0.2 (6.35). Let \( e_1, \ldots, e_k \) be an orthonormal basis of \( U \). Then,

\[
P_{U,U^\perp}(v) = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_k \rangle e_k = \text{proj}_{e_1} v + \cdots + \text{proj}_{e_k} v.
\]

Corollary (Theorem 6.17): when \( U = V \) and \( \dim V = n \), \( v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \).

Proof. • Let \( u \) be the RHS. Then \( \langle v - u, e_i \rangle = \langle v, e_i \rangle - \langle v, e_i \rangle = 0 \) for all \( i \).
  • Hence, \( v - u \in U^\perp \). Since \( u \in U \) and \( v = u + (v - u) \), we deduce \( u = P_{U,U^\perp}(v) \). \qed

Gram-Schmidt orthogonalization (12)

Theorem 9 (6.20: Gram-Schmidt). Given a basis \( (v_1, \ldots, v_n) \) of \( V \), there is an upper-triangular change of basis \( e_i := s_{i1} v_1 + \cdots + s_{ii} v_i \) so that \( (e_1, \ldots, e_n) \) is orthonormal. Moreover, the \( s_{ij} \) are computable by an effective algorithm!

Proof. • By induction on \( n = \dim V \); base case \( \dim V = 0 \).
  • Let \( U := \text{Span}(v_1, \ldots, v_{n-1}) \). Inductively, form the orthonormal basis \( (e_1, \ldots, e_{n-1}) \) of \( U \) via upper-triangular change of basis.
  • It suffices to find \( e_n \) such that \( \langle e_n, e_j \rangle = \delta_{nj} \) (then automatically upper-triangular).
  • Set \( e'_n := v_n - P_{U,U^\perp}(v_n) \in U^\perp \). So \( \langle e'_n, e_j \rangle = 0 \) for \( j \neq n \).
  • \( e'_n \) is nonzero since \( v_n \notin U \). Set \( e_n := e'_n / \|e'_n\| \). Then \( \langle e_n, e_j \rangle = \delta_{nj} \). \qed
Corollaries (13)

**Corollary 10** (Corollary 6.24). Every finite-dimensional inner-product space has an orthonormal basis.

Proof: Apply Gram-Schmidt to an arbitrary basis.

**Corollary 11** (Corollary 6.25). Every orthonormal list of vectors in \( V \) can be extended to an orthonormal basis of \( V \).

Proof: Extend to an arbitrary basis and perform Gram-Schmidt.

**Corollary 12** (Corollary 6.27). Suppose that \( M(T) \) is (block) upper-triangular in some basis. Then there exists an orthonormal basis where it is still (block) upper triangular.

Proof: Pick an arbitrary basis in which the matrix is as desired, and then apply Gram-Schmidt. It is an upper-triangular change of basis, so it preserves (block) upper-triangularity.

Further corollaries (14)

**Corollary 13** (Corollary 6.27+). Over \( \mathbb{C} \), there exists an orthonormal basis in which \( M(T) \) is upper-triangular. Over \( \mathbb{R} \), there exists an orthonormal basis in which it is block upper-triangular with diagonal blocks of size \( 1 \times 1 \) or \( 2 \times 2 \).

**Corollary 14** (Corollary 6.33). \( (U^\perp)^\perp = U \).

Proof: Clearly \( U \subseteq (U^\perp)^\perp \). They are both complements to \( U^\perp \), so they have the same dimension. Hence they are equal.

Norm formulas (15)

**Proposition 0.3** (Proposition 6.15). Let \( (e_1, \ldots, e_k) \) be an orthonormal list. Then \( |a_1 e_1 + \cdots + a_k e_k|^2 = |a_1|^2 + \cdots + |a_k|^2 \).

Proof: Repeated application of the Pythagorean theorem, since \( \langle a_i e_i, a_i e_i \rangle = |a_i|^2 \).

**Corollary 15.** Let \( (e_1, \ldots, e_k) \) be an orthonormal basis of \( U \). Then \( |P_{U,U^\perp}(v)|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_k \rangle|^2 \).

Proof: Apply the previous corollary and \( P_{U,U^\perp}(v) = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_k \rangle e_k \). When \( U = V \), we get \( |v|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2 \) (cf. Theorem 6.17).