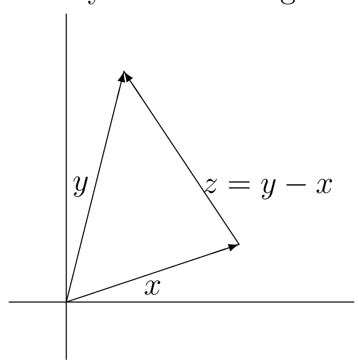
Chapter 6

Orthogonality

6.1 Orthogonal Vectors and Subspaces

Recall that if nonzero vectors $x, y \in \mathbb{R}^n$ are linearly independent then the subspace of all vectors $\alpha x + \beta y$, $\alpha, \beta \in \mathbb{R}$ (the space spanned by x and y) form a plane. All three vectors x, y, and $(x - y)$ lie in this plane and they form a triangle:



where this page represents the plane in which all of these vectors lie.

Vectors x and y are considered to be orthogonal (perpendicular) if they meet at a right angle. Using the Euclidean length $\|x\|_2 = \sqrt{\chi_0^2 + \cdots + \chi_{n-1}^2} = \sqrt{x^T x}$, we find that the Pythagorean Theorem dictates that *if* the angle in the triangle where x and y meet is a right angle, then $\|z\|_2^2 = \|x\|_2^2 + \|y\|_2^2$. In this case,

$$\begin{aligned}\|x\|_2^2 + \|y\|_2^2 &= \|z\|_2^2 = \|y - x\|_2^2 = (y-x)^T(y-x) = (y^T - x^T)(y-x) = (y^T - x^T)y - (y^T - x^T)x \\ &= \underbrace{y^T y}_{\|y\|_2} - \underbrace{(x^T y + y^T x)}_{2x^T y} + \underbrace{x^T x}_{\|x\|_2} = \|x\|_2 - 2x^T y + \|y\|_2,\end{aligned}$$

in other words,

$$\|x\|_2 \|y\|_2 = \|x\|_2 - 2x^T y + \|y\|_2.$$

Cancelling terms on the left and right of the equality, this implies that $x^T y = 0$. This motivates the following definition:

Definition 6.1 *Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal if $x^T y = 0$.*

Sometimes we will use the notation $x \perp y$ to indicate that x is perpendicular to y.

We can extend this to define orthogonality of two subspaces:

Definition 6.2 *Let $\mathbf{V}, \mathbf{W} \subset \mathbb{R}^n$ be subspaces. Then $\mathbf{V}$ and $\mathbf{W}$ are said to be orthogonal if $v \in \mathbf{V}$ and $w \in \mathbf{W}$ implies that $v^T w = 0$.*

We will use the notation $\mathbf{V} \perp \mathbf{W}$ to indicate that subspace $\mathbf{V}$ is orthogonal to subspace $\mathbf{W}$.

Definition 6.3 *Given subspace $\mathbf{V} \subset \mathbb{R}^n$, the set of all vectors in $\mathbb{R}^n$ that are orthogonal to $\mathbf{V}$ is denoted by $\mathbf{V}^\perp$ (pronounced as "V-perp").*

Exercise 6.4 *Let $\mathbf{V}, \mathbf{W} \subset \mathbb{R}^n$ be subspaces. Show that if $\mathbf{V} \perp \mathbf{W}$ then $\mathbf{V} \cap \mathbf{W} = \{0\}$, the zero vector.*

Whenever $\mathbf{V} \cap \mathbf{W} = \{0\}$ we will sometimes call this the *trivial intersection*. Trivial in the sense that it only contains the zero vector.

Exercise 6.5 *Show that if $\mathbf{V} \in \mathbb{R}^n$ is a subspace, then $\mathbf{V}^\perp$ is a subspace.*

Let us recall some definitions:

Definition 6.6 *The column space of a matrix $A \in \mathbb{R}^{m \times n}$, $\mathcal{C}(\mathcal{A})$, equals the set of all vectors in $\mathbb{R}^m$ that can be written as Ax: $\{y \mid y = Ax\}$.*

Definition 6.7 *The null space of a matrix $A \in \mathbb{R}^{m \times n}$, $\mathcal{N}(A)$, equals the set of all vectors in $\mathbb{R}^n$ that map to the zero vector: $\{x \mid Ax = 0\}$.*

Definition 6.8 *The row space of a matrix $A \in \mathbb{R}^{m \times n}$, $\mathcal{R}(\mathcal{A})$, equals the set of all vectors in $\mathbb{R}^n$ that can be written as $A^T x$: $\{y \mid y = A^T x\}$.*

Definition 6.9 *The left null space of a matrix $A \in \mathbb{R}^{m \times n}$ equals the set of all vectors x in $\mathbb{R}^m$ such $x^T A = 0$.*

Exercise 6.10 *Show that the left null space of a matrix $A \in \mathbb{R}^{m \times n}$ equals $\mathcal{N}(A^T)$.*

Theorem 6.11 *Let $A \in \mathbb{R}^{m \times n}$. Then the null space of A is orthogonal to the row space of A: $\mathcal{R}(\mathcal{A}) \perp \mathcal{N}(\mathcal{A})$.*

Proof: Assume that $y \in \mathcal{R}(\mathcal{A})$ and $x \in \mathcal{N}(A)$. Then there exists a vector $z \in \mathbb{R}^n$ such that $y = A^T z$. Now, $y^T x = (A^T z)^T x = (z^T A) x = z^T (Ax) = 0$.

Theorem 6.12 *The dimension of $\mathcal{R}(\mathcal{A})$ equals the dimension of $\mathcal{C}(\mathcal{A})$.*

Proof: We saw this in Chapter 2 of Strang's book: The dimension of the row space equals the number of linearly independent rows, which equals the number of nonzero rows that result from the Gauss-Jordan method, which equals the number of pivots that show up in that method, which equals the number of linearly independent columns.

Theorem 6.13 *Let $A \in \mathbb{R}^{m \times n}$. Then every $x \in \mathbb{R}^n$ can be written as $x = x_r + x_n$ where $x_r \in \mathcal{R}(\mathcal{A})$ and $x_n \in \mathcal{N}(A)$.*

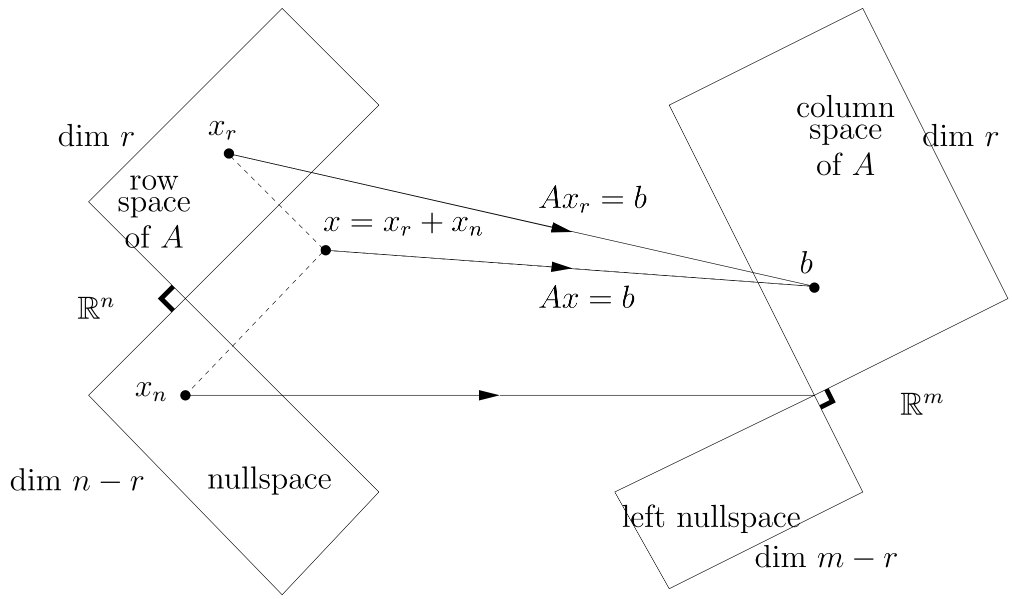


Figure 6.1: A pictorial description of how $x = x_r + x_n$ is transformed by $A \in \mathbb{R}^{m \times n}$ into $b = Ax = A(x_r + x_n)$. (Blatently borrowed from Strang's book.) Important to note: (1) The row space is orthogonal to the nullspace; (2) The column space is orthogonal to the left nullspace; (3) $n = r + (n - r) = \dim(\mathcal{R}(\mathcal{A})) + \dim(\mathcal{N}(\mathcal{A}))$; and (4) $m = r + (m - r) = \dim(\mathcal{C}(\mathcal{A})) + \dim(\mathcal{N}(\mathcal{A}^{\mathcal{T}}))$.

Proof: Recall from Chapter 2 of Strang's book that the dimension of $\mathcal{N}(A)$ and the dimension of $\mathcal{C}(\mathcal{A})$, r, add to the number of columns, n. Thus, the dimension of $\mathcal{R}(\mathcal{A})$ equals r and the dimension of $\mathcal{N}(A)$ equals $n - r$. If $x \in \mathbb{R}^n$ cannot be written as $x_r + x_n$ as indicated, then consider the set of vectors that consists of the union of a basis of $\mathcal{R}(\mathcal{A})$ and a basis of $\mathcal{N}(A)$, plus the vector x. This set is linearly independent and has $n+1$ vectors, contradicting the fact that $\mathbb{R}^n$ has dimension n.

Theorem 6.14 *Let $A \in \mathbb{R}^{m \times n}$. Then A is a one-to-one, onto mapping from $\mathcal{R}(\mathcal{A})$ to $\mathcal{C}(\mathcal{A})$.*

Proof: Let $A \in \mathbb{R}^{m \times n}$. We need to show that

- A maps $\mathcal{R}(\mathcal{A})$ to $\mathcal{C}(\mathcal{A})$. This is trivial, since any vector $x \in \mathbb{R}^m$ maps to $\mathcal{C}(\mathcal{A})$.
- Uniqueness: We need to show that if $x, y \in \mathcal{R}(\mathcal{A})$ and $Ax = Ay$ then $x = y$. Notice that $Ax = Ay$ implies that $A(x - y) = 0$, which means that $(x - y)$ is both in $\mathcal{R}(\mathcal{A})$ (since it is a linear combination of x and y, both of which are in $\mathcal{R}(\mathcal{A})$) and in $\mathcal{N}(A)$. Since we just showed that these two spaces are orthogonal, we conclude that $(x-y) = 0$, the zero vector. Thus $x = y$.

- Onto: We need to show that for any $b \in \mathcal{C}(\mathcal{A})$ there exists $x_r \in \mathcal{R}(\mathcal{A})$ such that $Ax_r = b$. Notice that if $b \in \mathcal{C}$, then there exists $x \in \mathbb{R}^n$ such that $Ax = b$. By Theorem 6.13, $x = x_r + x_n$ where $x_r \in \mathcal{R}(\mathcal{A})$ and $x_n \in \mathcal{N}(A)$. Then $b = Ax = A(x_r + x_n) = Ax_r + Ax_n = Ax_r$.

Theorem 6.15 *Let $A \in \mathbb{R}^{m\times n}$. Then the left null space of A is orthogonal to the columns space of A and the dimension of the left null space of A equals $m-r$, where r is the dimension of the column space of A.*

Proof: This follows trivially by applying the previous theorems to A^T.

The last few theorems are summarized in Figure 6.1.

6.2 Motivating Example, Part I

Let us consider the following set of points:

$$(\chi_0, \psi_0) = (1, 1.97), (\chi_1, \psi_1) = (2, 6.97), (\chi_2, \psi_2) = (3, 8.89), (\chi_3, \psi_3) = (4, 10.01),$$

which we plot in Figure 6.2. What we would like to do is to find a line that interpolates these points as near as is possible, in the sense that the sum of the square of the distances to the line are minimized. Let us express this with matrices and vectors.

Let

$$x = \begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \\ \chi_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \qquad y = \begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} 1.97 \\ 6.97 \\ 8.89 \\ 10.01 \end{pmatrix}.$$

If we give the equation of the line as $y = \gamma_0 + \gamma_1 x$ then, **IF** this line **COULD** go through all these points **THEN** the following equations would have to be simultaneously satified:

$$\begin{array}{rcl} \psi_0 &=& \gamma_0 + \gamma_1\chi_1 \\ \psi_1 &=& \gamma_0 + \gamma_1\chi_2 \\ \psi_2 &=& \gamma_0 + \gamma_1\chi_3 \\ \psi_3 &=& \gamma_0 + \gamma_1\chi_4 \end{array} \quad \text{or, specifically,} \quad \begin{array}{rcl} 1.97 &=& \gamma_0 + \gamma_1 \\ 6.97 &=& \gamma_0 + 2\gamma_1 \\ 8.89 &=& \gamma_0 + 3\gamma_1 \\ 10.01 &=& \gamma_0 + 4\gamma_1 \end{array}$$

which can be written in matrix notation as

$$\begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} 1 & \chi_0 \\ 1 & \chi_1 \\ 1 & \chi_2 \\ 1 & \chi_3 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} \quad \text{or, specifically,} \quad \begin{pmatrix} 1.97 \\ 6.97 \\ 8.89 \\ 10.01 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix}.$$

Figure 6.2: Left: A plot of the points. Right: Some first guess of a line that interpolates, and the distance from the points to the line.

Now, just looking at Figure 6.2, it is obvious that these point do not lie on the same line and that therefore all these equations cannot be simultaneously satified. **So, what do we do?**

How does it relate to column spaces? The first question we ask is "For what right-hand sides could we have solved all four equations simultaneously?" We would have had to choose y so that $Ac = y$, where

$$A = \begin{pmatrix} 1 & \chi_0 \\ 1 & \chi_1 \\ 1 & \chi_2 \\ 1 & \chi_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix}.$$

This means that y must be in the column space of A. It must be possible to express it as $y = \gamma_0 a_0 + \gamma_1 a_1$, where $A = \left(\begin{array}{c|c} a_0 & a_1 \end{array} \right)$! What does this mean if we relate this back to the picture? Only if $\{\psi_0, \cdots, \psi_3\}$ have the property that $\{(1, \psi_0), \cdots, (4, \psi_3)\}$ lie on a line can we find coefficients γ_0 and γ_1 such that $Ac = y$.

How does this problem relate to orthogonality and projection? The problem is that the given y does **not** lie in the column space of A. So, a question is what vector, z, that **does** lie in the column space we should use to solve $Ac = z$ instead so that we end up with a line that best interpolates the given points. Now, if z solves $Ac = z$ exactly, then $z = \left(a_0 \mid a_1 \right) \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} = \gamma_0 a_0 + \gamma_1 a_1$, which is of course just a repeat of the observation that z is in the column space of A. Thus, what we want is $y = z + w$, where w is as small (in length) as possible. This happens when w is orthogonal to z! So, $y = \gamma_0 a_0 + \gamma_1 a_1 + w$, with $a_0^T w = a_1^T w = 0$. The vector z in the column space of A that is closest to y is known as the projection of y onto the column space of A. So, it would be nice to have a way of finding a way to compute this projection.

6.3 Solving a Linear Least-Squares Problem

The last problem motivated the following general problem: Given m equations in n unknowns, we end up with a system $Ax = b$ where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$.

- This system of equations may have no solutions. This happens when b is not in the column space of A.
- This system may have a unique solution. This happens only when $r = m = n$, where r is the rank of the matrix (the dimension of the column space of A). Another way of saying this is that it happens only if A is square and nonsingular (it has an inverse).
- This system may have many solutions. This happens when b is in the column space of A and $r < n$ (the columns of A are linearly dependent, so that the null space of A is nontrivial).

Let us focus on the first case: b is not in the column space of A. In the last section, we argued that what we want is an approximate solution $\hat{x}$ such that $A\hat{x} = z$, where z is the vector in the column space of A that is "closest" to b: $b = z + w$ where $w^T v = 0$ for all $v \in \mathcal{C}(\mathcal{A})$. From Figure 6.1 we conclude that this means that w is in the left null space of A. So, $A^T w = 0$. But that means that

$$0 = A^T w = A^T(b - z) = A^T(b - A\hat{x})$$

which we can rewrite as

$$A^T A \hat{x} = A^T b. \tag{6.1}$$

Lemma 6.16 *If $A \in \mathbb{R}^{m \times n}$ has linearly independent columns, then $A^T A$ is nonsingular (equivalently, has an inverse, $A^T A\hat{x} = A^T b$ has a solution for all b, etc.).*

Proof: Proof by contradiction. Assume that $A \in \mathbb{R}^{m \times n}$ has linearly independent columns and $A^T A$ is singular. Then there exists $x \neq 0$ such that $A^T A x = 0$. Hence, there exists $y = Ax \neq 0$ such that $A^T y = 0$ (because A has linearly independent columns and $x \neq 0$). This means y is in the left null space of A. But y is also in the column space of A, since $Ax = y$. Thus, $y = 0$, since the intersection of the column space of A and the left null space of A only contains the zero vector. This contradicts the fact that A has linearly independent columns.

This means that if A has linearly independent columns, then the desired $\hat{x}$ in (6.1) is given by

$$\hat{x} = (A^T A)^{-1} A^T b$$

and the vector $z \in \mathcal{C}(\mathcal{A})$ closest to b is given by

$$z = A\hat{x} = A(A^T A)^{-1} A^T b.$$

This shows that *if* A has linearly independent columns, then $z = A(A^T A)^{-1} A^T b$ is the vector in the columns space closest to b. Think of this as the projection of z onto the column space of A.

Let us now formulate the above observations as a special case of a *linear least-squares* problem:

Theorem 6.17 *Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $x \in \mathbb{R}^n$ and assume that A has linearly independent columns. Then the solution that minimizes $\min_x \|b - Ax\|_2$ is given by $\hat{x} = (A^T A)^{-1} A^T b$.*

Definition 6.18 *Let $A \in \mathbb{R}^{m\times n}$.*

- *If A has linearly independent columns, then $(A^T A)^{-1} A^T$ is called the (left) pseudo inverse. Note that this means $m \geq n$ and $(A^T A)^{-1} A^T A = I$.*

- *If A has linearly independent rows, then $A^T (AA^T)^{-1}$ is called the right pseudo inverse. Note that this means $m \leq n$ and $AA^T (AA^T) = I$.*

Now, let us discuss the "least-squares" in the name of the section. Notice that we are trying to find $\hat{x}$ that minimizes $min_x\|b - Ax\|_2$. Now, if $\hat{x}$ minimizes $\min_x \|b - Ax\|_2$, it also minimizes the function $F(x) = \|b - Ax\|_2^2$ (since $\|b - Ax\|_2$ is always positive). But

$$F(x) \quad = \|b - Ax\|_2^2 = (b - Ax)^T (b - Ax) = b^T b - 2b^T Ax - x^T A^T Ax.$$

Recall how one would find the minimum of a function $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = \alpha^2 x^2 - 2\beta\alpha x + \beta^2$: One would take the derivative $f'(x) = 2\alpha^2 x - 2\beta\alpha$ and set it to zero: $2\alpha^2 x - 2\beta\alpha = 0$. The minimum is then attained by $\hat{x}$ that solves this equation: $\hat{x} = \beta/\alpha$. Now, here $F : \mathbb{R}^n \rightarrow \mathbb{R}$. Those who have taken multivariate calculus will know that this function is optimized when its gradient (essentially the derivative) equals zero: $0 - 2A^T b + 2A^T Ax = 0$, or, $A^T Ax = A^T b$. And thus we arrive at the same place as in (6.17)! We are looking for $\hat{x}$ that solves $A^T Ax = A^T b$.

Now, let $z = A\hat{x}$. Then $F(x) = \|b - z\|_2^2 = (b - z)^T (b - z) = \sum_i^{m-1} (\beta_i - \zeta_i)^2$ and hence the desired solution $\hat{x}$ minimizes the sum of the squares of the elements of $b - A\hat{x}$.

```
> x = [
   1
   2
   3
   4
];
> A = [
   1   1
   1   2
   1   3
   1   4
];
> y = [
  1.97
  6.97
  8.89
 10.01
];
> B = A' * A
B =

    4   10
   10   30

> yhat = A' * y
yhat =

   27.840
   82.620

> c = B \ yhat              % this solves B c = yhat
c =

   0.45000
   2.60400

> plot( x, y, 'o' )              %  plot the points ( x(1), y(1) ), ...
> axis( [ 0, 6, 0, 12 ] )        %  adjust the x and y ranges
> hold on                        %  plot the next graph over the last
> z = A * c;                     %  z = projection of y onto the column
                                 %  space of A
> plot( x, z, '-' )              %  plot the line through the points
                                 %  ( x(1), z(1) ), ( x(2), z(2) ), ...
> hold off
```

Figure 6.3: Solving the motivating problem in Section 6.4.

6.4 Motivating Example, Part II

Let us return to the example from Section 6.2. To find the best solution to $Ac = y$:

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} 1.97 \\ 6.97 \\ 8.89 \\ 10.01 \end{pmatrix}$$

we instead solve

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}^T \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}^T \begin{pmatrix} 1.97 \\ 6.97 \\ 8.89 \\ 10.01 \end{pmatrix}.$$

Let's use `octave` to do so, in Figure 6.3. By looking at the resulting plot, you will recognize that this is a nice fit to the data.

6.5 Computing an Orthonormal Basis

Definition 6.19 *Let $\{q_0, q_1, \ldots, q_{k-1}\}$ be a basis for subspace $\mathbf{V} \subset \mathbb{R}^m$. Then these vectors form an orthonormal basis if each vector is of unit length and $q_i^T q_j = 0$ if $i \neq j$.*

Exercise 6.20 *Let $\{q_0, q_1, \ldots, q_{k-1}\}$ be an orthonormal basis for subspace $\mathbf{V} \subset \mathbb{R}^m$. Show that*

1. $q_i^T q_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

2. *If $Q = \left(q_0 \mid \cdots \mid q_{k-1} \right)$ then $Q^T Q = I$, the identity.*

Given a basis for a subspace we show how to compute an orthonormal basis for that same space.

Let us start with two vectors $a_0, a_1 \in \mathbb{R}^m$ and matrix $A = \left(a_0 \mid a_1 \right)$ and let us assume these vectors are linearly independent. Then, by definition, they form a basis for the column space of A. Some observations:

- We can easily come up with a vector q_0 that is of length one and points in the same direction as a_0: $q_0 = a_0/\rho_{00}$ where $\rho_{00} = \|a_0\|_2$.

  Check: $\|q_0\|_2 = \|a_0/\|a_0\|_2\|_2 = \|a_0\|_2/\|a_0\|_2 = 1$.

- Now, we can write $a_1 = \rho_{01} q_0 + v$ where v is orthogonal to q_0, for some scalar ρ_{00}. Then
  - multiplying on the left by q_0^T we find that $q_0^T a_1 = q_0^T(\rho_{01} q_0 + v) = q_0^T \rho_{00} q_0 + q_0^T v = \rho_{01} q_0^T q_0 + q_0^T v = \rho_{01}$ because $q_0^T q_0 = 1$ and $q_0^T v = 0$.

- Once we have computed ρ_{01}, we can compute $v = a_1 - \rho_{01} q_0$.
- What we have now are two vectors, q_0 and v, the first of which has length 1 and the second of which is orthogonal to the first. If we now scale the second one to have length 1, then we have found two orthonormal vectors (vectors that are of length 1 that are mutually orthogonal): Compute $\rho_{11} = \|v\|_2$ and let $q_1 = v/\rho_{11}$.
- Notice that $a_1 = \rho_{01} q_0 + v = \rho_{01} q_0 + \rho_{11} q_1$ so that $q_1 = (a_1 - \rho_{01} q_0)/\rho_{11} = (a_1 - \frac{\rho_{01}}{\rho_{00}} a_0)/\rho_{11}$.

Now, the important observation is that the column space of $A = \left(\begin{array}{c|c} a_0 & a_1 \end{array} \right)$ is equal to the column space of the just computed $Q = \left(\begin{array}{c|c} q_0 & q_1 \end{array} \right)$.

$\Rightarrow$: Let z be in the column space of A. Then there exist $\gamma_0, \gamma_1 \in \mathbb{R}$ such that

$$z = \gamma_0 a_0 + \gamma_1 a_1 = \gamma_0(\rho_{00} q_0) + \gamma_1(\rho_{01} q_0 + \rho_{11} q_1) = \underbrace{(\gamma_0 \rho_{00} + \gamma_1 \rho_{01})}_{\beta_0} q_0 + \underbrace{(\gamma_1 \rho_{11})}_{\beta_1} q_1 = \beta_0 q_0 + \beta_1 q_1.$$

In other words, there exist $\beta_0, \beta_1 \in \mathbb{R}$ such that $z = \beta_0 q_0 + \beta_1 q_1$. Hence z is in the column space of Q.

$\Leftarrow$: Let z be in the column space of Q. Then there exist $\beta_0, \beta_1 \in \mathbb{R}$ such that

$$z = \beta_0 q_0 + \beta_1 q_1 = \beta_0(a_0/\rho_{00}) + \beta_1(a_1 - \frac{\rho_{01}}{\rho_{00}} a_0)/\rho_{11} = \text{etc.} = \gamma_0 a_0 + \gamma_1 a_1.$$

In other words, there exist $\gamma_0, \gamma_1 \in \mathbb{R}$ such that $z = \gamma_0 a_0 + \gamma_1 a_1$. Hence z is in the column space of A.

6.6 Motivating Example, Part III

Let us again return to the example from Section 6.2. We wanted to find z that was the projection of b onto (the closest point in) the column space of A. What we just found is that this is the same as finding z that is the projection of y onto the column space of Q, as computed in Section 6.5.

So, this brings up the question: If we have a matrix Q with orthonormal columns, how do we compute the projection onto the column space of Q? We will work with a matrix Q with two columns for now.

We want to find an approximate solution to $Ac = y$. Now, $y = z + w$ where z is in the column space of Q and w is orthogonal to the column space of Q. Thus

$$y = z + w = \beta_0 q_0 + \beta_1 q_1 + w$$

where $q_0^T w = 0$ and $q_1^T w = 0$. Now

- $q_0^T y = q_0(\beta_0 q_0 + \beta_1 q_1 + w) = \beta_0 q_0^T q_0 + \beta_1 q_0^T q_1 + q_0^T w = \beta_0$ since $q_0^T q_0 = 1$ and $q_0^T q_1 = q_0^T w = 0$.
- $q_1^T y = q_1(\beta_0 q_0 + \beta_1 q_1 + w) = \beta_0 q_1^T q_0 + \beta_1 q_1^T q_1 + q_1^T w = \beta_1$ since $q_1^T q_1 = 1$ and $q_1^T q_0 = q_1^T w = 0$.

So, we have a way of computing z:

$$z = \beta_0 q_0 + \beta_1 q_1 = \underbrace{q_0^T y \; q_0}_{\substack{\text{projection}\\ \text{of } y\\ \text{onto } q_0}} + \underbrace{q_1^T y \; q_1.}_{\substack{\text{projection}\\ \text{of } y\\ \text{onto } q_1}}$$

Thus, we have a systematic way of solving our problem. First we must compute Q:

- The matrix is given by

$$A = \begin{pmatrix} 1 & \chi_0 \\ 1 & \chi_1 \\ 1 & \chi_2 \\ 1 & \chi_3 \end{pmatrix} \quad \text{so that} \quad a_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad a_1 = \begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \\ \chi_3 \end{pmatrix}.$$

- $\rho_{00} = \|a_0\|_2 = \sqrt{a_0^T a_0} = \sqrt{\sum_{i=0}^{3} 1} = \sqrt{4} = 2$.
- $q_0 = a_0/\rho_{00} = (0.5, 0.5, 0.5, 0.5)^T$.
- $\rho_{01} = q_0^T a_1 = (0.5, 0.5, 0.5, 0.5)(1, 2, 3, 4)^T = 5$
- $v = a_1 - \rho_{01} q_0 = (1, 2, 3, 4)^T - 5(0.5, 0.5, 0.5, 0.5, 0.5)^T = (-1.5, -0.5, 0.5, 1.5)^T$.
- $\rho_{11} = \|v\|_2 = \sqrt{(-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2} = \sqrt{2\left(\frac{3}{2}\right)^2 + 2\left(\frac{1}{2}\right)^2} = \sqrt{5}$.
- $q_1 = v/\rho_{11} = \sqrt{5}/5(-1.5, -0.5, 0.5, 1.5)^T$.

Next, we compute $z = q_0^T y \; q_0 + q_1^T y \; q_1$. For this, we again use `octave`, in Figure 6.4.

Now, it turns out there is an easier way to proceed than to compute z and then solve, from scratch, $Ac = z$.

Recall that $a_0 = \rho_{00} q_0$ and $a_1 = \rho_{01} q_0 + v = \rho_{01} q_0 + \rho_{11} q_1$. Now, we can write this as

$$\left(\begin{array}{c|c} a_0 & a_1 \end{array} \right) = \left(\begin{array}{c|c} q_0 & q_1 \end{array} \right) \left(\begin{array}{c|c} \rho_{00} & \rho_{01} \\ \hline 0 & \rho_{11} \end{array} \right) = \left(\begin{array}{c|c} \rho_{00} q_0 & \rho_{01} q_0 + \rho_{11} q_1 \end{array} \right).$$

Also, we know that

$$z = q_0^T y q_0 + q_1^T y q_1 = \left(\begin{array}{c|c} q_0 & q_1 \end{array} \right) \left(\begin{array}{c} q_0^T y \\ \hline q_1^T y \end{array} \right).$$

```
> q0 = 0.5 * ones( 4, 1 )
q0 =

   0.50000
   0.50000
   0.50000
   0.50000

> q1 = sqrt(5)/5 * [ -1.5 -0.5 0.5 1.5 ]'
q1 =

  -0.67082
  -0.22361
   0.22361
   0.67082
> y = [ 1.97 6.97 8.89 10.01 ]'
y =

    1.9700
    6.9700
    8.8900
   10.0100

> z = q0' * y * q0 + q1' * y * q1
z =

    3.0540
    5.6580
    8.2620
   10.8660
> c = A \ z        % Solve A c = z
```

Figure 6.4: Solution of using the Q computed in Section 6.6.

Thus, $Ac = z$ can be rewritten as $QRc = Q\hat{z}$ where $\hat{z} = \begin{pmatrix} q_0^T y \\ q_1^T y \end{pmatrix} = Q^T z$. Now, multiply both side from the left by Q^T: $Q^T Q R c = Q^T Q \hat{z}$ and recognize that $Q^T Q = I$, the identity, since Q has orthonormal columns! We are left with $Rc = \hat{z}$. In other words

$$\begin{pmatrix} \rho_{00} & \rho_{01} \\ 0 & \rho_{11} \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} q_0^T y \\ q_1^T y \end{pmatrix}.$$

For our specific example this means we need to solve the upper triangular system

$$\begin{pmatrix} 2 & 5 \\ 0 & \sqrt{5} \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} 13.92 \\ 5.8227 \end{pmatrix}.$$

We can then solve the system with octave via the steps

```
> R = [
> 2 5
> 0 sqrt(5)
> ]
R =

   2.00000   5.00000
   0.00000   2.23607

> zhat = [
> q0' * y
> q1' * y
> ]
zhat =

   13.9200
    5.8227

> c = R \ zhat
c =

   0.45000
   2.60400
```

In Figure 6.5 we plot the line $y = \gamma_0 + \gamma_1 x$ with the resulting coefficients.

6.7 What does this all mean?

Let's summarize a few observations:

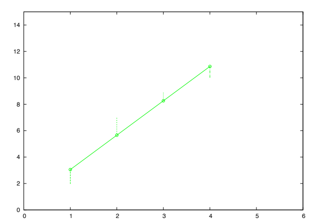


Figure 6.5: Least-squares fit to the data.

Component in the direction of a unit vector Given $q, y \in \mathbb{R}^m$ where q has unit length ($\|q\|_2 = 1$), the component of y in the direction of q is given by $q^T y q$. In other words, $y = q^T y q + z$ where z is orthogonal to q ($q^T z = z^T q = 0$).

Orthonormal basis Let $\{q_0, \ldots, q_{k-1}\} \in \mathbb{R}^m$ form an orthonormal basis for $\mathbf{V}$. Let $Q = \left(\begin{array}{c|c|c} q_0 & \cdots & q_{k-1} \end{array} \right)$. Then

- $Q^T Q = I$. Why? The (i, j) entry of $Q^T Q = q_i^T q_j = \left\{ \begin{array}{ll} 0 & \text{if } i \neq j \\ 1 & \text{otherwise} \end{array} \right.$.
- Given $y \in \mathbb{R}^m$, the vector z in $\mathbf{V}$ closest to y is given by $\hat{y} = q_0^T y q_0 + \cdots + q_{k-1}^T y q_{k-1} = Q Q^T y$. The matrix $Q Q^T$ is the *projection of y onto* $\mathbf{V}$.
- Given $y \in \mathbb{R}^m$, the component of this vector that is orthogonal to $\mathbf{V}$ is given by $z = y - \hat{y} = y - (q_0^T y q_0 + \cdots + q_{k-1}^T y q_{k-1}) = y - Q Q^T y = (I - Q Q^T) y$. The matrix $I - Q Q^T$ is the *projection of y onto the space orthonormal to* $\mathbf{V}$, $\mathbf{V}^\perp$.

Computing an orthonormal basis Given linearly independent vectors $\{a_0, \ldots, a_{k-1}\} \subset \mathbb{R}^m$ that form a basis for subspace $\mathbf{V}$, the following procedure will compute an orthonormal basis for $\mathbf{V}$, $\{q_0, \ldots, q_{k-1}\}$:

- $\rho_{00} = \|a_0\|$ (the length of a_0).

- $q_0 = a_0/\rho_{00}$ (the vector in the direction of a_0 of unit length).
- $\rho_{01} = q_0^T a_1$.
- $v_1 = a_1 - \rho_{01} q_0 = a_1 - q_0^T a_1 q_0$ (= the component of a_1 orthogonal to q_0).
- $\rho_{11} = \|v_1\|_2$ (= the length of the component of a_1 orthogonal to q_0).
- $q_1 = v_1/\rho_{11}$ (= the unit vector in the direction of the component of a_1 orthogonal to q_0).
- $\rho_{02} = q_0^T a_2$.
- $\rho_{12} = q_1^T a_2$.
- $v_2 = a_2 - \rho_{02} q_0 - \rho_{12} q_1 = a_2 - (q_0^T a_2 q_0 + q_1^T a_2 q_1)$ (= the component of a_2 orthogonal to q_0 and q_1).
- $\rho_{22} = \|v_2\|_2$ (= the length of the component of a_2 orthogonal to q_0 and q_1).
- $q_2 = v_2/\rho_{22}$ (= the unit vector in the direction of the component of a_2 orthogonal to q_0 and q_1).
- etc.

This procedure is known as *Classical Gram-Schmidt*. It is stated as an algorithm in Fig. 6.6. It can be verified that $A = QR$, where

$$A = \left(\begin{array}{c|c|c} a_0 & \cdots & a_{n-1} \end{array} \right), \quad Q = \left(\begin{array}{c|c|c} q_0 & \cdots & q_{n-1} \end{array} \right), \quad \text{and} \quad R = \left(\begin{array}{c|c|c|c} \rho_{00} & \rho_{01} & \cdots & \rho_{0(n-1)} \\ \hline 0 & \rho_{11} & \cdots & \rho_{1(n-1)} \\ \hline \vdots & \ddots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & \rho_{(n-1)(n-1)} \end{array} \right)$$

so that

$$\left(\begin{array}{c|c|c} a_0 & \cdots & a_{n-1} \end{array} \right) = \left(\begin{array}{c|c|c} q_0 & \cdots & q_{n-1} \end{array} \right) \left(\begin{array}{c|c|c|c} \rho_{00} & \rho_{01} & \cdots & \rho_{0(n-1)} \\ \hline 0 & \rho_{11} & \cdots & \rho_{1(n-1)} \\ \hline \vdots & \ddots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & \rho_{(n-1)(n-1)} \end{array} \right).$$

for $j = 0, \ldots, k-1$
  Compute component of a_j orthogonal to $\{q_0, \ldots, q_{j-1}\}$:
    $v_j = a_j$
    for $i = 0, \ldots, j-1$
      Let $\rho_{ij} = q_i^T a_j$
      $v_j = v_j - \rho_{ij} q_i$
    endfor
  Normalize v_j to unit length:
    $\rho_{jj} = \|v_j\|_2$
    $q_j = v_j / \rho_{jj}$
endfor

Figure 6.6: Gram-Schmidt procedure for computing an orthonormal basis for the space spanned by linearly independent vectors $\{a_0, \ldots, a_{k-1}\}$, which are now taken to be the columns of matrix A.

for $j = 0, \ldots, k-1$
  Compute component of a_j orthogonal to $\{q_0, \ldots, q_{j-1}\}$:

$$\left(\begin{array}{c} \rho_{0j} \\ \hline \vdots \\ \hline \rho_{(j-1)j} \end{array} \right) = \left(\begin{array}{c|c|c} q_0 & \cdots & q_{j-1} \end{array} \right)^T a_j = \left(\begin{array}{c} q_0^T \\ \hline \vdots \\ \hline q_{j-1}^T \end{array} \right) a_j = \left(\begin{array}{c} q_0^T a_j \\ \hline \vdots \\ \hline q_{j-1}^T a_j \end{array} \right)$$

$$v_j = a_j - \left(\begin{array}{c|c|c} q_0 & \cdots & q_{j-1} \end{array} \right) \left(\begin{array}{c} \rho_{0j} \\ \vdots \\ \rho_{(j-1)j} \end{array} \right) = a_j - (\rho_{0j} q_0 + \cdots + \rho_{(j-1)j} q_j)$$

  Normalize v_j to unit length:
    $\rho_{jj} = \|v_j\|_2$
    $q_j = v_j / \rho_{jj}$
endfor

Figure 6.7: Gram-Schmidt procedure for computing an orthonormal basis for the space spanned by linearly independent vectors $\{a_0, \ldots, a_{k-1}\}$.

Two ways for solving linear least-squares problems We motivated two different ways for computing the best solution, $\hat{x}$, to the linear least-squares problem $\min_x \|Ax = b\|_2$ where $A \in \mathbb{R}^{m\times n}$ has linearly independent columns:

- $\hat{x} = (A^TA)^{-1}A^Tb$.
- Compute Q with orthonormal columns and upper triangular R such that $A = QR$. Solve $R\hat{x} = Q^Tb$. In other words, $\hat{x} = R^{-1}Q^Tb$.

These, fortunately, yield the same result:

$$\begin{aligned}
\hat{x} &= (A^TA)^{-1}A^Tb \\
&= ((QR)^T(QR))^{-1}(QR)^Tb \\
&= (R^TQ^TQR)^{-1}R^TQ^Tb \\
&= (R^TR)^{-1}R^TQ^Tb \\
&= R^{-1}(R^T)^{-1}R^TQ^Tb \\
&= R^{-1}Q^Tb.
\end{aligned}$$

6.8 Exercises

(Some of these exercises have been inspired by similar exercises in Strang's book.')

1. Consider $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.
   (a) Determine if b is in the column space of A.
   (b) Compute the approximate solution, in the least squares sense, of $Ax = b$. (This means: solve $A^T Ax = A^T b$.)
   (c) What is the project of b onto the column space of A?
   (d) Compute the QR factorization of A and use it to solve $Ax = b$. (In other words, compute QR using the Gram-Schmidt process and then solve $Rx = Q^T b$.)
2. Consider $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $b = \begin{pmatrix} 4 \\ 5 \\ 9 \end{pmatrix}$.
   (a) Determine if b is in the column space of A.
   (b) Compute the approximate solution, in the least squares sense, of $Ax = b$. (This means: solve $A^T Ax = A^T b$.)
   (c) What is the project of b onto the column space of A?
   (d) Compute the QR factorization of A and use it to solve $Ax = b$. (In other words, compute QR using the Gram-Schmidt process and then solve $Rx = Q^T b$.)
3. Consider $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -2 & 4 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix}$.
   (a) Find the projection of b onto the column space of A.
   (b) Split b into $u+v$ where u is in the column space and v is perpendicular (orthogonal) to that space.
   (c) Which of the four subspaces ($C(A)$, $R(A)$, $\mathcal{N}(A)$, $\mathcal{N}(A^T)$) contains q?
4. What 2×2 matrix A projects the x-y plane onto the line $x + y = 0$?
5. Find the best straight line fit to the following data:

$$\begin{aligned} y &= 2 \text{ at } t = -1 \\ y &= -3 \text{ at } t = 1 \\ y &= 0 \text{ at } t = 0 \\ y &= -5 \text{ at } t = 2 \end{aligned}$$