1 Last time: least-squares problems

Definition. If A is an $m \times n$ matrix and $b \in \mathbb{R}^m$, then a *least-squares solution* to the linear system $Ax = b$ is a vector $\widehat{x} \in \mathbb{R}^n$ such that $\|b - A\widehat{x}\| \leq \|b - Ax\|$ for all $x \in \mathbb{R}^n$.

If $Ax = b$ is a consistent linear system then every least-squares solution will be an exact solution. For any linear system $Ax = b$, there is **always** at least one least-squares solution.

Theorem. Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. A vector $\widehat{x} \in \mathbb{R}^n$ is a least-square solutions to $Ax = b$ if and only if $A^T A\widehat{x} = A^T b$. The linear system $A^T Ax = A^T b$ is always consistent.

Theorem. Let A be an $m \times n$ matrix. The following are then equivalent:

(a) $Ax = b$ has a unique least-squares solution for each $b \in \mathbb{R}^m$.

(b) The columns of A are linearly independent.

(c) $A^T A$ is invertible.

When these properties hold, the unique least-squares solution to $Ax = b$ is the vector

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

(which is the unique exact solution to $A^T Ax = A^T b$).

In applications, the matrix A usually has many more rows than columns, so m is much larger than n. In this case, the $n \times n$ matrix $A^T A$ is much smaller than A, and we can find a least-squares solution by row reducing $\begin{bmatrix} A^T A & A^T b \end{bmatrix}$ to echelon form to find the exact solutions to $A^T Ax = A^T b$ in the usual way.

Example (Lines of best fit)**.** Suppose we have n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that appear to be close to a line. We want to find parameters $\beta_0, \beta_1 \in \mathbb{R}$ such that $y = \beta_0 + \beta_1 x$ describes the *line of best fit* for this data. If our points are collinear, then for some $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \in \mathbb{R}^2$ we would have exactly

$$y_i = \beta_0 + \beta_1 x_i \quad \text{for } i = 1, 2, \ldots, n,$$

meaning that there is an exact solution to the linear system $X\beta = Y$ where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \text{and} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

If the given points are not already on a line, then no exact solution to $X\beta = Y$ exists, and we should instead try to find a least-squares solution to this linear system.

To be concrete, suppose we have four points $(2, 1), (5, 2), (7, 3)$, and $(8, 3)$ so that

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \qquad \text{and} \qquad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

The least-squares solutions to $X\beta = Y$ are the exact solutions to $X^T X\beta = X^T Y$. We have

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

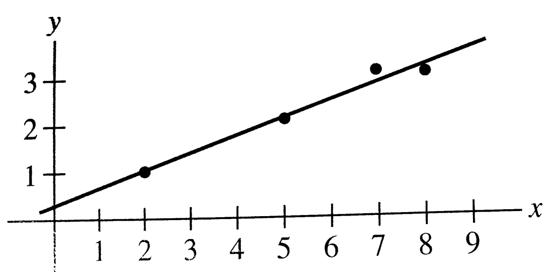
and

$$X^TY = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

The matrix X^TX is invertible. (Why?) It follows that a least-squares solution $\beta \in \mathbb{R}^2$ is provided by

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}.$$

Thus our line of best fit for the data is $y = \frac{2}{7} + \frac{5}{14}x$:



2 Symmetric matrices

A matrix A is *symmetric* if $A^T = A$.

This happens if and only if A is square and $A_{ij} = A_{ji}$ for all i, j.

Example. $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$ and $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}$ and $\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$ are symmetric matrices.

$\begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 6 & -6 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix}$ are not symmetric.

Proposition. If A is a symmetric matrix and k is a positive integer then A^k is also symmetric.

Proof. If $A = A^T$ then $(A^k)^T = (AA\cdots A)^T = A^T \cdots A^T A^T = (A^T)^k = A^k$. □

Proposition. If A is an invertible symmetric matrix then A^{-1} is also symmetric.

Proof. This is because $(A^{-1})^T = (A^T)^{-1}$. □

Recall how we can diagonalise a matrix.

Example. Let $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$.

Then $\det(A - xI) = (8-x)(6-x)(3-x)$ so the eigenvalues of A are 8, 6, and 3. By constructing bases for the null spaces of $A - 8I$, $A - 5I$, and $A - 3I$, we find that the following are eigenvectors of A:

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \text{ with eigenvalue 8.}$$

$$v_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \text{ with eigenvalue 6.}$$

$$v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ with eigenvalue 3.}$$

These eigenvectors are actually an orthogonal basis for $\mathbb{R}^3$.

Converting these vectors to unit vectors gives an orthonormal basis of eigenvectors:

$$u_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \qquad u_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \qquad u_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

We then have $A = PDP^{-1}$ where

$$P = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \qquad \text{and} \qquad D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

(Do you remember why this holds? It is enough to check that $PDP^{-1}v = Av$ for $v \in \{u_1, u_2, u_3\}$.)

Since the columns of P are orthonormal, we actually have $P^T = P^{-1}$ so $A = PDP^T$.

The special properties in this example will turn out to hold for all symmetric matrices.

Theorem. Suppose A is a symmetric matrix. Then any two eigenvectors from different eigenspaces of A are orthogonal. In other words, if A is $n \times n$ and $u, v \in \mathbb{R}^n$ are such that $Au = au$ and $Av = bv$ for numbers $a, b \in \mathbb{R}$ with $a \neq b$, then $u \bullet v = 0$.

Proof. Let u and v be eigenvectors of A with eigenvalues a and b, where $a \neq b$.

Then $au \bullet v = Au \bullet v = (Au)^T v = u^T A^T v = u^T Av = u \bullet Av = u \bullet bv$.

But $au \bullet v = a(u \bullet v)$ and $u \bullet bv = b(u \bullet v)$, so this means $a(u \bullet v) = b(u \bullet v)$ and therefore

$$(a - b)(u \bullet v) = 0.$$

Since $a - b \neq 0$, it follows that $u \bullet v = 0$. □

Definition. A matrix P is *orthogonal* if P is invertible and $P^{-1} = P^T$.

Definition. A matrix A is *orthogonally diagonalisable* if there is an orthogonal matrix P and a diagonal matrix D such that $A = PDP^{-1} = PDP^T$.

When A is orthogonally diagonalisable and $A = PDP^{-1} = PDP^T$, the diagonal entries of D are the eigenvalues of A, and the columns of P are the corresponding eigenvectors; moreover, these eigenvectors form an orthonormal basis of $\mathbb{R}^n$.

In fact, it follows by the arguments in our earlier lectures about diagonalisable matrices that an $n \times n$ matrix A is orthogonally diagonalisable if and only if there is an orthogonal basis for $\mathbb{R}^n$ consisting of eigenvectors for A.

Surprisingly, there is a much more direct characterisation of orthogonally diagonalisable matrices:

Theorem. A square matrix is orthogonally diagonalisable if and only if it is symmetric.

We prove this after a sequence of lemmas.

Lemma. If A is orthogonally diagonalisable then A is symmetric.

Proof. Note that if X, Y, Z are $n \times n$ matrices the $(XYZ)^T = Z^T(XY)^T = Z^T Y^T X^T$.

Suppose $A = PDP^T$ where D is diagonal. Then $D = D^T$ and $(P^T)^T = P$, so

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A.$$

□

Lemma. All (complex) eigenvalues of an $n \times n$ symmetric matrix A with real entries belong to $\mathbb{R}$.

Proof. Suppose A is a symmetric $n \times n$ matrix with real entries, so that $A = A^T = \overline{A}$.

Let $v \in \mathbb{C}^n$. Then $\overline{v}^T A v$ is some complex number.

For example, if $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $v = \begin{bmatrix} 1+i \\ 1-i \end{bmatrix}$ then

$$\begin{aligned} \overline{v}^T A v &= \begin{bmatrix} 1-i & 1+i \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} \\ &= \begin{bmatrix} 3+i & 3-i \end{bmatrix} \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = (3+i)(1+i) + (3-i)(1-i) = 4. \end{aligned}$$

In fact, the number $\overline{v}^T A v$ belongs to $\mathbb{R}$ since

$$\overline{\overline{v}^T A v} = v^T A \overline{v} = \left(\overline{v}^T A v\right)^T = \overline{v}^T A v.$$

(The last equality holds since both sides are 1×1 matrices, i.e., scalars.)

Now suppose $v \in \mathbb{C}^n$ is an eigenvector for A with eigenvalue $\lambda \in \mathbb{C}$. Then $\overline{v}^T A v = \overline{v}^T(\lambda v) = \lambda(\overline{v}^T v) \in \mathbb{R}$. The complex number $\overline{v}^T v$ always belongs to $\mathbb{R}$ (why?) so it must also hold that $\lambda \in \mathbb{R}$. □

Lemma. An $n \times n$ matrix A with all real eigenvalues can be written as $A = URU^T$ where U is an $n \times n$ orthogonal matrix (i.e., has orthonormal columns) and R is an $n \times n$ upper-triangular matrix.

One calls $A = URU^T$ with U and R of this form a *Schur factorisation* of A.

Proof. Suppose A is an $n \times n$ matrix with all real eigenvalues.

Let $u_1 \in \mathbb{R}^n$ be a unit eigenvector for A with eigenvalue $\lambda \in \mathbb{R}$.

Let $u_2, \dots, u_n \in \mathbb{R}^n$ be any vectors such that $u_1, u_2, \dots, u_n$ is an orthonormal basis for $\mathbb{R}^n$. (One way to construct these vectors: let $u_1 = x_1, x_2, \dots, x_n$ be any basis, apply Gram-Schmidt to get $u_1 = v_1, v_2, \dots, v_n$, then convert each v_i to a unit vector.)

Define $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ so that $U^T = U^{-1}$.

By considering the product $U^T A U e_i$ for $i = 1, 2, \dots, n$, one finds that $U^T A U$ has the form

$$U^T A U = \begin{bmatrix} \lambda & * \\ 0 & B \end{bmatrix}$$

for some $(n-1) \times (n-1)$ matrix B. Here, $*$ stands for $n-1$ arbitrary entries.

The matrix $U^TAU = U^{-1}AU$ has the same characteristic polynomial as A.

This polynomial is $(\lambda - x)\det(B - xI)$, i.e., $\lambda - x$ times the characteristic polynomial of B.

Since the characteristic polynomial of A has all real roots, the same must be true of the characteristic polynomial of B. Thus B must also have all real eigenvalues.

By repeating the argument above, we deduce that there is an eigenvalue $\mu \in \mathbb{R}$ for B, an $(n-1)\times(n-1)$ orthogonal matrix V, and an $(n-2)\times(n-2)$ matrix C with all real eigenvalues such that

$$V^TBV = \begin{bmatrix} \mu & * \\ 0 & C \end{bmatrix}.$$

The matrix $\begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$ is also orthogonal, and the product of orthogonal matrices is orthogonal. (Why?)

It follows for the orthogonal matrix $W = U\begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$ that

$$W^TAW = \begin{bmatrix} \lambda & * & * \\ 0 & \mu & * \\ 0 & 0 & C \end{bmatrix}.$$

By continuing in this way, we will eventually construct an orthogonal matrix X and an upper-triangular matrix R such that $X^TAX = R$, in which case $A = XX^TAXX^T = XRX^T$. □

Now we can prove the theorem.

Proof of theorem. The first lemma shows that if A is orthogonally diagonalisable then A is symmetric.

Suppose conversely that A is symmetric.

Then A has all real eigenvalues, so there exists a Schur factorisation $A = URU^T$.

We then have $A^T = (URU^T)^T = UR^TU^T$ but also $A^T = A = URU^T$.

Since $U^T = U^{-1}$, it follows that $R = R^T$.

Since R is upper-triangular, this can only hold if R is diagonal.

But if R is diagonal then $A = URU^T$ is evidently orthogonally diagonalisable. □

To orthogonally diagonalise an $n \times n$ symmetric matrix A, we just need to find an orthogonal basis of eigenvectors $v_1, v_2, \dots, v_n$ for $\mathbb{R}^n$. Then $A = UDU^T$ with $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ where $u_i = \frac{1}{\|v_i\|}v_i$ and D is the diagonal matrix of the corresponding eigenvalues.

If all eigenspaces of A are 1-dimensional, then any basis of eigenvectors will be orthogonal. If A has an eigenspace of dimension greater than one, then after finding a basis for this eigenspace, it is necessary to apply the Gram-Schmidt process to covert this basis to one which is orthogonal.

Corollary. If $A = UDU^T$ where $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ has orthonormal columns and

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

is diagonal, then

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \dots + \lambda_n u_n u_n^T. \qquad (*)$$

Each product $u_i u_i^T$ is an $n \times n$ matrix of rank 1. One calls (*) a *spectral decomposition* of A.

Example. Let $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$. A spectral decomposition of A is given by

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

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

$$= \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix}.$$