LINEAR ALGEBRA. LECTURE 12

Symmetric matrices and positive definiteness

Symmetric matrices are good – their eigenvalues are real and each has a complete set of orthonormal eigenvectors. Positive definite matrices are even better.

Eigenvalues of A^T

The eigenvalues of A and the eigenvalues of A^T are the same:

$$(A - \lambda I)^T = A^T - \lambda I,$$

so property 10 of determinants tells us that $\det(A - \lambda I) = \det(A^T - \lambda I)$. If λ is an eigenvalue of A then $\det(A^T - \lambda I) = 0$ and λ is also an eigenvalue of A^T.

Symmetric matrices

A *symmetric matrix* is one for which $A = A^T$. If a matrix has some special property (e.g. it's a Markov matrix), its eigenvalues and eigenvectors are likely to have special properties as well. For a symmetric matrix with real number entries, the eigenvalues are real numbers and it's possible to choose a complete set of eigenvectors that are perpendicular (or even orthonormal).

If A has n independent eigenvectors we can write $A = S\Lambda S^{-1}$. If A is symmetric we can write $A = Q\Lambda Q^{-1} = Q\Lambda Q^T$, where Q is an orthogonal matrix. Mathematicians call this the *spectral theorem* and think of the eigenvalues as the "spectrum" of the matrix. In mechanics it's called the *principal axis theorem*.

In addition, any matrix of the form $Q\Lambda Q^T$ will be symmetric.

Real eigenvalues

Why are the eigenvalues of a symmetric matrix real? Suppose A is symmetric and $A\mathbf{x} = \lambda\mathbf{x}$. Then we can conjugate to get $\overline{A}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. If the entries of A are real, this becomes $A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. (This proves that complex eigenvalues of real valued matrices come in conjugate pairs.)

Now transpose to get $\overline{\mathbf{x}}^T A^T = \overline{\mathbf{x}}^T\overline{\lambda}$. Because A is symmetric we now have $\overline{\mathbf{x}}^T A = \overline{\mathbf{x}}^T\overline{\lambda}$. Multiplying both sides of this equation on the right by $\mathbf{x}$ gives:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x}.$$

On the other hand, we can multiply $A\mathbf{x} = \lambda\mathbf{x}$ on the left by $\overline{\mathbf{x}}^T$ to get:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}.$$

Comparing the two equations we see that $\overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}$ and, unless $\overline{\mathbf{x}}^T\mathbf{x}$ is zero, we can conclude $\lambda = \overline{\lambda}$ is real.

How do we know $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$?

$$\overline{\mathbf{x}}^T\mathbf{x} = \begin{bmatrix} \overline{x}_1 & \overline{x}_2 & \cdots & \overline{x}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2.$$

If $\mathbf{x} \neq \mathbf{0}$ then $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$.

With complex vectors, as with complex numbers, multiplying by the conjugate is often helpful.

Symmetric matrices with real entries have $A = A^T$, real eigenvalues, and perpendicular eigenvectors. If A has complex entries, then it will have real eigenvalues and perpendicular eigenvectors if and only if $A = \overline{A}^T$. (The proof of this follows the same pattern.)

Projection onto eigenvectors

If $A = A^T$, we can write:

$$\begin{aligned} A &= Q\Lambda Q^T \\ &= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\ &= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T \end{aligned}$$

The matrix $\mathbf{q}_k \mathbf{q}_k^T$ is the projection matrix onto $\mathbf{q}_k$, so every symmetric matrix is a combination of perpendicular projection matrices.

Information about eigenvalues

If we know that eigenvalues are real, we can ask whether they are positive or negative. (Remember that the signs of the eigenvalues are important in solving systems of differential equations.)

For very large matrices A, it's impractical to compute eigenvalues by solving $|A - \lambda I| = 0$. However, it's not hard to compute the pivots, and the signs of the pivots of a symmetric matrix are the same as the signs of the eigenvalues:

$$\text{number of positive pivots} = \text{number of positive eigenvalues.}$$

Because the eigenvalues of $A + bI$ are just b more than the eigenvalues of A, we can use this fact to find which eigenvalues of a symmetric matrix are greater or less than any real number b. This tells us a lot about the eigenvalues of A even if we can't compute them directly.

Positive definite matrices and minima

Studying positive definite matrices brings the whole course together; we use pivots, determinants, eigenvalues and stability. The new quantity here is $\mathbf{x}^T A\mathbf{x}$; watch for it.

This lecture covers how to tell if a matrix is positive definite, what it means for it to be positive definite, and some geometry.

Positive definite matrices

Here is the main theorem on positive definiteness, and a reasonably detailed proof:

Each of the following tests is a necessary and sufficient condition for the real symmetric matrix A to be ***positive definite***:

(I) $x^{\mathrm{T}}kx > 0$ for all nonzero real vectors x.

(II) All the eigenvalues of A satisfy $\lambda_i > 0$.

(III) All the upper left submatrices A_k have positive determinants.

(IV) All the pivots (without row exchanges) satisfy $d_k > 0$.

Proof. Condition I defines a positive definite matrix. Our first step shows that each eigenvalue will be positive:

$$\text{If} \quad Ax = \lambda x, \quad \text{then} \quad x^{\mathrm{T}}Ax = x^{\mathrm{T}}\lambda x = \lambda \|x\|^2.$$

A positive definite matrix has positive eigenvalues, since $x^{\mathrm{T}}Ax > 0$.

Now we go in the other direction. If all $\lambda_i > 0$, we have to prove $x^{\mathrm{T}}Ax > 0$ for every vector x (not just the eigenvectors). Since symmetric matrices have a full set of orthonormal eigenvectors, any x is a combination $c_1x_1 + \cdots + c_nx_n$. Then

$$Ax = c_1Ax_1 + \cdots + c_nAx_n = c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n.$$

Because of the orthogonality $x_i^{\mathrm{T}}x_i = 0$, and the normalization $x_i^{\mathrm{T}}x_i = 1$,

$$\begin{aligned} x^{\mathrm{T}}Ax &= \left(c_1x_1^{\mathrm{T}} + \cdots + c_nx_n^{\mathrm{T}}\right)\left(c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n\right) \\ &= c_1^2\lambda_1 + \cdots + c_n^2\lambda_n. \end{aligned}$$

If every $\lambda_i > 0$, then equation (2) shows that $x^{\mathrm{T}}Ax > 0$. Thus condition II implies condition I.

If condition I holds, so does condition III: The determinant of A is the product of the eigenvalues. And if condition I holds, we already know that these eigenvalues are positive. But we also have to deal with every upper left submatrix A_k. The trick is to look at all nonzero vectors whose last $n-k$ components are zero:

$$x^{\mathrm{T}}Ax = \begin{bmatrix} x_k^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^{\mathrm{T}}A_kx_k > 0.$$

Thus A_k is positive definite. Its eigenvalues (not the same λ_1!) must be positive. Its determinant is their product, so all upper left determinants are positive.

If condition III holds, so does condition IV: According to Section 4.4, the kth pivot d_k is the ratio of $\det A_k$ to $\det A_{k-1}$. If the determinants are all positive, so are the pivots.

If condition IV holds, so does condition I: We are given positive pivots, and must deduce that $x^{\mathrm{T}}Ax > 0$. This is what we did in the 2 by 2 case, by completing the square. The pivots were the numbers outside the squares. To see how that happens for symmetric matrices of any size, we go back to *elimination on a symmetric matrix*: $A = LDL^{\mathrm{T}}$.

Example 1. Positive pivots 2, $\frac{3}{2}$, and $\frac{4}{3}$:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} = LDL^{\mathrm{T}}.$$

I want to split $x^{\mathrm{T}}Ax$ into $x^{\mathrm{T}}LDL^{\mathrm{T}}x$:

$$\text{If} \quad x = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \quad \text{then} \quad L^{\mathrm{T}}x = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} u - \frac{1}{2}v \\ v - \frac{2}{3}w \\ w \end{bmatrix}.$$

So $x^{\mathrm{T}}Ax$ is a sum of squares with the pivots 2, $\frac{3}{2}$, and $\frac{4}{3}$ as coefficients:

$$x^{\mathrm{T}}Ax = (L^{\mathrm{T}}x)^{\mathrm{T}}D(L^{\mathrm{T}}x) = \mathbf{2}\left(u - \frac{\mathbf{1}}{\mathbf{2}}v\right)^2 + \frac{\mathbf{3}}{\mathbf{2}}\left(v - \frac{\mathbf{2}}{\mathbf{3}}w\right)^2 + \frac{\mathbf{4}}{\mathbf{3}}(w)^2.$$

Those positive pivots in D multiply perfect squares to make $x^{\mathrm{T}}Ax$ positive. Thus condition IV implies condition I, and the proof is complete.

It is beautiful that elimination and completing the square are actually the same. Elimination removes x_1 from all later equations. Similarly, the first square accounts for all terms in $x^{\mathrm{T}}Ax$ involving x_1. The sum of squares has the pivots outside. *The multipliers ℓ_{ij} are inside*! You can see the numbers $-\frac{1}{2}$ and $-\frac{2}{3}$ inside the squares in the example.

Every diagonal entry a_{ii} must be positive. As we know from the examples, however, it is far from sufficient to look only at the diagonal entries.

The pivots d_i are not to be confused with the eigenvalues. For a typical positive definite matrix, they are two completely different sets of positive numbers, In our 3 by 3 example, probably the determinant test is the easiest:

$$\textbf{Determinant test} \qquad \det A_1 = 2, \quad \det A_2 = 3, \quad \det A_3 = \det A = 4.$$

The pivots are the ratios $d_1 = 2$, $d_2 = \frac{3}{2}$, $d_3 = \frac{4}{3}$. Ordinarily the eigenvalue test is the longest computation. For this A we know the λ's are all positive:

$$\textbf{Eigenvalue test} \qquad \lambda_1 = 2 - \sqrt{2}, \quad \lambda_2 = 2, \quad \lambda_3 = 2 + \sqrt{2}.$$

Even though it is the hardest to apply to a single matrix, eigenvalues can be the most useful test for theoretical purposes. ***Each test is enough by itself***.

Positive Definite Matrices and Least Squares

I hope you will allow one more test for positive definiteness. It is already close. We connected positive definite matrices to pivots, determinants and eigenvalues.
Now we see them in the least-squares problems, coming from the rectangular matrices

The rectangular matrix will be R and the least-squares problem will be $Rx = b$. It has m equations with $m \geq n$ (square systems are included). *The least-square choice $\widehat{x}$ is the solution of* $R^{\mathrm{T}}R\widehat{x} = R^{\mathrm{T}}b$. That matrix $AR^{\mathrm{T}}R$ is not only symmetric but positive definite, as we now show—provided that the n columns of R are linearly independent:

6C The symmetric matrix A is positive definite if and only if

(V) There is a matrix R with independent columns such that $A = R^{\mathrm{T}}R$.

The key is to recognize $x^{\mathrm{T}}Ax$ *as* $x^{\mathrm{T}}R^{\mathrm{T}}Rx = (Rx)^{\mathrm{T}}(Rx)$. This squared length $||Rx||^2$ is positive (unless $x = 0$), because R has independent columns. (If x is nonzero then Rx is nonzero.) Thus $x^{\mathrm{T}}R^{\mathrm{T}}Rx > 0$ and $R^{\mathrm{T}}R$ is positive definite.

It remains to find an R For which $A = R^{\mathrm{T}}R$. We have almost done this twice already:

Elimination $\quad A = LDL^{\mathrm{T}} = (L\sqrt{D})(\sqrt{D}L^{\mathrm{T}})$. So take $R = \sqrt{D}L^{\mathrm{T}}$.

This ***Cholesky decomposition*** has the pivots split evenly between L and L^{T}.

Eigenvalues $\quad A = Q\Lambda Q^{\mathrm{T}} = (Q\sqrt{\Lambda})(\sqrt{\Lambda}Q^{\mathrm{T}})$. So take $R = \sqrt{\Lambda}Q^{\mathrm{T}}$.

A third possibility is $R = Q\sqrt{\Lambda}Q^{\mathrm{T}}$, the ***symmetric positive definite square root*** of A. There are many other choices, square or rectangular, and we can see why. If you multiply any R by *a matrix Q with orthonormal columns*, then $(QR)^{\mathrm{T}}(QR) = R^{\mathrm{T}}Q^{\mathrm{T}}QR = R^{\mathrm{T}}IR = A$. Therefore QR is another choice.

Minimum Principles

In this section we escape for the first time from linear equations. The unknown x will not be given as the solution to $Ax = b$ or $Ax = \lambda x$. Instead, the vector x will be determined by a minimum principle.

It is astonishing how many natural laws can be expressed as minimum principles. Just the fact that heavy liquids sink to the bottom is a consequence of minimizing their potential energy. And when you sit on a chair or lie on a bed, the springs adjust themselves so that the energy is minimized. A straw in a glass of water looks bent because light reaches your eye as quickly as possible. Certainly there are more highbrow examples: The fundamental principle of structural engineering is the minimization of total energy.

We have to say immediately that these "energies" are nothing but *positive definite quadratic functions*. And the derivative of a quadratic is linear. We get back to the familiar linear equations, when we set the first derivatives to zero. Our first goal in this section is ***to find the minimum principle that is equivalent to $Ax = b$, and the minimization equivalent to $Ax = \lambda x$.*** We will be doing in finite dimensions exactly what the theory of optimization does in a continuous problem, where "first derivatives $= 0$" gives a differential equation. In every problem, we are free to solve the linear equation or minimize the quadratic.

The first step is straightforward: We want to find the "parabola" $P(x)$ whose minimum occurs when $Ax = b$. If A is just a scalar, that is easy to do:

$$\text{The graph of} \quad P(x) = \frac{1}{2}Ax^2 - bx \quad \text{has zero slope when} \quad \frac{dP}{dx} = Ax - b = 0.$$

This point $x = A^{-1}b$ will be a minimum if A is positive. Then the parabola $P(x)$ opens upward (Figure 6.4). In more dimensions this parabola turns into a parabolic bowl (a paraboloid). To assure a minimum of $P(x)$, not a maximum or a saddle point, A must be positive definite!

Theorem. If A is symmetric positive definite, then $P(x) = \frac{1}{2}x^{\mathrm{T}}Ax - x^{\mathrm{T}}b$ reaches its minimum at the point where $Ax = b$. At that point $P_{\min} = -\frac{1}{2}b^{\mathrm{T}}A^{-1}b$.

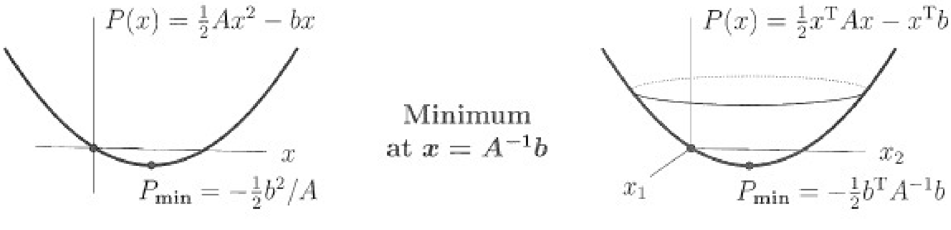

Figure: The graph of a positive quadratic $P(x)$ is a parabolic bowl.

Proof. Suppose $Ax = b$. For any vector y, we show that $P(y) \geq P(x)$:

$$\begin{aligned}
P(y) - P(x) &= \frac{1}{2}y^{\mathrm{T}}Ay - y^{\mathrm{T}}b - \frac{1}{2}x^{\mathrm{T}}Ax + x^{\mathrm{T}}b \\
&= \frac{1}{2}y^{\mathrm{T}}Ay - y^{\mathrm{T}}Ax + \frac{1}{2}x^{\mathrm{T}}Ax \quad (\text{set } b = Ax) \\
&= \frac{1}{2}(y-x)^{\mathrm{T}}A(y-x).
\end{aligned}$$

This can't be negative since A is positive definite—and it is zero only if $y - x = 0$. At all other points $P(y)$ is larger than $P(x)$, so the minimum occurs at x.

Example 1. Minimize $P(x) = x_1^2 - x_1x_2 + x_2^2 - b_1x_1 - b_2x_2$. The usual approach, by calculus, is to set the partial derivatives to zero. This gives $Ax = b$:

$$\begin{aligned}
\partial P/\partial x_1 &= 2x_1 - x_2 - b_1 = 0 \\
\partial P/\partial x_2 &= -x_1 + 2x_2 - b_2 = 0
\end{aligned} \qquad \text{means} \qquad \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Linear algebra recognizes this $P(x)$ as $\frac{1}{2}x^{\mathrm{T}}Ax - x^{\mathrm{T}}b$, and knows immediately that $Ax = b$ gives the minimum. Substitute $x = A^{-1}b$ into $P(x)$:

$$\textbf{Minimum value} \qquad P_{\min} = \frac{1}{2}(A^{-1}b)^{\mathrm{T}}A(A^{-1}b) - (A^{-1}b)^{\mathrm{T}}b = -\frac{1}{2}b^{\mathrm{T}}A^{-1}b.$$

In applications, $\frac{1}{2}x^{\mathrm{T}}Ax$ is the internal energy and $-x^{\mathrm{T}}b$ is the external work. The system automatically goes to $x = A^{-1}b$, where the total energy $P(x)$ is a minimum.