1 Extensions of Sylvester’s criterion

Last time, we showed that if $\Delta_1, \Delta_2, \ldots, \Delta_n$ are the principal minors of an $n \times n$ symmetric matrix $A$, then we can write the quadratic form associated to $A$ as

$$x^T A x = \Delta_1 y_1^2 + \frac{\Delta_2}{\Delta_1} y_2^2 + \frac{\Delta_3}{\Delta_2} y_3^2 + \cdots + \frac{\Delta_n}{\Delta_{n-1}} y_n^2,$$

after a substitution $y = U x$ for some invertible matrix $U$. In particular, Sylvester’s criterion follows from this: $A$ is positive definite if and only if $\Delta_1, \ldots, \Delta_n > 0$.

Similarly, we can test if $A$ is negative definite. For this, all the coefficients $\Delta_1, \frac{\Delta_2}{\Delta_1}, \ldots, \frac{\Delta_n}{\Delta_{n-1}}$ should be negative, which means

$$\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 > 0, \ldots, (-1)^n \Delta_n > 0.$$

In other words, the principal minors should alternate signs. (We can also derive this rule by testing if $-A$ is positive definite.)

In cases where $\Delta_1, \Delta_2, \ldots, \Delta_n$ are all nonzero, this finishes the classification:

- If they’re all positive, then $A \succ 0$.
- If $\Delta_1 < 0$ and then they alternate signs, then $A \prec 0$.
- If neither of these patterns holds, then $A$ is indefinite.

If any of the principal minors are 0, then the matrix might be positive semidefinite or negative semidefinite (it can’t be positive definite or negative definite). But testing for that is harder. There is a version of Sylvester’s criterion for that, but it requires taking many more determinants.

Another consequence of Sylvester’s criterion is that we can now prove a version of the second-derivative test for local minimizers and local maximizers.

**Theorem 1.1.** Let $f : \mathbb{R}^n \to \mathbb{R}$ have continuous second derivatives, and let $x^* \in \mathbb{R}^n$ be a critical point of $f$. Then:

(a) If $Hf(x^*) \succ 0$, then $x^*$ is a strict local minimizer.

(b) If $Hf(x^*) \prec 0$, then $x^*$ is a strict local maximizer.

(c) If $Hf(x^*)$ is indefinite, then $x^*$ is neither a local minimizer nor a local maximizer.

---

1This document comes from the Math 484 course webpage: https://faculty.math.illinois.edu/~mlavrov/courses/484-fall-2018.html
Proof. Pick a small radius \( r > 0 \), to be decided later, and take any point \( x \) such that \( \|x - x^*\| < r \).

By the Taylor series expansion we used a week ago, we have

\[
f(x) = f(x^*) + (x - x^*)^T Hf(x^*)(x - x^*)
\]

for some \( \xi \) between \( x \) and \( x^* \). (We stated this in terms of a one-dimensional “slice” of \( f \), but this statement is equivalent.) In particular, \( \|\xi - x^*\| < r \) as well.

In case (a), since \( Hf(x^*) > 0 \), the principal minors \( \Delta_1, \Delta_2, \ldots, \Delta_n \) of \( Hf(x^*) \) are also all positive. If the second derivatives of \( f \) are continuous, then \( \Delta_1, \ldots, \Delta_n \) are also continuous functions of \( x \), so they will remain positive very close to \( x^* \). In other words, we can choose \( r \) small enough that \( Hf(\xi) \) is also guaranteed to be positive definite.

But then, \((x - x^*)^T Hf(\xi)(x - x^*) > 0\), which means \( f(x) > f(x^*) \), and \( x^* \) is a local minimizer.

The argument for case (b) is the same, except that the condition on \( \Delta_1, \ldots, \Delta_n \) has different signs.

Finally, in case (c), since \( Hf(x^*) \) is indefinite, we can pick two directions \( u, v \in \mathbb{R}^n \) such that \( u^T Hf(x^*)u > 0 \) but \( v^T Hf(x^*)v < 0 \). Essentially, this means that going a tiny step in the direction \( u \) will increase \( f \), but going a tiny step in the direction \( v \) will decrease \( f \). This means that \( x^* \) is neither a local minimizer nor a local maximizer.

To make this formal: the function \( u^T Hf(x^* + tu)u \) is a continuous function of \( x \). It’s positive when \( t = 0 \), so there is some \( r > 0 \) such that it’s positive for \( t \in [0, r] \). Set \( x = x^* + ru \). Then

\[
f(x) = f(x^*) + (x - x^*)^T Hf(x^* + tu)(x - x^*) \quad \text{for some } t \in [0, r]
\]

\[
= f(x^*) + (ru)^T Hf(x^* + tu)(ru)
\]

\[
= f(x^*) + r^2 (u^T Hf(x^* + tu)u) > f(x^*).
\]

If we repeat the same argument with \( v \) instead of \( u \), we get a point \( x \) with \( f(x) < f(x^*) \). \( \square \)

2 Using eigenvalues

For an \( n \times n \) matrix \( A \), if a nonzero vector \( x \in \mathbb{R}^n \) satisfies \( Ax = \lambda x \) for some scalar \( \lambda \in \mathbb{R} \), we call \( \lambda \) an eigenvalue of \( A \) and \( x \) its associated eigenvector.

When \( A \) is symmetric, we are guaranteed good behavior from eigenvalues: we are guaranteed \( n \) real eigenvalues, possibly with repetition. They will also allow us to classify the matrix \( A \) as positive definite, positive semidefinite, and so on.

Intuitively, this will happen because, if \( x \) is an eigenvector of \( A \) with eigenvalue \( \lambda \), then

\[
x^T Ax = x^T \lambda x = x \cdot \lambda x = \lambda \|x\|^2.
\]

Since \( \|x\|^2 \) is always positive, the sign of \( x^T Ax \) depends on the sign of \( \lambda \).

To turn this into a proof, we essentially need to know that the behavior of \( x^T Ax \) when \( x \) is an eigenvector (which is really a very specific case) actually describes \( x^T Ax \) for all \( x \).

This is true by the spectral theorem, which we will use as a black box.
Theorem 2.1 (Spectral theorem for symmetric matrices). Let $A$ be a symmetric $n \times n$ matrix. Then $A$ has an orthonormal basis of eigenvectors $q^{(1)}, \ldots, q^{(n)}$: that is, we can choose these vectors to satisfy $\|q^{(i)}\| = 1$, $q^{(i)} \cdot q^{(j)} = 0$ when $i \neq j$, and $Aq^{(i)} = \lambda_i q^{(i)}$ for some $\lambda_1, \ldots, \lambda_n$.

Equivalently, we can factor

$$A = Q\Lambda Q^T$$

where $Q$ is an orthogonal matrix (it satisfies $QQ^T = Q^TQ = I$) and $\Lambda$ is a diagonal matrix whose entries are the eigenvalues of $A$.

If you have not encountered the spectral theorem before, it is a good exercise in working with matrices to check that the second form of the spectral theorem is equivalent. To turn the first form of the theorem into the second, let $Q$ be the matrix whose columns are the eigenvectors $q^{(1)}, \ldots, q^{(n)}$, and let $\Lambda$ be the diagonal matrix whose main diagonal has the values $\lambda_1, \ldots, \lambda_n$.

Theorem 2.2. Let $A$ be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then:

- $A \succeq 0$ if $\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0$.
- $A \succ 0$ if $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$.
- $A \preceq 0$ if $\lambda_1, \lambda_2, \ldots, \lambda_n \leq 0$.
- $A \prec 0$ if $\lambda_1, \lambda_2, \ldots, \lambda_n < 0$.
- $A$ is indefinite if it has both positive and negative eigenvalues.

Proof. Either form of the spectral theorem can be used to prove this result. First, here’s a proof using the first form.

If the eigenvectors $q^{(1)}, \ldots, q^{(n)}$ form a basis, then we can write any vector $x$ as

$$x = c_1 q^{(1)} + \cdots + c_n q^{(n)}$$

for some constants $c_1, \ldots, c_n$. Then

$$x^T A x = (c_1 q^{(1)} + \cdots + c_n q^{(n)})^T A (c_1 q^{(1)} + \cdots + c_n q^{(n)})$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} (c_i q^{(i)})(c_j q^{(j)}) A_{ij}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} (c_i q^{(i)})^T \lambda_j c_j q^{(j)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j c_i c_j (q^{(i)} \cdot q^{(j)})$$

$$= \lambda_1 c_1^2 + \lambda_2 c_2^2 + \cdots + \lambda_n c_n^2$$

where the last step is true because $q^{(i)} \cdot q^{(j)}$ simplifies to 1 when $i = j$ and 0 otherwise.

Then we proceed as usual: if $\lambda_1, \ldots, \lambda_n \geq 0$, then we can conclude that $x^T A x \geq 0$ no matter what $x$ is. If in fact $\lambda_1, \ldots, \lambda_n > 0$, then we can conclude that $x^T A x > 0$ no matter what $x$ is—unless $c_1 = \cdots = c_n = 0$, in which case $x = 0$. The same argument holds for the next two cases.
Second, here’s an entirely separate proof using the second form. Given the factorization $A = Q\Lambda Q^T$, if $\lambda_1, \ldots, \lambda_n \geq 0$, then we can define $\Lambda^{1/2}$ as the diagonal matrix with entries $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$. This lets us factor $A$ as

$$A = QA^{1/2}\Lambda^{1/2}Q^T = (QA^{1/2})(QA^{1/2})^T$$

which means $A = B^TB$ for $B = (QA^{1/2})^T$. As we saw before, such a factorization guarantees that $A$ is positive semidefinite.

The matrix $Q$ is always invertible: its inverse is $Q^T$. If, moreover, $\lambda_1, \ldots, \lambda_n > 0$, then $\Lambda^{1/2}$ is invertible: just take the reciprocal of each diagonal entry. So in this case, $B$ is invertible, and this means $A$ is positive definite.

If $\lambda_1, \ldots, \lambda_n \leq 0$, then we can write $-A = Q(-\Lambda)Q^T$ and apply the preceding argument to classify $-A$, finishing the other two cases.

For both proofs, the indefinite case is handled the same way, and doesn’t need to use the spectral theorem at all. If $x$ is an eigenvector of $A$ with eigenvalue $\lambda > 0$, and $y$ is an eigenvector of $A$ with eigenvalue $\mu < 0$, then $x^TAx = x^T\lambda x = \lambda \|x\|^2 > 0$, but $y^TAy = y^T\mu y = \mu \|y\|^2 < 0$, so $A$ is indefinite.

To find the eigenvalues in order to apply this test, we do the following. Suppose $x$ is an eigenvector with eigenvalue $\lambda$. Then $Ax = \lambda x$ means that $(A - \lambda I)x = 0$, so $A - \lambda I$ is singular: this means that the determinant $\det(A - \lambda I)$ is 0. So we solve the equation $\det(A - \lambda I) = 0$ for $\lambda$.

In general, this is harder to do by hand than Sylvester’s criterion, because even for a $3 \times 3$ matrix it requires solving a cubic equation. The eigenvalue test is useful for other reasons:

- Computers are very good at solving polynomial equations numerically, so they usually have no trouble with the eigenvalue test.
- The eigenvalue test can tell if a matrix is positive semidefinite but not positive definite (or negative semidefinite but not negative definite), while Sylvester’s criterion becomes much more complicated in that case.
- When we prove theorems that require working with positive definite matrices, the eigenvalue test is usually more useful.
- In cases where the matrix is indefinite, the eigenvectors give us specific vectors $x$ that make the quadratic form positive and negative.

We can use the eigenvalue test to fill in the proof of Sylvester’s criterion when a principal minor $\Delta_k$ is 0: we need to confirm that in such cases, $A$ cannot be positive definite.

First, we show that if $A \succ 0$, then $\Delta_n = \det A$ cannot be zero. This is because $\det A$ is the product $\lambda_1\lambda_2\cdots\lambda_n$ of all the eigenvalues of $A$. If all eigenvalues are positive, then $\det A > 0$.

Second, if $A \succ 0$, then the $k \times k$ submatrix whose determinant is $\det A$ must also be positive definite. To see this, consider $x^TAx$ when the last $n - k$ entries of $x$ are 0. This product ignores all but the top left $k \times k$ entries of $A$, so it’s the same as the quadratic form for that $k \times k$ submatrix.

So by the argument that $\Delta_n > 0$, we must also have $\Delta_k > 0$. 

4