The principle submatrix trick (Sylvester's criterion)

Andrew Guy*

When testing whether a matrix is positive definite ($x^t Ax > 0$), the course today only teaches to diagonalize and check the eigenvalues. Back when I took 3M1, I was introduced to another method, in which we test whether all principle submatrices have positive determinant.

Strategy

We are going to take our matrix $A \in \mathbb{R}^{n\times n}$ and break it up as follows:

$$A = \begin{bmatrix} M_k & B_k \\ B_k^t & N_k \end{bmatrix} \tag{1}$$

with $1 \leq k \leq n$. Then, by induction, we are going to show that the matrix M_k being positive definite (and therefore having positive determinant as $\det M_k = \prod_{i=1}^k \lambda_i$ and all $\lambda_i > 0$) and M_{k+1} having positive determinant implies M_{k+1} is positive definite. We start with $M_1 = a_{1,1} > 0$ and induct, but first need a little trick regarding the determinant of block matrices...

Determinant of block matrices

Consider the expression

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Au + Bv \\ Cu + Dv \end{bmatrix} \tag{2}$$

and the change in volume associated with transforming a unit cube by this matrix — the determinant. We motivate this with a simple example in 3 dimensions, then we generalize[1]. Suppose that the matrix is given by

$$M = \begin{bmatrix} A & v \\ 0 & 1 \end{bmatrix} \tag{3}$$

with A being 2×2. We have that our first two unit vectors are stretched into a parallelogram with a certain area $\det A$, then layers of this are stacked in the direction v. But note that the total volume is not related to the length of the final column vector $[v^t, 1]^t$ — only the orthogonal component matters. We could

*aag43 (at) cam (dot) ac (dot) uk

[1] Actually, this example is quite close to our intended use case as we will be expanding our matrices 1 row/column at a time

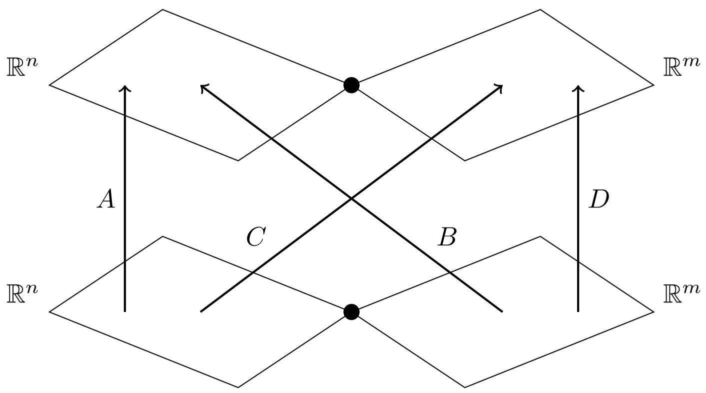

Figure 1: The mappings contained within the block matrix.

shear this back such that the layers were stacked along the direction $[0, 0, 1]^t$, and the volume remains.

Our strategy is therefore to shear the components of the two orthogonal subspaces into a form in which we can treat the total volume change as the product of the volume changes in the two orthogonal components. Further motivated by our use case, we want our expression to have a term of $\det A$, so we want to first shear in a way that multiplication by A alone gives the required $Au + Bv$ — we are going to use the pathway $CA^{-1}B$ from Fig. 1 to achieve this, and compensate for that by our lower right transform becoming $D - P$ for some P, which we shall see is simply the pathway $P = CA^{-1}B$ we have used.

Thus, we can decompose the transform into

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \tag{4}$$

and the determinant is $\det A \det(D - CA^{-1}B)$.

Sylvester's criterion

We are now ready to induct. Knowing that M_k is positive definite, we expand our matrix

$$M_{k+1} = \begin{bmatrix} M_k & a \\ a^t & b \end{bmatrix} \tag{5}$$

and evaluate $x^t M_{k+1} x$ for a vector $x = \begin{bmatrix} u^t & v^t \end{bmatrix}^t$ with v scalar:

$$x^t M_{k+1} x = u^t M_k u + u^t a v + v a^t u + b v^2. \tag{6}$$

Much like quadratics in the single variable case, we can complete the square:

$$x^t M_{k+1} x = \left(u + v M_k^{-1} a\right)^t M_k (u + v M_k^{-1} a) + v^2 (b - a^t M_k^{-1} a) \tag{7}$$

We have a term of $c^t M_k c$ which we know is positive by the hypothesis of the induction, so we require the term $b - a^t M_k^{-1} a > 0$ for M_{k+1} to be positive definite. Now we use the block matrix determinant trick: $\det M_{k+1} = \det M_k \det(b - a^t M_k^{-1} a) = (b - a^t M_k^{-1} a) \det M_k$ (as it is a scalar), so if we

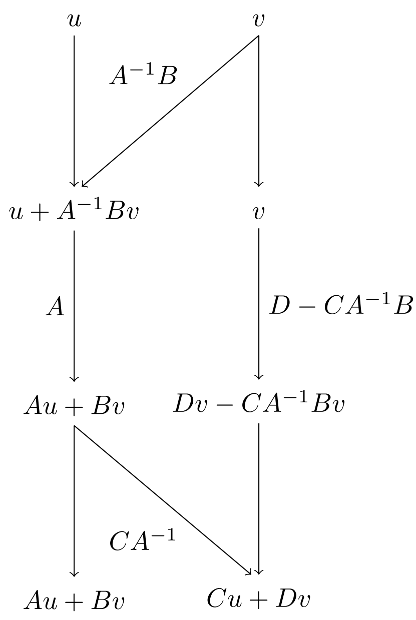

Figure 2: Two shears and two self contained transforms

.

test and find $\det M_{k+1} > 0$ we know that $b - a^t M_k^{-1} a > 0$ (as $\det M_k > 0$ by hypothesis) hence that M_{k+1} is positive definite.

Therefore, we know by induction that if all the principle submatrices have positive determinant, the matrix itself is positive definite.

Picking the submatrices

In the supervision, I told students that I strongly suspected that you do not have to march down the diagonal when picking submatrices, and that if there were more favourable submatrices for evaluation you could start anywhere and expand by picking another index, then adding in the row/column of that index. The proof of this relies on the fact that we can pick our indices beforehand, then swap rows and columns such that the matrix is reordered in a way that we can take principle submatrices down the diagonal.

Let us swap the desired first index into first place by the permutation: $A' = P_1 A P_1$. As we are swapping two elements, we have $P_1^t = P_1$. Now we can bring our desired second index into second position by another permutation $A'' = P_2 P_1 A P_1 P_2$, and so on until we have our overall permutation P and the transform $A' = P^t A P$. Now we find that positive definiteness of A' is equivalent to positive definiteness of A, as we can equivalently wrap the P into the test vector as such: $x^t A' x = (Px)^t A (Px)$.

That approach is ok, but it requires reordering the submatrices when you write them out to tack the selected row and column elements in the correct order, which may be confusing. So, can we just take the submatrix as-is from the active indices at stage k? The answer, naturally, is yes — we have already seen that a permutation of both rows and columns does not change the positive definiteness of a matrix, and the determinants also remain unchanged (we could see this by noting that the required permutation would have determinant ± 1 and is applied twice, so any negation would cancel).