Theorem 0.1. (Cauchy-Binet) Let $A$ be a $k \times n$ matrix and $B$ be an $n \times k$ matrix. Then

$$\det(AB) = \sum_J \det(A(J)) \det(B(J)),$$

where $J = (j_1, j_2, \ldots, j_k)$, $1 \leq j_1 < j_2 < \cdots < j_k \leq n$, runs through all such multi-indices, $A(J)$ denotes the matrix formed from $A$ using columns $J$ (in that order), and $B(J)$ denotes the matrix formed using rows $J$ of $B$ in that order.

Proof. By definition of matrix product

$$\det AB = \det \begin{bmatrix} \sum_{j_1=1}^n a_{1j_1} b_{j_11} & \cdots & \sum_{j_k=1}^n a_{1j_k} b_{j_k1} \\ \vdots & & \vdots \\ \sum_{j_1=1}^n a_{kj_1} b_{j_11} & \cdots & \sum_{j_k=1}^n a_{kj_k} b_{j_k1} \end{bmatrix}$$

by the multi-linearity of the determinant. Since $\det(A(j_1, j_2, \ldots, j_k) = 0$ if the indices $j_i$ are not all distinct, only those sets of indices occur in the sum. For a fixed multi-index $J' = (j'_1, j'_2, \ldots, j'_k)$ with $1 \leq j'_1 < j'_2 < \cdots < j'_k \leq n$ and $J$ some multi-index with these indices in some order, let $j'_i = j_{\sigma(i)}$ where $\sigma$ is a permutation of $[n]$. Then

$$\det(A(j_1, j_2, \ldots, j_k) = sgn(\sigma) \det(A(j'_1, j'_2, \ldots, j'_k)).$$

Now let $J'$ be fixed, and sum over all $J$ which are permutations of $J'$. Let $\tau$ be the inverse of $\sigma$. Then $j_i = j_{\sigma(i)} = j'_{\tau(i)}$. So the sum multiplying $\det(A(j'_1, j'_2, \ldots, j'_k) = \det(A(J'))$ is

$$\sum_{\sigma} sgn(\sigma)b_{j'_{\tau(1)}1}b_{j'_{\tau(2)}2} \cdots b_{j'_{\tau(k)k}}$$

$$= \sum_{\tau} sgn(\tau)b_{j'_11}b_{j'_22} \cdots b_{j'_kk}$$

$$= \det B(J').$$

Hence

$$\det(AB) = \sum_{J'} \det(A(J')) \det(B(J')).$$

Corollary 0.1.
\[ \det AA^T = \sum_J (\det A(J))^2. \]

Here’s an application.

Corollary 0.2. Let \( \Pi \) be a \( k \)-parallelepiped in \( \mathbb{R}^n \) and let \( \Pi_J \) be the orthogonal projection of \( \Pi \) onto the \( k \)-dimensional subspace spanned by the \( x_J \) axes. Let \( m_J = \mu(\Pi_J) \) be the \( k \)-dimensional measure of this \( k \)-parallelepiped. Then
\[ (\mu(\Pi))^2 = \sum_J m_J^2 = \sum_J \mu(\Pi_J)^2. \]

Proof. Recall that if \( v_1, v_2, \ldots, v_k \) are the \( k \) row vectors which are the spanning edges of \( \Pi \) and \( V \) is the \( k \times n \) matrix defined by
\[
V = \begin{bmatrix}
v_1 \\
v_2 \\
\vdots \\
v_k
\end{bmatrix},
\]
then the measure of \( \Pi \) is \( \sqrt{\det VV^T} \). Now use the previous corollary and interpret each term in the sum as the square of the measure of a \( k \)-parallelepiped in \( \mathbb{R}^k \).

This is a sort of Pythagorean theorem, generalizing the length (1-dimensional measure) formula for a line segment in \( \mathbb{R}^n \).

Application 0.1. Let the surface \( S \) in \( \mathbb{R}^4 \) be defined by the parameterization
\[
(x, y) \rightarrow (x, y, f(x, y), g(x, y)), (x, y) \in D \subset \mathbb{R}^2.
\]
Then the area of \( S \) is
\[
\int_D \left( 1 + f_x^2 + f_y^2 + g_x^2 + g_y^2 + \left( \frac{\partial(f, g)}{\partial(x, y)} \right)^2 \right)^{1/2} dx dy.
\]
The general result for this type of parameterization is as follows. Let \( (x_1, x_2, \ldots, x_k) = x_K \in D \subset \mathbb{R}^k \) and let \( (f_1(x_K), \ldots, f_m(x_K)) = f_M(x_K) \) be an \( m \)-tuple of differentiable functions defined on \( D \). Let \( \mathcal{M} = \{(x_K, f_M(x_K) : x_K \in D\} \). Then the \( k \)-dimensional measure of \( \mathcal{M} \) is
\[
\int_D \left( 1 + \sum_{I, J, 1 \leq |I| = |J| \leq k} \left( \frac{\partial f_I}{\partial x_J} \right)^2 \right)^{1/2} dx_K.
\]