Scaling Factors for Measure

John McCuan

July 25, 2014

1 The Question

I mentioned in class that given two vectors v_1 and v_2 in $\mathbb{R}^3$, the area of the parallelogram spanned by v_1 and v_2 is

$$\sqrt{A^T A} \tag{1}$$

where A is the 3×2 matrix with v_1 and v_2 in the columns. More generally, if $v_1, \ldots, v_k$ are vectors in $\mathbb{R}^n$, then they span a k-dimensional parallelopiped

$$\left\{ \sum_{j=1}^{k} a_j v_j : a_1, \ldots, a_k \in [0, 1] \right\}$$

in $\mathbb{R}^n$, and the k-dimensional measure of that set is given by the same formula (1) if we take A to be the $n \times k$ matrix with $v_1, \ldots, v_k$ in the columns. In particular, this will work for two vectors v_1 and v_2 to give the area of a parallelogram in a Euclidean space of any dimension.

One student, Yong Jea Kim, asked me to justify these assertions.

Using integration, one can extend these formulas to areas/lower dimensional measures of a variety of sets in Euclidean spaces of higher dimension. I'll record some of these formula at the end.

2 Area in $\mathbb{R}^3$

We will take as a definition the formula

$$\text{Area}(\mathcal{P}) = |\det P| \tag{2}$$

Where

$$\mathcal{P} = \{ap_1 + bp_2 : 0 \le a, b \le 1\} \subset \mathbb{R}^2$$

denotes the planar parallelogram spanned by two vectors p_1 and p_2 in $\mathbb{R}^2$, and P denotes the 2×2 matrix with those two vectors as columns.

Let us begin with a simple justification in the initial case $k = 2$ and $n = 3$. In this special case,

$$\det A^T A = \begin{vmatrix} \|v_1\|^2 & v_1 \cdot v_2 \\ v_1 \cdot v_2 & \|v_2\|^2 \end{vmatrix} = \|v_1\|^2 \|v_2\|^2 - (v_1 \cdot v_2)^2. \tag{3}$$

Notice that in the last expression, some decidedly geometric quantities appear. Namely, $\|v_j\|$ is the length of the vector v_j for $j = 1, 2$, and $v_1 \cdot v_2$ is the product of the two lengths and the cosine of the angle between the vectors.

Now, let us imagine $\mathcal{P}$ to be the parallelogram spanned by v_1 and v_2 in the two dimensional plane $V = \text{span}\{v_1, v_2\}$ considered as an abstract vector space. There is some orthonormal basis for V. For example, we could apply the Gram-Schmidt procedure to $\{v_1, v_2\}$ to get one. We don't need to write down this basis explicitly; it is just enough to know it's there.

Next, given the orthonormal basis $\{u_1, u_2\}$ for V, we can define a linear transformation $T : V \to V$ by assigning images for u_1 and u_2. Let us consider the particular linear transformation with

$$T(u_1) = v_1 \qquad \text{and} \qquad T(u_2) = v_2.$$

Notice that there is a square Q spanned by u_1 and u_2 in V, and the parallogram $\mathcal{P}$ in which we are interested is the image $T(Q)$ of Q under the transformation T.

We can express T in coordinates. In particular, let us take the basis $\{u_1, u_2\}$ of V. Then there is a matrix

$$P = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$$

associated with the linear transformation T. We know the absolute value of the determinant of that matrix is the area of $\mathcal{P}$. Therefore,

$$[\text{Area}(\mathcal{P})]^2 = (w_{11}w_{22} - w_{12}w_{21})^2 = w_{11}^2 w_{22}^2 - 2w_{11}w_{22}w_{12}w_{21} + w_{12}^2 w_{21}^2.$$

On the other hand, what is the meaning of the matrix $P = (w_{ij})$? One answer is that the columns give the coordinates of v_1 and v_2 in the basis $\{u_1, u_2\}$. This means

$$v_1 = w_{11}u_1 + w_{21}u_2 \quad \text{and} \quad v_1 = w_{12}u_1 + w_{22}u_2.$$

Given that u_1 and u_2 are orthonormal, we can easily compute the geometric quantities appearing in (3) in terms of the entires w_{ij}. That computation is as follows:

$$\begin{aligned} \|v_1\|^2 \|v_2\|^2 - (v_1 \cdot v_2)^2 &= (w_{11}^2 + w_{21}^2)(w_{12}^2 + w_{22}^2) - (w_{11}w_{12} + w_{21}w_{22})^2 \\ &= w_{11}^2 w_{22}^2 + w_{21}^2 w_{12}^2 - 2w_{11}w_{12}w_{21}w_{22} \\ &= (\det \mathcal{P})^2 . \end{aligned}$$

Thus, we have shown by explicit calculation that formula (1) holds in the special case where we have two vectors in $\mathbb{R}^3$.

This seems a little long, but I have given *all* the details. If you understand the background, it's really just a couple lines of calculations.

3 Other Dimensions

The general problem is a little more tricky, and there may be a simpler way to see it than the following. However, we will not use anything other than material covered in class. Again, we take the full dimension definition of measure as given:

$$\mathcal{H}^k(\mathcal{P}) = |\det(P)|$$

where $\mathcal{H}^k$ is k-dimensional measure, $\mathcal{P}$ is a k dimensional parallelopiped in $\mathbb{R}^k$, and P is the $k \times k$ matrix whose columns are the vectors spanning $\mathcal{P}$.

The approach is similar to the one in the last section, but we will not make an explicit calculation but use instead some important facts we learned about A^T. To be precise, remember that given vectors $v_1, \dots, v_k \in \mathbb{R}^n$ which we take as the columns of a matrix A, the operation of exchanging "inner products" for "coordinates" with respect to $\{v_1, \dots, v_k\}$ is an invertible linear operation.

To be more precise, let $V = \text{span}\{v_1, \dots, v_k\}$, and let us assume $\{v_1, \dots, v_k\}$ is a basis. (The situation in which $\{v_1, \dots, v_k\}$ is linearly dependent can be considered as a degenerate special case.) We know there is a linear coordinate transformation $\chi : V \to \mathbb{R}^k$ which gives coordinates. That is, given a vector $w \in V$, we get a vector $\chi(w)$ in $\mathbb{R}^k$ whose entries are the coordinates of w with respect to the basis $\{v_1, \dots, v_k\}$. On the other hand, there is another vector in $\mathbb{R}^k$ which can be associated with w. That vector has as its entries the inner products $w \cdot v_1$, $w \cdot v_2$, ..., $w \cdot v_k$. This also clearly defines a linear transformation from V to $\mathbb{R}^k$. In fact, this linear transformation extends to all of $\mathbb{R}^n$ and is then precisely the transformation given by multiplication by A^T.

We considered these transformations carefully in class, and they were closely related to the projection operator onto V. In particular, they are each invertible on V and we observed the following:

> *The linear transformation from $\mathbb{R}^k$ to $\mathbb{R}^k$ given by multiplication by A^TA takes the coordinates of a vector in V and returns the inner products with respect to $\{v_1, \dots, v_k\}$.*
>
> *Moreover, the linear transformation from $\mathbb{R}^k$ to $\mathbb{R}^k$ given by multiplication by $(A^TA)^{-1}$ takes the vector of inner products with respect to $\{v_1, \dots, v_k\}$ and returns the coordinates of a vector in V.*

Since A^Tw gives the inner products, we see that

$$A(A^TA)^{-1}A^Tw = w$$

for each w in V. (And the same formula gives the projection onto V defined on all of $\mathbb{R}^n$.)

If we understand this, we are in a position to give the general argument. As before, we consider an orthonormal basis $\{u_1, \dots, u_k\}$ for V, and the linear transformation $T : V \to V$

for which $T(u_j) = v_j$ for $j = 1, \ldots k$. We express this linear transformation with respect to the basis $\{u_1, \ldots, u_k\}$ obtaining a $k \times k$ matrix P. We then know

$$[\mathcal{H}^k(\mathcal{P})]^2 = [\det P]^2.$$

Since coordinates in the basis $\{u_1, \ldots, u_k\}$ are just the same as inner products, we see that the i, j-th entry of P is $v_j \cdot u_i$. If we let Q denote the matrix with $u_1, \ldots, u_k$ in the columns, then another way to say this is that $P = Q^T A$. We note also that $P^T = A^T Q$.

This agrees with what we did before. But now we will do something a little different. Let's complete $\{u_1, \ldots, u_k\}$ to a basis for $\mathbb{R}^n$ by adjoining $n - k$ more orthonormal vectors $u_{k+1}, \ldots, u_n$. Denoting the $n \times (n - k)$ matrix with $u_{k+1}, \ldots, u_n$ in the columns by $\tilde{Q}$, we see that the $n \times n$ block matrix $\bar{Q} = (Q|\tilde{Q})$ is an orthogonal matrix. We will also consider the $n \times n$ block matrix $\bar{A} = (A|\tilde{Q})$. It is easy to check that

$$\bar{Q}^T \bar{A} = \left(\frac{Q^T}{\tilde{Q}^T}\right)(A|\tilde{Q}) = \begin{pmatrix} P & 0 \\ \tilde{Q}^T A & I \end{pmatrix} = \begin{pmatrix} P & 0 \\ 0 & I \end{pmatrix}$$

where I represents the $(n - k) \times (n - k)$ identity matrix. It follows that $\det(\bar{Q}^T \bar{A}) = \det P$. Similarly,

$$\bar{A}^T \bar{Q} = \begin{pmatrix} P^T & A^T \tilde{Q} \\ 0 & I \end{pmatrix} = \begin{pmatrix} P^T & 0 \\ 0 & I \end{pmatrix},$$

and $\det(\bar{A}^T \bar{Q}) = \det P^T = \det P$.

Therefore,

$$[\det P]^2 = \det(\bar{A}^T \bar{Q})\det(\bar{Q}^T \bar{A}) = \det(\bar{A}^T \bar{Q} \bar{Q}^T \bar{A})$$

by the product formula for determinants. Since $\bar{Q}$ is an orthogonal matrix, the middle factors are inverses, and

$$[\det P]^2 = \det(\bar{A}^T \bar{A}).$$

Finally,

$$\bar{A}^T \bar{A} = \left(\frac{A^T}{\tilde{Q}^T}\right)(A|\tilde{Q}) = \begin{pmatrix} A^T A & 0 \\ 0 & I \end{pmatrix}.$$

Thus, $[\mathcal{H}^k(\mathcal{P})]^2 = \det(\bar{A}^T \bar{A}) = \det(A^T A)$, which is what we wanted to prove.

4 Applications

These formulas were mentioned in class, and are straightforward applications of the discussion above when applied to the definition of integration. In each case, the area of a parallelopiped acts as a scaling factor for a measure.

1. If $f : \Omega \to \mathbb{R}^2$ is differentiable function defined on an open set $\Omega \subset \mathbb{R}^2$ and f has derivative
$$Df = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix},$$
then the area of $f(\Omega)$ is given by
$$\int_\Omega |\det Df|.$$
This means that in the particular case in which f is a linear function given by multiplication by a 2×2 matrix A, then
$$\text{Area}\, f(\Omega) = |\det A| \text{Area}\,(\Omega).$$

2. If $f : \Omega \to \mathbb{R}^n$ is a differentiable function defined on an open set $\Omega \subset \mathbb{R}^k$ and the derivative of f is the $n \times k$ matix
$$Df = \left(\frac{\partial f_i}{\partial x_j} \right),$$
then
$$\mathcal{H}^k[f(\Omega)] = \int_\Omega \sigma$$
where $\sigma = \sigma(x)$ is the scaling factor
$$\sigma = \sqrt{\det (Df)^T Df}.$$