AN ELEMENTARY DEFINITION OF SURFACE AREA IN $E^{n+1}$ FOR SMOOTH SURFACES

LOUIS I. ALPERT AND L. V. TORALBALLA

The present paper concerns the difficulty which one encounters in text books of Advanced Calculus of giving a simple and elementary definition of area of a smooth non-parametric surface in $E^{n+1}$ such that, within the same elementary framework, one can then prove that the area so defined is equal to the classical area integral.

The authors were first made aware of the considerable interest of such a task in 1955 with the publication of Angus Taylor’s now classic textbook “Advanced Calculus”. The following statement is taken from page 384 of this book:

“...It is logically and aesthetically desirable to have a definition of surface area which is directly geometric, and which does not put too many restrictions on the surface. A good definition ought not to depend upon the method of representing the surface analytically, and should not be limited to smooth surfaces. The demand for such a definition poses a very difficult problem, however. It may surprise the student to know that the problem has occupied the attention of many able mathematicians over the last fifty years, and that the end of research on the question is not yet in sight.”

In the present paper we present an idea which seems to answer the questions raised by Angus Taylor for surfaces $S: z=f(x_1, \ldots, x_n)$, which are continuous with their first order partial derivatives. The idea is to develop a scheme for the construction of sequences of suitably chosen polyhedra inscribed within the given surface, such that the corresponding sequences of the polyhedral areas converge to the classical area integral for the surface, and hence to the Lebesgue area of $S$.

In previous papers [1], [7] we discussed our definition of area for surfaces $S: z=f(x_1, x_2)$. In [7] we took in consideration surfaces $z=f(x_1, x_2)$ with $f$ continuous with its first order partial derivatives. In [1] we gave a necessary and sufficient condition in order that for a surface $z=f(x_1, x_2)$ there are sequences of inscribed polyhedra satisfying the requirements of our definitions (see [1]).

J. A. Serret [6] in 1868 proposed a geometric definition of area, but H. A. Schwartz [5] in 1882 proved that Serret’s definition was incorrect. Other geometric definitions of area and constructions have been proposed, and we mention here for example the ones of S. Kempisty [3] for surfaces $S: z=f(x_1, x_2)$ with $f$ absolutely continuous in the sense of Tonelli. For general expositions concerning area, in particular, Le-
The n-ary vector product. Consider the \((n+1)\)-dimensional Euclidean space \(E^{n+1}\), \(n \geq 2\). Let \(\{V_1, V_2, \cdots, V_n\}\), where for 

\[
i = 1, 2, \cdots, n, \quad V_i = (a_{i1}, a_{i2}, \cdots, a_{i,n+1}),
\]

be a set of \(n\) linearly independent vectors in \(E^{n+1}\). For an arbitrary vector \(X = (x_1, x_2, \cdots, x_{n+1})\) in \(E^{n+1}\) define

\[
\varphi(X) = \begin{vmatrix}
x_1 & x_2 & \cdots & x_{n+1} \\
a_{11} & a_{12} & \cdots & a_{1,n+1} \\
a_{21} & a_{22} & \cdots & a_{2,n+1} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{n,n+1}
\end{vmatrix}
\]

By elementary properties of the determinant, \(\varphi\) is a linear function from \(E^{n+1}\) into the reals; i.e. \(\varphi(X_1 + X_2) = \varphi(X_1) + \varphi(X_2)\) for every pair of vectors \(X_1\) and \(X_2\) in \(E^{n+1}\), and \(\varphi(aX) = a\varphi(X)\) for every vector \(X\) in \(E^{n+1}\) and every real number \(a\). Hence, there is a unique vector \(Z = (z_1, z_2, \cdots, z_{n+1})\) in \(E^{n+1}\) such that

\[
\varphi(X) = X \cdot Z = x_1z_1 + x_2z_2 + \cdots + x_{n+1}z_{n+1}
\]

for every vector \(X\) in \(E^{n+1}\). We denote this vector \(Z\) by

\[
V_1 \times V_2 \times \cdots \times V_n
\]

and call it the \(n\)-ary vector product of \(V_1, V_2, \cdots, V_n\).

It is clear from elementary properties of the determinant that \(V_1 \times V_2 \times \cdots \times V_n\) is orthogonal to each \(V_i\). Moreover, if \(i_1, i_2, \cdots, i_{n+1}\) is the natural vector basis of \(E^{n+1}\), then \(z_j = i_j \cdot Z = \varphi(i_j)\) for each \(j = 1, 2, \cdots, n + 1\), and \(V_1 \times V_2 \times \cdots \times V_n\) can be expressed by the formal determinant

\[
V_1 \times V_2 \times \cdots \times V_n = \begin{vmatrix}
i_1 & i_2 & \cdots & i_{n+1} \\
a_{11} & a_{12} & \cdots & a_{1,n+1} \\
a_{21} & a_{22} & \cdots & a_{2,n+1} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{n,n+1}
\end{vmatrix}
\]

The subspace of \(E^{n+1}\) which is spanned by the vectors \(V_1, V_2, \cdots, V_n\) is called an \(n\)-hyperplane in \(E^{n+1}\). We say that the vector

\[
V_1 \times V_2 \times \cdots \times V_n
\]

is normal to this hyperplane.
2. The \( n \)-hedra. Given a set of \( n + 1 \) points in \( E^{n+1} \), if the matrix of the coordinates of these \( n + 1 \) points is of rank \( n \), this set determines an \( n \)-hedron, or \( n \)-simplex. This is the closed convex subset of \( E^{n+1} \) which is bounded by the \( n + 1 \) \((n-1)\)-hyperplanes determined by the given set of \( n + 1 \) points. An \( n \)-hedron determines the \( n \)-hyperplane in which it lies.

Given two vectors \( U \) and \( V \) in \( E^{n+1} \), the angle \( \alpha = (U, V) \) between \( U \) and \( V \) is determined from the equation \( U \cdot V = (U)(V) \cos \alpha \).

Given two \( n \)-hyperplanes in \( E^{n+1} \) by their dihedral angle we shall mean the acute angle between their normals. An \( n \)-hedron \( n \) of whose faces \((n-1)\)-hedras) are at right angles is called a right \( n \)-hedron. Given an \( n \)-hedron \( T \) in \( E^{n+1} \), we define the area of \( T \) to be the \( n \)-dimensional volume of \( T \) in the standard manner.

3. Projections. We distinguish \( x_{n+1} \) and call it \( z \). Given an \( n \)-hedron \( T \) in \( E^{n+1} \), its projection on the \((x_1, x_2, \ldots, x_n)\) hyperplane need not be an \( n \)-hedron. This occurs, for instance, if \( T \) is orthogonal to the hyperplane. We assume here that \( T \) lies on an hyperplane \( H: z = c + m_1x_1 + \cdots + m_nx_n \). Then, the projection of \( T \) on the \((x_1, x_2, \ldots, x_n)\) hyperplane, or \( \text{Proj} \ T \), is also an \( n \)-hedron. If \( \alpha \) is the dihedral angle between the hyperplane \( H \) determined by \( T \) and the \((x_1, x_2, \ldots, x_n)\) hyperplane, \( A \) is the area of \( T \), and \( A' \) is the area of \( \text{Proj} \ T \), then \( A = A' \sec \alpha \), where

\[
\sec \alpha = (1 + m_1^2 + \cdots + m_n^2)^{1/2}.
\]

4. Surfaces in \( E^{n+1} \). Let \( E \) be an open and connected set in the \((x_1, x_2, \ldots, x_n)\) hyperplane such that its closure \( \overline{E} \) is capable of being decomposed as the union of \( n \)-hedra in the natural manner. We say that \( \overline{E} \) is polyhedral. Let \( f \) be a real-valued function defined and continuous on \( \overline{E} \). The locus in \( E^{n+1} \) of the points \((x_1, x_2, \ldots, x_n, z)\), where \( z = f(x_1, x_2, \ldots, x_n) \), a function having \( \overline{E} \) for domain, is called an \( n \) surface in \( E^{n+1} \), or more briefly a surface. We wish to give a definition of the area of this surface in the case where \( f \) is continuously partially differentiable on \( \overline{E} \). We refer to such a surface as a continuously partially differentiable surface.

Let \( \Gamma: x_1 = F_1(t), \ldots, x_n = F_n(t), a \leq t \leq b \), be any parametric curve in \( \overline{E} \) passing through \((x_1^0, \ldots, x_n^0)\) for \( t = t_0 \). Then its image on \( S \), or

\[
C: x_1 = F_1(t), \ldots, x_n = F_n(t), z = f[F_1(t), \ldots, F_n(t)],
\]

\( a \leq t \leq b \), is a curve on \( C \) passing through \( Q \) for \( t = t_0 \). Assuming that each \( dx_i/dt \) exists and is continuous on \([a, b]\), it follows that there exists a tangent vector \( v \) to \( \Gamma \) at \((x_1^0, \ldots, x_n^0)\), and a tangent
Vector $V$ to $C$ at $Q$. If $\Gamma_1, \cdots, \Gamma_n$ are $n$ such curves in $E$, if $C_1, \cdots, C_n$ are the corresponding curves on $S$, and if we have chosen the curves $\Gamma$ in such a way that the $n$ vectors $v_1, \cdots, v_n$ are linearly independent, then the corresponding $n$ tangent vectors $V_1, \cdots, V_n$ determine an $n$-hyperplane $H$ in $E^{n+1}$.

One shows that for all such sets of curves in $S$, the corresponding $n$-hyperplane is unique. We refer to its normal line as the normal to $S$ at $Q$.

If $T$ is an $n$-hedron all of whose vertices are in $S$, we say that $T$ is inscribed in $S$. By $D(T)$, the deviation of $S$ on $T$, we mean the supremum of the set of the acute angles between the normal to $T$ and the normals to the portion of $S$ which is subtended by $T$ (i.e., the portion of $S$ whose projection on the $x_1, x_2, \cdots, x_n$ hyperplane is identical to that of $T$).

Let $\{P_1, P_2, \cdots, P_m\}$ be the vertices of a decomposition of $E$ into a finite set of $n$-hedra. For each $i$, let $Q_i = f(P_i)$. The set

$$\{Q_1, Q_2, \cdots, Q_m\}$$

determines a polyhedron which is inscribed on $S$. This polyhedron is composed of a finite set of $n$-hedra which are inscribed on $S$. By the norm of such a polyhedron we mean the largest of the diameters of its faces. By the deviation norm of the polyhedron we mean the largest of the deviations on its faces. By the area of this polyhedron we mean the sum of the areas of the $n$-hedra which compose it. We refer to these $n$-hedra as the faces of the polyhedron.

5. The geometric basis. We make use of the following additional properties of $E^{n+1}$.

(a) Let $U_1, U_2, \cdots, U_n$ be any $n$ vectors in $E^{n+1}$, such that $|\cos(U_i, U_j)| < k$, $0 < k < 1$, for every $i \neq j$. Then for each $\varepsilon > 0$ there exists $\delta > 0$ such that if $U_1', U_2', \cdots, U_n'$ are any $n$ vector such that $|\sin(U_i, U_i')| < \delta$, for each $i$, then

$$|\sin(U_1 \times U_2 \times \cdots U_n, U_1' \times U_2' \times \cdots U_n')| < \varepsilon.$$

(b) Let $P \in E$. Let $U$ be any vector in the $x_1, x_2, \cdots, x_n$ plane. We define the directional derivative of $f$ in the direction $U$ in the standard manner.

Under the hypothesis that $f$ is continuously partially differentiable on $E$, then the directional derivative of $f$ is uniformly continuous on $E$, i.e., for each $\varepsilon > 0$ there exists $\delta > 0$ such that if $(x_1', x_2', \cdots, x_n')$ and $(x_1'', x_2'', \cdots, x_n'')$ are in $E$ and $\rho((x_1', x_2', \cdots, x_n'), (x_1'', x_2'', \cdots, x_n'')) < \delta$, then $|f(x_1', x_2', \cdots, x_n') - f(x_1'', x_2'', \cdots, x_n'')| < \varepsilon$.
then the absolute value of the difference between the directional derivatives at \((x', x'', \ldots, x_n')\) and \((x', x'', \ldots, x_n'')\) in the direction of the vector from the first point to the latter, is less than \(\varepsilon\).

The directional derivative is uniformly Lipschitzian on \(\bar{E}\).

(c) There exist positive numbers \(k\) and \(\delta\), \(k < 1\), such that if \(P, P_1\) and \(P_2\) are any three distinct points in \(\bar{E}\) such that

1. \(\rho(P, P_1) < \delta\)
2. \(\rho(P, P_2) < \delta\)
3. \(\cos(\overrightarrow{PP_1}, \overrightarrow{PP_2}) = 0\)

then \(\cos(\overrightarrow{QQ_1}, \overrightarrow{QQ_2}) < k\), where \(Q = f(P), Q_1 = f(P_1), Q_2 = f(P_2)\).

(d) Let \(P, P_1\) be any two distinct points in \(\bar{E}\). \(Q_1 = f(P_1)\) and \(Q_2 = f(P_2)\). Let \(\overrightarrow{P_1P_2}\) be the linear interval determined by \(P_1\) and \(P_2\) and \(\overrightarrow{Q_1Q_2}\) be the linear interval determined by \(Q_1\) and \(Q_2\). Let the curve \(C = f(\overrightarrow{P_1P_2})\). Then there exists a point \(R\) in \(C\) such that the tangent line to \(C\) at \(R\) is parallel to \(\overrightarrow{Q_1Q_2}\).

(e) With the notation as in (d), let the deviation \(D(P_1, P_2)\) denote the supremum of the acute angles \(\phi\) between \(\overrightarrow{Q_1Q_2}\) and any tangent line to \(C\). Then for every \(\varepsilon > 0\), there exists \(\delta > 0\) such that if \(0 < \rho(P_1, P_2) < \delta\), then \(D(P_1, P_2) < \varepsilon\).

(f) For each \(\varepsilon > 0\) there exists \(\delta > 0\) such that if \(P_1\) and \(P_2\) are any two distinct points of \(\bar{E}\) such that \(\rho(P_1, P_2) < \delta\), then \(\psi < \varepsilon\), where \(\psi\) is the acute angle between the normals to \(S\) at \(f(P_1)\) and \(f(P_2)\).

(g) If \(\bar{E}\) is polyhedral, it can be decomposed into a set of \(n\)-hedra each of which is a right \(n\)-hedron. Moreover, for each real number \(r\), there exists a decomposition of \(\bar{E}\) into a set of right \(n\)-hedra, the diameter of each of which is less than \(r\).

We now proceed to the main theory.

We consider infinite sequences of polyhedra inscribed on \(S\). A sequence \((\Pi_1, \Pi_2, \cdots)\) of such polyhedra is said to be a proper sequence if the corresponding sequence \((N_1, N_2, \cdots)\) of norms and the corresponding sequence \((\phi_1, \phi_2, \cdots)\) of deviation norms both converge to zero.

We now give our definition of the area of the surface \(S = f(\bar{E})\), where \(f\) is continuously partially differentiable on \(\bar{E}\). If to every proper sequence \((\Pi_1, \Pi_2, \cdots)\) of polyhedra inscribed on \(S\) the corresponding sequence \((A_1, A_2, \cdots)\) of the polyhedral areas converges, then we say that \(S\) is quadrable and refer to the necessarily unique limit of \((A_1, A_2, \cdots)\) as the area of the surface \(S\).

**Theorem 1.** Let \(f(x_1, x_2, \ldots, x_n)\) be defined and be continuously partially differentiable on \(\bar{E}\). Then there exists a proper sequence \((\Pi_1, \Pi_2, \cdots)\) of polyhedra inscribed on \(S\).
Proof. For every positive number \( r \), there exists a decomposition of \( \overline{E} \) into a finite set of right \( n \)-hedra whose diameters are all less than \( r \). Their vertices determine a finite set of points in \( S \) whose projection is precisely the set of these vertices. This set of points in \( S \) determines a polyhedron \( \Pi \) which is inscribed on \( S \). We show that by making the norm of the decomposition of \( \overline{E} \) sufficiently small, we can make the deviation norm of \( \Pi \) arbitrarily small.

Let \( \varepsilon > 0 \) be given.

By property (g) there exists a decomposition of \( \overline{E} \) into a set of right \( n \)-hedra the diameter of each of which is arbitrarily small. By property (c) there exist real numbers \( k, \delta_1, k < 1 \), such that if \( PP_1 P_2 \cdots P_n \) is a right \( n \)-hedron in \( \overline{E} \) (with \( P \) the right angled vertex) of diameter \( < \delta_1 \), then \( |\cos(QQ_i, QQ_j)| < k \) for \( i \neq j \), where \( Q_i = f(P_i) \) and \( Q_j = f(P_j) \). Let the decomposition of \( \overline{E} \) be by right \( n \)-hedra each of diameter less than \( \delta_1 \).

By property (a), there exists a positive number \( \theta \) such that if \( \sin(QQ_i, QQ_i') < \theta \) for each \( i \), then the acute angle between

\[
QQ_1 \times QQ_2 \times \cdots \times QQ_n
\]

and \( QQ'_1 \times QQ'_2 \times \cdots \times QQ'_n \) is less than \( \varepsilon/3 \).

By properties (d) and (e) there exists a positive number \( \delta_2 \) such that if \( PP_1 P_2 \cdots P_n \) is a right \( n \)-hedron of diameter less than \( \delta_2 \), then, for each \( i \), the acute angle between the chord \( QQ_i \) and the tangent line at \( Q \) to the curve in \( S \) subtended by \( QQ_i \) is less than \( \theta \). It follows that the acute angle between the normal to the polyhedral face \( QQ_1 Q_2 \cdots Q_n \) and the surface normal at \( Q \) is less than \( \varepsilon/3 \).

By property (f) there exists a positive number \( \delta_3 \) such that if the diameter of the \( n \)-hedron \( PP_1 \cdots P_n \) is less than \( \delta_3 \), then the angle between the surface normals at any two points of the portion of \( S \) which is subtended by the polyhedral face \( QQ_1 \cdots Q_n \) is less than \( \varepsilon/3 \).

Let \( \delta \) be the least of \( \delta_1, \delta_2, \delta_3 \). If \( D \) is any decomposition of \( \overline{E} \) into right \( n \)-hedra each of diameter less than \( \delta \), then if \( QQ_1 \cdots Q_n \) is any of the polyhedral faces, the supremum of the angles between the normal to the \( n \)-hedron \( QQ_1 \cdots Q_n \) and the portion of \( S \) which is subtended by \( QQ_1 \cdots Q_n \) is less than \( \varepsilon \).

Thus, corresponding to a sequence \( (\varepsilon_1, \varepsilon_2, \cdots) \) converging to zero, there exists a sequence of polyhedra with corresponding sequence of norms converging to zero and also with corresponding sequence of deviation norms converging to zero.

**Theorem 2.** Let \( f(x_1, x_2, \cdots, x_n) \) be defined and continuously
partially differentiable on \( E \). Then, for every proper sequence \((Π₁, Π₂, \cdots)\) of \( n \)-hedra inscribed on \( S = f(E) \), the corresponding sequence \((A₁, A₂, \cdots)\) of polyhedral areas converges to the multiple integral
\[
\int_E (1 + z₁² + \cdots + zₙ²)^{1/2} \, dx₁ \, dx₂ \cdots dxₙ.
\]

Proof. For the sake of notations, let \( z_{x,h} \) denote \( \partial z/\partial x_h \), \( h = 1, \cdots, n \). For each \( m \), the projection of \( Π_m \) constitutes a decomposition \( D_m \) of \( E \) into a finite set of \( n \)-hedra. Let the \( n \)-hedron \( Δ_m = QQₖ \cdots Qₙ \) be a face of \( Π_m \) and let \( Δ'_m = \text{Proj}(QQₖ \cdots Qₙ) = PPₖ \cdots Pₙ \). Let \( β_{mr} \) be the acute angle between the normals to \( Δ_m \) and \( Δ'_m \). Then \( A_m = A'_m \sec β_{mr} \) where \( β_{mr} \) is the angle between the \( z \)-axis and the normal to \( Δ_m \). The area of \( A_m \) of \( Π_m \) is given by \( \sum r A'_r \sec β_{mr} \).

Let \( P_{mr} \) be any point in \( Δ'_m \) and let \( Q_{mr} \) be the point of \( S \) whose projection is \( P_{mr} \). Let \( θ_{mr} \) denote the acute angle between the surface normal at \( Q_{mr} \) and the \( z \)-axis.

We associate to the sequence \((Π₁, Π₂, \cdots)\) certain related sequences:
\[
(Π₁, Π₂, \cdots),
(φ₁, φ₂, \cdots),
(Σ₁, Σ₂, \cdots),
(Σ', Σ'', \cdots).
\]

The sequence \((φ₁, φ₂, \cdots)\) is the corresponding sequence of deviation norms. The sequence \((Σ₁, Σ₂, \cdots)\) is the corresponding sequence of polyhedral areas, \( Σ_m = \sum r A'_r \sec β_{mr} \). In the fourth sequence, \( Σ'_m = \sum r A'_r \sec θ_{mr} \). Here, \( sec θ_{mr} \) is the value of \( (1 + z₁² + \cdots + zₙ²)^{1/2} \) at some point of \( Δ'_m \). Thus the sequence \((Σ', Σ'', \cdots)\) is a sequence of Riemann sum of the function \((1 + z₁² + \cdots + zₙ²)^{1/2} \) on \( E \) with corresponding sequence of norms converging to zero. Since
\[
(1 + z₁² + \cdots + zₙ²)^{1/2}
\]

is continuous on \( E \), the sequence \((Σ', Σ'', \cdots)\) converges to the multiple integral
\[
\int_E (1 + z₁² + \cdots + zₙ²)^{1/2} \, dx₁ \, dx₂ \cdots dxₙ.
\]

We now consider the sequence \((Σ₁, Σ₂, \cdots)\). Let \( θ \) denote the acute angle between the surface normal at a point in \( S \) and the \( z \)-axis. \( sec θ = (1 + z₁² + \cdots + zₙ²)^{1/2} \) is bounded on \( E \). Thus there exists an angle \( θ^* > 0 \) such that \( θ < θ^* \) for all points of \( E \) (i.e., for all points of \( S \)). Since \( sec θ \) is uniformly continuous on the closed
interval \([\theta, \theta^*]\) for every \(\eta > 0\) there exists \(\tau > 0\) such that if \(0 < \theta_1 < \theta^*, \theta < \theta_2 < \theta^*\) and \(|\theta_1 - \tau_2| < \tau\), then \(|\sec \theta_1 - \sec \theta_2| < \eta\).

We now compare the sequences \((\Sigma, \Sigma_2, \cdots)\) and \((\Sigma', \Sigma'_2, \cdots)\).

Let \(\varepsilon > 0\) be given. Take \(\varepsilon/2V\) where \(V\) is the volume (area) of \(E\). There exists \(\tau > 0\) such that if \(|\theta_1 - \theta_2| < \tau\), then

\[|\sec \theta_1 - \sec \theta_2| < \frac{\varepsilon}{2V}.
\]

Since \((\phi_1, \phi_2, \cdots)\) converges to zero, there exists a positive integer \(N_1\) such that if \(m > N_1\) then \(\phi_m < \tau\). Thus if \(m > N_1\), then

\[|\Sigma_m - \Sigma'_m| = |\Sigma_\tau \alpha^\tau_m (sec \beta_{\tau m} - \sec \theta_{\tau m})| < \frac{\varepsilon}{2V} \Sigma A^\tau_m = \frac{\varepsilon}{2}.
\]

Since \((\Sigma', \Sigma'_2, \cdots)\) converges to \(\int\), there exists a positive integer \(N_2\) such that if \(m > N_2\) then \(|\Sigma'_m - \int| < \varepsilon/2\). Let \(N\) be the larger of \(N_1\) and \(N_2\). If \(m > N\), then

\[|\Sigma_m - \int| = |\Sigma_m - \Sigma'_m + \Sigma'_m - \int| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\]

Thus, \((\Sigma, \Sigma_2, \cdots)\) converges to \(\int\).

REFERENCES


Received February 5, 1970.

MARIST COLLEGE