Linearization and Scaling Factors for change of variables in multiple integrals

John McCuan

November 12, 2019

If $\Psi : \mathcal{W} \to \mathcal{V}$ is a valid change of variables from an object $\mathcal{W}$ onto an object $\mathcal{V}$, and these are objects on which one can integrate, then we can change variables to express an integral over $\mathcal{V}$ as an equivalent integral over $\mathcal{W}$. More precisely, if $f : \mathcal{V} \to \mathbb{R}$ is a real valued function on $\mathcal{V}$ for which the Riemann integral

$$\int_{\mathcal{V}} f = \lim_{\|\mathcal{P}\| \to 0} \sum_j f(\mathbf{x}_j^*) \operatorname{meas}(\mathcal{V}_j) \tag{1}$$

is well-defined, then we have the change of variables formula

$$\int_{\mathcal{V}} f = \int_{\mathcal{W}} (f \circ \Psi)\, \sigma$$

where σ is an appropriate **scaling factor**. It is our objective here to explain the nature and origin of the scaling factor in terms of linearization of the transformation Ψ. We will assume (and use) intuition concerning Riemann sum approximation/definition of integrals as in (1), though that topic will also be briefly discussed. The concept of change of variables will be presented in the abstract context[1] of integration on "objects" $\mathcal{V}$ and $\mathcal{W}$, though we will use specific examples for the details concerning the scaling factor and, as such, the discussion is not entirely general. We should like to include the following cases at least:

1. $\mathcal{V}$ and $\mathcal{W}$ are regions in $\mathbb{R}^n$ and Ψ is a diffeomorphism.

2. $\mathcal{V} = \Gamma$ is a **curve** in $\mathbb{R}^n$ and $\Psi^{-1} : [a, b] \to \mathbb{R}^n$ is a parameterization of that curve on an interval $\mathcal{W} = [a, b]$.

[1] See my discussion of integration on objects if this is not familiar.

3. $\mathcal{V} = \mathcal{S}$ is a **surface** in $\mathbb{R}^n$ and $\Psi^{-1} : \mathcal{U} \to \mathbb{R}^n$ is a parameterization of that surface on a domain $\mathcal{W} = \mathcal{U}$ in $\mathbb{R}^2$.

Perhaps the first point to recall/understand is that the measure of a region/object $\mathcal{V}$ can be represented as the sum of the measures of many small regions in a partition $\{\mathcal{V}_j\}$ of $\mathcal{V}$. In particular, we can often take these **partition pieces** $\mathcal{V}_j$ to have some standard form. We know the integral

$$\int_{\mathcal{V}} f$$

is approximated by a Riemann sum:

$$\int_{\mathcal{V}} f \sim \sum_j f(\mathbf{x}_j^*)\ \mathrm{meas}(\mathcal{V}_j).$$

Taking the special case in which $f \equiv 1$ is a constant function, we get

$$\mathrm{meas}(\mathcal{V}) \sim \sum_j \mathrm{meas}(\mathcal{V}_j).$$

In the case of a change of variables we can also write

$$\mathrm{meas}(\mathcal{V}) \sim \sum_j \mathrm{meas}(\Psi(\mathcal{W}_j)) \tag{2}$$

where $\mathcal{W}_j = \Psi^{-1}(\mathcal{V}_j)$. Now, it seems like we've started with the partition pieces $\mathcal{V}_j$, but there is no reason we can't assume we've chosen a partition $\{\mathcal{W}_j\}$ of $\mathcal{W}$ first and then obtained the partition $\{\mathcal{V}_j\}$ of $\mathcal{V}$ consisting of the **image pieces** $\mathcal{V}_j = \Psi(\mathcal{W}_j) = \{\Psi(w) : w \in \mathcal{W}_j\}$. From this point of view, it is crucial to understand the measure of $\Psi(\mathcal{W}_j)$, at least when $\mathcal{W}_j$ is a piece of $\mathcal{W}$ with a particularly simple form. We return to this point below; see (3).

Let's illustrate what we're saying here with a specific example which is found on page 957 of the Thomas Calculus text and in problem 15.8.15 of the same section. In these problems the domain

$$\mathcal{U} = \{(x, y) : 1/y < x < y,\ 1 < y < 2\}$$

is transformed to the domain

$$\mathcal{W} = \{(u, v) : 1 < u < 2,\ 1 < v < 2/u\}$$

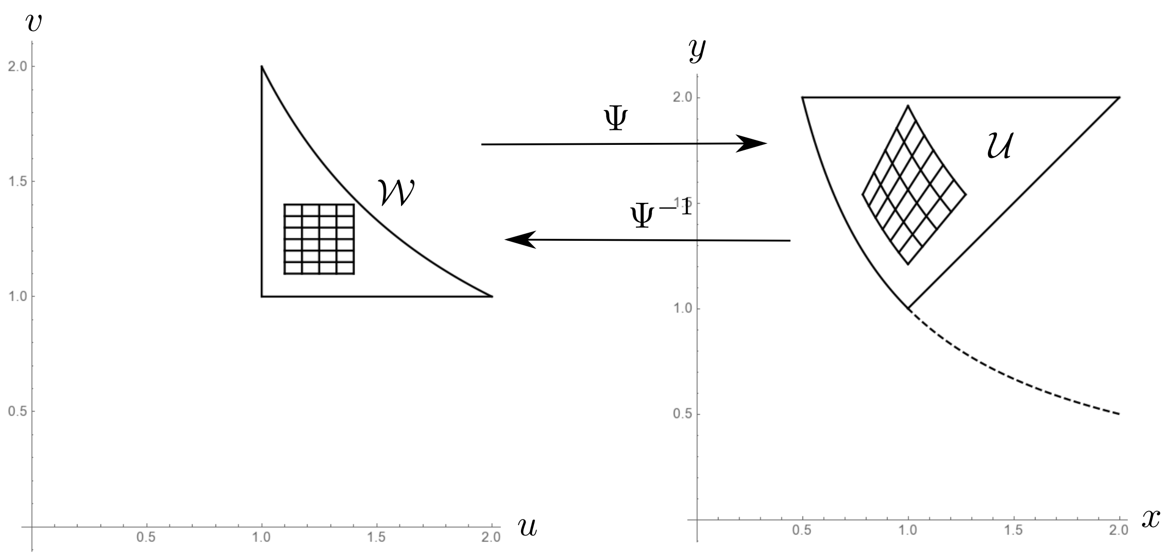

Figure 1: transformation of domains

by the transformation

$$\Psi^{-1}(x, y) = \left(\sqrt{xy}, \sqrt{\frac{y}{x}}\right).$$

This transformation of domains involving the mapping $\Psi : \mathcal{W} \to \mathcal{U}$ by $\Psi(u, v) = (u/v, uv)$ is indicated in Figure 1. Given a real valued function $f : \mathcal{U} \to \mathbb{R}$ with $f = f(x, y)$, the integral over $\mathcal{U}$ of f is approximated by a Riemann sum:

$$\int_{\mathcal{U}} f \sim \sum_{i,j} f(x_i^*, y_j^*)\,\text{area}(\Psi(\mathcal{W}_{ij})).$$

Here, we have taken a partition of $\mathcal{W}$ consisting of rectangular regions

$$\mathcal{W}_{ij} = [u_i, u_{i+1}] \times [v_j, v_{j+1}] = \{(u, v) : u_i \le u \le u_{i+1},\ v_j \le v \le v_{j+1}\}$$

some of which are shown in Figure 1.

The relation between the area of the entire domain $\mathcal{U}$ and the area of $\mathcal{W}$ is not very obvious. The global transformation is somewhat complicated. If we take the rectangular subregion which is the union of all the rectangles in $\mathcal{W}$ indicated in the figure, then the image looks roughly like a parallelogram, but it is clearly not exactly a parallelogram. Not only is the image shape not exactly a parallelogram, but if we

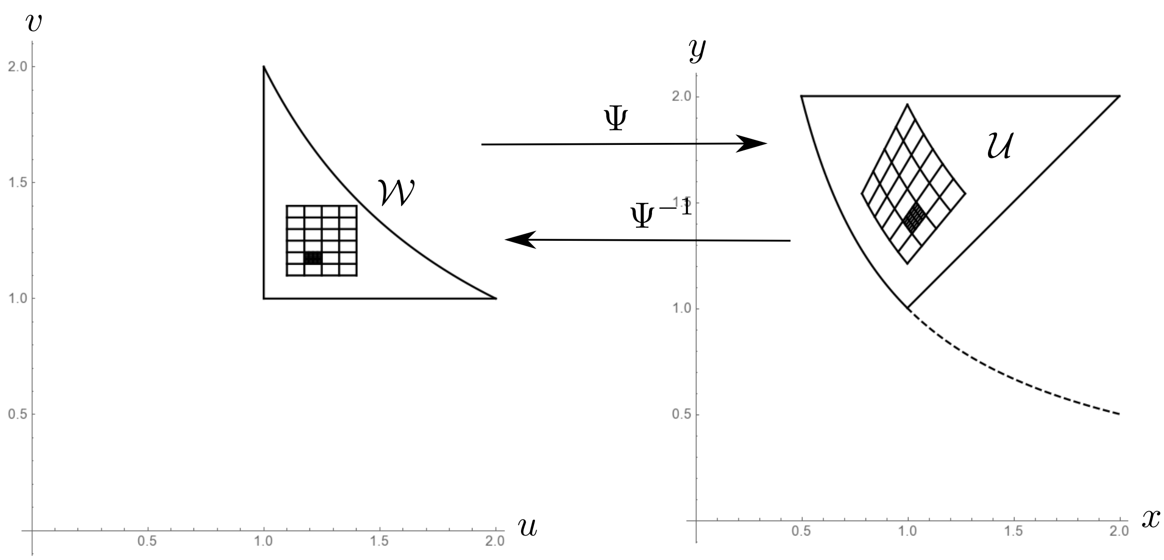

Figure 2: focusing on one (shaded) rectangle

filled out the domain $\mathcal{W}$ with more rectangles, notice that their union would no longer be a rectangle (the union would be closer to the shape of $\mathcal{W}$), so the image of the union would be more complicated than a parallelogram in more than one way.

Nevertheless if instead of looking at what happens more globally with the mapping, we look more locally, say at one of the rectangles partitioning $\mathcal{W}$ as indicated in Figure 2, then the image is even closer to a parallelogram.

In figure 3 we have zoomed in on the shaded rectangle and its image. We have denoted the corner of this rectangle $\mathcal{W}_{ij}$ by (u_i, v_j) and the image of this corner in the x, y-plane by $(x_i, y_j) = \Psi(u_j, v_j)$. This series of figures is intended to suggest that the image of a very small rectangle $\mathcal{W}_{ij}$ is very close to a parallelogram. In particular, the area $\text{area}(\Psi(\mathcal{W}_{ij}))$ appearing in the Riemann sum

$$\int_{\mathcal{U}} f \sim \sum_{i,j} f(x_i^*, y_j^*)\,\text{area}(\mathcal{U}_{ij}) =\sim \sum_{i,j} f \circ \Psi(u_i^*, v_j^*)\,\text{area}(\Psi(\mathcal{W}_{ij}))$$

is approximated in the limit by the area of **some parallelogram**. Thus, the big question is: What is this parallelogram, and what is its area? That is,

$$\text{area}(\Psi(\mathcal{W}_{ij})) \sim \;?? \tag{3}$$

The next key ideas are that

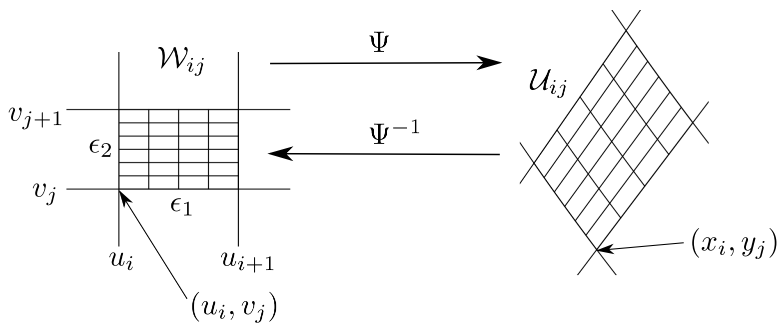

Figure 3: zooming in on the (shaded) rectangle

1. The image $\Psi(\mathcal{W}_{ij})$ is determined by the behavior of the mapping Ψ (locally) near the point (u_i, v_j), and
2. The local behavior of the mapping Ψ near a point (u_i, v_j) is given by the **linearization** (or first order approximation) of Ψ.

The **first order approximation formula** for a mapping Ψ at a point (u_i, v_j) is given by

$$\Psi(u, v) \sim A(u, v) = \Psi(u_i, v_j) + D\Psi(u_i, v_j) \begin{pmatrix} u - u_i \\ v - v_j \end{pmatrix} \tag{4}$$

where $D\Psi$ is the **matrix** of partial derivatives of the components of $\Psi = (\Psi_1, \Psi_2)$:

$$D\Psi = \begin{pmatrix} \dfrac{\partial \Psi_1}{\partial u} & \dfrac{\partial \Psi_1}{\partial v} \\ \\ \dfrac{\partial \Psi_2}{\partial u} & \dfrac{\partial \Psi_2}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{v} & -\dfrac{u}{v^2} \\ \\ v & u \end{pmatrix}. \tag{5}$$

We use the letter A for this map because it is an **affine** map. The affine map differs from a linear map by translations in both the independent variable and the dependent variable. More precisely, if $L : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear map given by

$$L(\xi, \eta) = D\Psi(u_i, v_j) \begin{pmatrix} \xi \\ \eta \end{pmatrix},$$

then

$$A(u, v) = (x_i, y_j) + L[(u, v) - (u_i, v_i)].$$

Technically, the matrix multiplication applied to column vectors we have used in (4) requires some transposes in these formulas at various points, but we have ignored this minor abuse of notation.

Now, let us assume a uniform rectangular grid in $\mathcal{W}$ with $u_{i+1} - u_i = \epsilon_1$ and $v_{j+1} - v_j = \epsilon_2$, so the area of $\mathcal{W}_{ij}$ is $\epsilon_1\epsilon_2$ as indicated in Figure 3. Denoting the columns of $D\Psi(u_i, u_j)$ by $\mathbf{v}$ and $\mathbf{w}$, we see $A(\mathcal{W}_{ij})$ is indeed a parallelogram. This may be seen precisely as follows. The rectangle

$$\mathcal{W}_{ij} = \{(u_i, v_j) + s\epsilon_1\mathbf{e}_1 + t\epsilon_2\mathbf{e}_2 : 0 \le s, t \le 1\}$$

where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ as usual. Therefore,

$$\begin{aligned} A(\mathcal{W}_{ij}) &= \{(x_i, y_j) + s\epsilon_1 L(\mathbf{e}_1) + t\epsilon_2 L(\mathbf{e}_2) : 0 \le s, t \le 1\} \\ &= \{(x_i, y_j) + s\epsilon_1\mathbf{v} + t\epsilon_2\mathbf{w} : 0 \le s, t \le 1\}. \end{aligned}$$

Notice the affine translations make no difference in the shape of the image $A(\mathcal{W}_{ij})$. In summary

$$\operatorname{area}(\Psi(\mathcal{W}_{ij})) \sim \operatorname{area}(L([0, \epsilon_1] \times [0, \epsilon_2]))$$

and $L([0, \epsilon_1] \times [0, \epsilon_2]$ is a parallelogram spanned by the vectors $\epsilon_1\mathbf{v}$ and $\epsilon_2\mathbf{w}$. This parallelogram has area $\epsilon_1\epsilon_2$ times the area of the parallelogram spanned by the vectors $\mathbf{v}$ and $\mathbf{w}$ since

$$\{s\epsilon_1\mathbf{v} + t\epsilon_2\mathbf{w} : 0 \le s, t \le 1\} = \{s\mathbf{v} + t\mathbf{w} : 0 \le s \le \epsilon_1,\ 0 \le t \le \epsilon_2\}.$$

Finally, then, we see the origin of the scaling factor. The area of the parallelogram

$$\{s\mathbf{v} + t\mathbf{w} : 0 \le s, t \le 1\}$$

spanned by the vectors

$$\mathbf{v} = \begin{pmatrix} \dfrac{\partial\Psi_1}{\partial u} \\ \\ \dfrac{\partial\Psi_2}{\partial u} \end{pmatrix} \qquad \text{and} \qquad \mathbf{w} = \begin{pmatrix} \dfrac{\partial\Psi_1}{\partial v} \\ \\ \dfrac{\partial\Psi_2}{\partial v} \end{pmatrix}$$

is the absolute value of the determinant of the matrix $D\Psi$ evaluated at (u_i, v_j).

Before we review (and generalize) this line of reasoning as applied to a transformation $\Psi : \mathcal{W} \to \mathcal{V}$ where $\mathcal{W}$ and $\mathcal{V}$ are domains in $\mathbb{R}^n$, let us give some more focused attention of the area approximation process in our example.

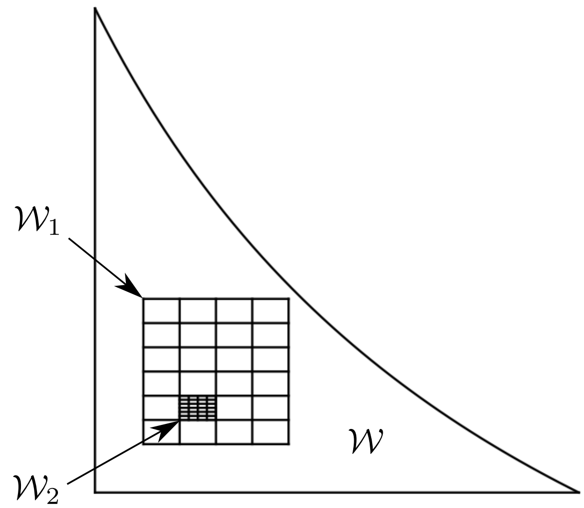

Figure 4: two specific rectangles in $\mathcal{W}$

Numerical Area Approximation

Let us denote by $\mathcal{W}_1$ the first large rectangle (composed of 24 smaller rectangles and appearing on the left in Figures 1 and 2. This rectangle (actually a square) has lower left corner $(u, v) = (1.1, 1.1)$ and is shown as it sits in the domain $\mathcal{W}$ in Figure 4. The entire rectangle may be written as

$$\mathcal{W}_1 = [1.1, 1.4] \times [1.1, 1.4] = \{(1.1, 1.1) + 0.3s(1, 0) + 0.3t(0, 1) : 0 \le s, t \le 1\}.$$

Similarly, we set

$$\begin{aligned} \mathcal{W}_2 &= [1.175, 1.25] \times [1.15, 1.2] \\ &= \{(1.175, 1.15) + 0.075s(1, 0) + 0.05t(0, 1) : 0 \le s, t \le 1\}. \end{aligned}$$

This is the (shaded) rectangle in Figure 4 with lower left corner $(1.175, 1.15)$ and side lengths $\epsilon_1 = 0.075$ and $\epsilon_2 = 0.05$. It also appears on the left in Figures 1, 2, and 3.

The linearization (i.e., first order affine approximation) of Ψ at the left corner of $\mathcal{W}_1$ is

$$A_1(u, v) = (1, 1.21) + (0.9\overline{09}(u - 1.1) - 0.9\overline{09}(v - 1.1), 1.1(u - 1.1) + 1.1(v - 1.1))$$

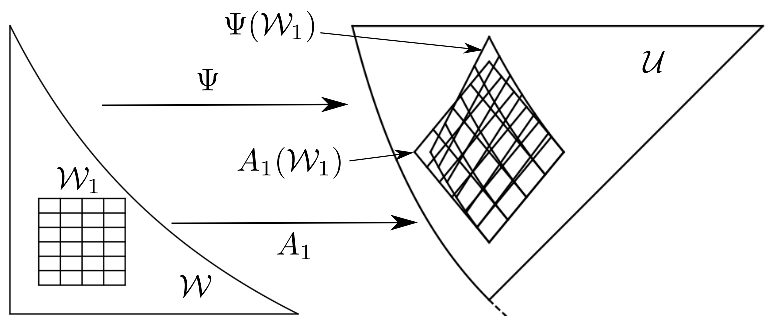

Figure 5: the affine mapping of $\mathcal{W}_1$

where $0.9\overline{09} = 1/1.1 = 1.1/(1.1)^2 = 0.9090909\ldots$. The image of the square $\mathcal{W}_1$ under this map may be compared to the image under Ψ in Figure 5. The image under Ψ is also shown on the right in Figures 1 and 2.

Now we turn to the smaller rectangle $\mathcal{W}_2$. The affine approximation of Ψ at the lower left corner $(1.175, 1.15)$ of $\mathcal{W}_2$ is

$$\begin{aligned} A_2(u,v) \sim\ & (1.0217, 1.351) \\ & + (0.870(u-1.175) - 0.888(v-1.15),\ 1.15(u-1.175) + 1.175(v-1.15)). \end{aligned}$$

We have approximated some of the coefficients to three decimal places because the exact expressions would be cumbersome to express. The image of $\mathcal{W}_2$ under A_2 and Ψ is illustrated in Figure 6. As in Figure 3 we have zoomed in and the "shading" is clearly seen to be another subdivision of $\mathcal{W}_2$ into 24 subrectangles. The image of $\mathcal{W}_2$ under Ψ, which is not a parallelogram, is clearly seen to be very close to the parallelogram $A_2(\mathcal{W}_2)$.

Finally, we make this area comparison precise by using the change of variables with integrand $f \equiv 1$ mentioned in connection with (2). Note from (5) that the scaling factor for the mapping Ψ is

$$\sigma = \det D\Psi = \frac{2u}{v}.$$

Evaluating this expression at $(u_1, v_1) = (1.1, 1.1)$, we get $\sigma_1 = 2$. The image of the square $\mathcal{W}_1$ under the affine approximation map A_1 has area

$$\text{area}(A_1(\mathcal{W}_1)) = (0.3)(0.3)\det D\Psi(1.1, 1.1) = (0.3)(0.3)(2) = 0.18.$$

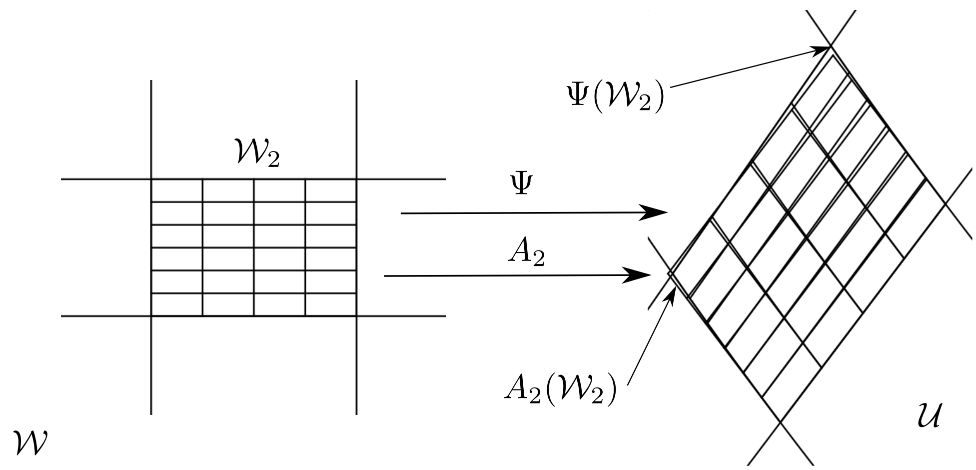

Figure 6: the affine mapping of $\mathcal{W}_2$

The area of $\Psi(\mathcal{W}_1)$ on the other hand, is computed exactly by the integral

$$\begin{aligned}
\operatorname{area}(\Psi(\mathcal{W}_1)) &= \int_{\mathcal{W}_1} \sigma \\
&= \int_{1.1}^{1.4} \int_{1.1}^{1.4} \frac{2u}{v}\, du dv \\
&= \int_{1.1}^{1.4} \frac{(1.4)^2 - (1.1)^2}{v}\, dv \\
&= [(1.4)^2 - (1.1)^2][\ln(1.4) - \ln(1.1)] \\
&\sim 0.180872.
\end{aligned}$$

Thus, the error in the affine approximation of the area at this scale is approximately 0.000872 or about 0.48 %.

Making the same calculations for $\mathcal{W}_2$, we find $\sigma_2 = 2(1.175)/(1.15) \sim 2.04348$, so

$$\operatorname{area}(A_2(\mathcal{W}_2)) = (0.075)(0.05)\sigma_2 \sim 0.00766304.$$

$$\begin{aligned}
\text{area}(\Psi(\mathcal{W}_2)) &= \int_{\mathcal{W}_2} \sigma \\
&= \int_{1.15}^{1.2} \int_{1.175}^{1.25} \frac{2u}{v}\, dudv \\
&= \int_{1.15}^{1.2} \frac{(1.25)^2 - (1.175)^2}{v}\, dv \\
&= [(1.25)^2 - (1.175)^2][\ln(1.2) - \ln(1.15)] \\
&\sim 0.00774053.
\end{aligned}$$

The absolute error is, in this case, approximately 0.0000774864 which is about 1 % of the total area. Well that's rather irritating. This discussion was intended to show that when you use smaller rectangles, the approximation improves. Instead, we used a smaller rectangle, and the approximation got worse. Maybe I made an error somewhere, but I don't see it. I am assuming that the image parallelogram just happens to be a better approximation of the nonlinear image of $\mathcal{W}_1$ than we get for $\mathcal{W}_2$. How irritating!

At least we're set up to pretty easily try some even smaller rectangles:

$$\text{area}(A_1([1.1, 1.1+\epsilon] \times [1.1, 1.1+\epsilon])) = 2\epsilon^2.$$

$$\begin{aligned}
\text{area}(\Psi([1.1, 1.1+\epsilon] \times [1.1, 1.1+\epsilon])) &= \int_{1.1}^{1.1+\epsilon} \int_{1.1}^{1.1+\epsilon} \frac{2u}{v}\, dudv \\
&= \int_{1.1}^{1.1+\epsilon} \frac{(1.1+\epsilon)^2 - (1.1)^2}{v}\, dv \\
&= [(1.1+\epsilon)^2 - (1.1)^2][\ln(1.1+\epsilon) - \ln(1.1)].
\end{aligned}$$

Thus, the percentage error is given by

$$\begin{aligned}
p_1 =& 100\, \frac{\text{area}(\Psi([1.1, 1.1+\epsilon] \times [1.1, 1.1+\epsilon])) - \text{area}(A_1([1.1, 1.1+\epsilon] \times [1.1, 1.1+\epsilon]))}{\text{area}(\Psi([1.1, 1.1+\epsilon] \times [1.1, 1.1+\epsilon])} \\
&= 100\, \frac{(2.2+\epsilon)\ln(1+\epsilon/1.1) - 2\epsilon}{(2.2+\epsilon)\ln(1+\epsilon/1.1)}.
\end{aligned}$$

You can see the plot of p_1 on the left in Figure 7. Notice the value $p_1(0.3) \sim 0.04$ corresponding to $\mathcal{W}_1$.

Similarly,

$$\text{area}(A_2([1.175, 1.175+\epsilon] \times [1.15, 1.15+2\epsilon/3])) = \frac{2\sigma_2\epsilon^2}{3} \sim 1.36232\epsilon^2.$$

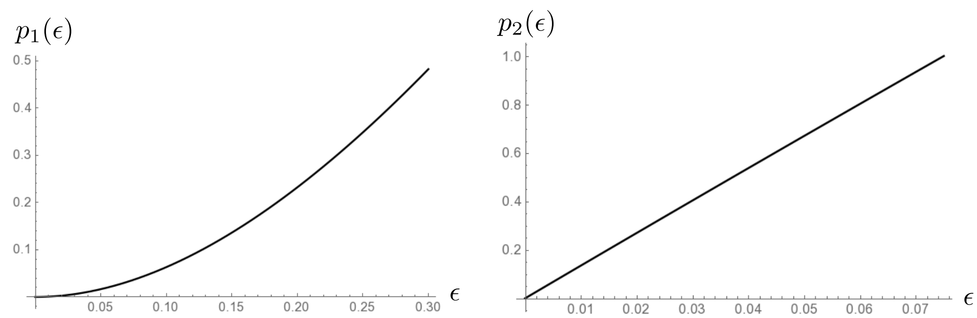

Figure 7: percentage error in affine approximation of area

and

$$
\begin{aligned}
&\text{area}(\Psi([1.175, 1.175+\epsilon] \times [1.15, 1.15+2\epsilon/3])) \\
&\qquad = \int_{1.15}^{1.15+2\epsilon/3} \int_{1.175}^{1.175+\epsilon} \frac{2u}{v}\, du dv \\
&\qquad = \int_{1.15}^{1.15+2\epsilon/3} \frac{(1.175+\epsilon)^2 - (1.175)^2}{v}\, dv \\
&\qquad = [(1.175+\epsilon)^2 - (1.175)^2][\ln(1.15+2\epsilon/3) - \ln(1.15)] \\
&\qquad = \epsilon(2.35+\epsilon)\ln(1+2\epsilon/3.45).
\end{aligned}
$$

Thus, the percentage error for rectangles like $\mathcal{W}_2$ is given by

$$
p_2 \sim 100\, \frac{(2.35+\epsilon)\ln(1+2\epsilon/3.45) - 1.36232\epsilon}{(2.35+\epsilon)\ln(1+2\epsilon/3.45)}.
$$

We have plotted p_2 as a function of ϵ on the right in Figure 7 with the 1 % error associated with $\mathcal{W}_2$ given for $\epsilon = \epsilon_1 = 0.075$. In both cases, one sees the error diminishes with the size of the square or rectangle, but indeed our smaller rectangle $\mathcal{W}_2$ starts with a larger precentage error than the large squre $\mathcal{W}_1$. Mathematics/arithmetic is sometimes irritating like that!

Summary for General Dimensions

Let's say we have a rectangular parallelopiped Q in $\mathbb{R}^n$ with side lengths ϵ_1, ϵ_2, $\ldots \epsilon_n$ so that its area is $\epsilon_1\epsilon_2 \cdots \epsilon_n$. If $\Psi : \mathcal{W} \to \mathcal{V}$ is a change of variables between domains

$\mathcal{W} \subset \mathbb{R}^n$ and $\mathcal{V} \subset \mathbb{R}^n$, and our parallelopiped happens to lie in $\mathcal{W}$ with a point $\mathbf{p} \in Q \cap \mathcal{W}$, then the can consider the areas of the images

$$\text{area}(\Psi(Q)) \qquad \text{and} \qquad \text{area}(A(Q))$$

where A is the affine approximation of Ψ at $\mathbf{p}$ given by

$$A(\mathbf{x}) = \Psi(\mathbf{p}) + D\Psi(\mathbf{p})(\mathbf{x} - \mathbf{p}).$$

If the side lengths $\epsilon_1, \ldots \epsilon_n$ are small, then the entire set Q will be close to $\mathbf{p}$. Taylors formula says that for $\mathbf{x}$ close to $\mathbf{p}$ (but not equal to $\mathbf{p}$) the quantity

$$\frac{|\Psi(\mathbf{x}) - A(\mathbf{x})|}{|\mathbf{x} - \mathbf{p}|} \quad \text{is small.}$$

In fact, the limit of this quantity as $\mathbf{x}$ tends to $\mathbf{p}$ is zero. In particular, if the entire parallelopiped Q is close to $\mathbf{p}$, then every point $\mathbf{x} \in Q$ will map under A very close to where it maps under Ψ, so the images $A(Q)$ and $\Psi(Q)$ will be close. Thus, the term

$$f(\Psi(\mathbf{p}^*))\, \text{volume}(\Psi(Q))$$

in the Riemann sum for $\int_{\mathcal{V}} f$ is approximated by

$$\begin{aligned} f(\Psi(\mathbf{p}^*))\, \text{volume}(\Psi(Q)) &\sim f(\Psi(\mathbf{p}^*))\, \text{volume}(A(Q)) \\ &= f(\Psi(\mathbf{p}^*))\ |\det D\Psi(\mathbf{p}^*)|\, \text{volume}(Q) \\ &= f(\Psi(\mathbf{p}^*))\ \sigma(\mathbf{p}^*)\ \epsilon_1\epsilon_2 \cdots \epsilon_n. \end{aligned} \tag{6}$$

This is precisely, the term one expects in the Riemann sum for $\int_{\mathcal{W}}(f \circ \Psi)\ \sigma$.

Exercise 1 *Show the formula given in (6) holds for a general set $Q \subset \mathcal{W}$ (not just a parallelopiped) in the sense that*

$$\text{volume}(\Psi(Q)) \sim \text{volume}(A(Q)) = |\det D\Psi(\mathbf{p})|$$

when $A(\mathbf{x}) = \Psi(\mathbf{p}) + D\Psi(\mathbf{p})(\mathbf{x} - \mathbf{p})$ and Q is a set with all its points close to $\mathbf{p}$.

Curves and Surfaces

The change of variables formula can be used to give computable expressions for integration on curves and surfaces, though the scaling factor takes a somewhat different form. When we have a curve $\Gamma \subset \mathbb{R}^n$, it is usually given to us using a parameterization. Sometimes it's not. For example, we can talk about the circle $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = a^2\}$ in $\mathbb{R}^2$ without a parameterization, but it's also easy to give a parameterization $\mathbf{r}(t) = (a\cos t, a\sin t)$. Let's say we start with a curve given by a parameterization $\mathbf{r} : [a, b] \to \mathbb{R}^n$ and using the arclength formula,

$$s = \int_a^t |\mathbf{r}'(\tau)|\, d\tau$$

we can also construct a parameterization $\gamma : [0, \ell] \to \mathbb{R}^2$ by $\gamma(s) = \mathbf{r}(t(s))$ where s is arclength. When we integrate a real valued function $f : \Gamma \to \mathbb{R}$ on the curve, we break it up into pieces Γ_j and consider a Riemann sum

$$\sum_j f(p_j^*)\, \mathrm{length}(\Gamma_j).$$

Since the derivative $\mathbf{r}'(t)$ gives the velocity vector and $|\mathbf{r}'(t)|$ is the speed, we can approximate $\mathrm{length}(\Gamma_j)$ by $|\mathbf{r}'(t_j^*)|(t_{j+1} - t_j)$ where

$$\mathbf{r}(t_j^*) = p_j^* \quad \text{and} \quad [t_j, t_{+1}j] \text{ is the interval in } [a, b] \text{ mapping to } \Gamma_j.$$

Using this approximation and our understanding of Riemann sums we see

$$\int_\Gamma f = \int_a^b f \circ \mathbf{r}(t)\, \sigma(t)\, dt$$

where $\sigma(t) = |\mathbf{r}'(t)|$. It's like we "changed variables" from Γ to the parameter interval $[a, b]$. (You should draw pictures of Γ and mappings/functions from $[a, b]$ into Γ and from Γ into $\mathbb{R}$ to illustrate this change of variables.)

Exercise 2 *Show integration on Γ is also given by integration with respect to arclength:*

$$\int_\Gamma f = \int_0^\ell f \circ \gamma(s)\, ds.$$

Surfaces are rather similar to curves. They ususally come via a parameterization $X : \mathcal{U} \to \mathbb{R}^n$ where $\mathcal{U}$ is a domain in $\mathbb{R}^2$. There are some nondegeneracy conditions required to make sure the mapping X actually produces a surface. We won't really get into those, but suffice it to say we want the vectors

$$X_u = \frac{\partial X}{\partial u} \qquad \text{and} \qquad X_v = \frac{\partial X}{\partial v}$$

to be linearly independent so that they determine and span a tangent plane to the image surface at each point—much the same way $\mathbf{r}'(t)$ spans a tangent line to the curve parameterized by $\mathbf{r} = \mathbf{r}(t)$ as long as $\mathbf{r}'(t) \neq 0$. In fact, not only do linearly independent vectors X_u and X_v span a tangent plane to the surface, they determine a parallelogram in $\mathbb{R}^n$ given by

$$P = \{X(u_0, v_0) + sX_u(u_0, v_0) + tX_v(u_0, v_0) : 0 \leq s, t \leq 1\}.$$

Notice this parallelogram P has one corner at the point $X(u_0, v_0)$ on the surface, and it is tangent to the surface. In fact, if we take a small square

$$Q = \{(u_0, v_0) + s\epsilon_1(1, 0) + t\epsilon_2(0, 1) : 0 \leq s, t \leq 1\} \subset \mathcal{U},$$

then the affine approximation of X given by

$$A(u, v) = X(u_0, v_0) + DX(u_0, v_0)(u - u_0, v - v_0)$$

(where DX, as usual, is the matrix of partial derivatives of the component functions of X) has

$$\begin{aligned} A(Q) &= \{A(u, v) : (u, v) \in Q\} \\ &= \{X(u_0, v_0) + s\epsilon_1 X_u(u_0, v_0) + t\epsilon_2 X_v(u_0, v_0) : 0 \leq s, t \leq 1\} \\ &= \{X(u_0, v_0) + sX_u(u_0, v_0) + tX_v(u_0, v_0) : 0 \leq s \leq \epsilon_1,\ 0 \leq t \leq \epsilon_2\}. \end{aligned}$$

Exercise 3 *What is the area of this parallelogram?*

Now, let's take a term in the Riemann sum for $\int_{\mathcal{S}} f$ where $\mathcal{S} \subset \mathbb{R}^n$ is the name of our surface parameterized by X on $\mathcal{U} \subset \mathbb{R}^2$. Such a term looks like

$$f(\mathbf{p}^*)\,\text{area}(X(Q)) \sim f(\mathbf{p}^*)\,\text{area}(A(Q)),$$

and we see it is important to know how to compute area$(A(Q))$. First of all, you may have noticed that area$(Q) = \epsilon_1\epsilon_2\,$area$(P)$. Remember P is the parallelogram spanned by the tangent vectors X_u and X_v. The area of this parallelogram turns out to satisfy

$$\sigma = |X_u \times X_v| = \sqrt{\det[(DX)^T\,DX]}. \tag{7}$$

Here we take DX as the matrix with X_u and X_v in the columns. Then DX is a 3×2 matrix and $(DX)^T\,DX$ is a 2×2 matrix with positive determinant. The expression on the left of (7) is valid for surfaces, but the expression on the right will work any time you have a domain $\mathcal{U}$ of lower dimension which is nondegenerately mapped onto a manifold of integration $\mathcal{V}$ in a higher dimensional space. In summary,

$$\int_{\mathcal{S}} f = \int_{\mathcal{U}} (f \circ X)\,\sigma$$

expresses integration on the surface $\mathcal{S}$ in terms of an integral over a planar region $\mathcal{U}$.