# Supplement 4: Differentiating under an integral sign

## S4.1 Differentiating through an integral

In the derivation of Maximum Likelihood Estimators, or the Cramér–Rao Lower Bound, we "differentiated under an integral sign," and I never told you explicitly when that is allowed. Here is a theorem that gives conditions under which it is permissible.

We start with a function $g$ of a vector variable $\boldsymbol{x}$ and a single parameter $\theta$. Let the domain of $g$ be $\Theta \times X$, where $X$ is a rectangle in $\boldsymbol{R}^{\mathrm{n}}$ and $\Theta$ is an interval in $\boldsymbol{R}$. So

$$g\colon \Theta \times X \to \boldsymbol{R}.$$

Define the function $G\colon \Theta \to \boldsymbol{R}$ by

$$G(\theta) = \int_X g(\theta; \boldsymbol{x})\, d\boldsymbol{x}.$$

In order for this to make sense we assume:

**S4.1.1 Assumption** *Let $X$ be a rectangle in $\boldsymbol{R}^{\mathrm{n}}$, let $\Theta$ be an open interval in $\boldsymbol{R}$. Assume that for every $\theta \in \Theta$ the function $\boldsymbol{x} \mapsto g(\theta; \boldsymbol{x})$ is continuous and has a finite integral. That is,*

$$\int_X |g(\theta; \boldsymbol{x})|\, d\boldsymbol{x} < \infty \qquad \big(\theta \in \Theta\big).$$

We would like to show that

$$G'(\theta) = \int_X D_1 g(\theta; \boldsymbol{x})\, d\boldsymbol{x},$$

where $D_1 g$ indicates the partial derivative of $g$ with respect to its first argument $\theta$.

In order for this to be true, the partial derivative has to exist and be integrable with respect to $\boldsymbol{x}$.

**S4.1.2 Assumption** *Assume that every $\boldsymbol{x} \in X$, the function $\theta \mapsto g(\theta; \boldsymbol{x})$ is continuous, and for every $\theta \in \Theta$,*

*the partial derivative $D_1 g(\theta; \boldsymbol{x})$ exists,*

*and is integrable with respect to $\boldsymbol{x}$. That is,*

$$\int_X |D_1 g(\theta; \boldsymbol{x})|\, d\boldsymbol{x} < \infty.$$

But this is not enough. As you may remember, a partial derivative is a limit of the form $\lim_{h\to 0}\big(g(\theta + h, \boldsymbol{x}) - g(\theta, \boldsymbol{x})\big)/h$. You may have heard that passing limits through an integral sign requires a boundedness condition (the Lebesgue Dominated Convergence Theorem). It is beyond the scope of this course and this note to go through the details, but let me say that what is required is called a **uniform local integrability condition**. To get a feeling for what this means, realize that Assumption S4.1.2 requires that at each $\theta$ the function $D_1 g(\theta; \boldsymbol{x})$ is bounded in absolute value by an integrable function of $\boldsymbol{x}$, namely $|D_1 g(\theta; \boldsymbol{x})|$ itself. Uniform local integrability requires that for each $\theta_0$, there is little strip $(\theta_0 - \varepsilon, \theta + \varepsilon) \times X$ on which $|D_1 g(\theta; \boldsymbol{x})|$ is bounded uniformly in $\theta$ for each $\boldsymbol{x}$ by an integrable function of $\boldsymbol{x}$.

**S4.1.3 Assumption** *Assume that for every $\theta_0 \in \Theta$ there is an $\varepsilon > 0$ and a nonnegative function $h\colon X \to \boldsymbol{R}$ (depending on $\theta_0$) such that for all $\boldsymbol{x} \in X$,*

$$\sup_{\theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)} |D_1 g(\theta; \boldsymbol{x})| \leqslant h(\boldsymbol{x}).$$

*and*

$$\int_X h(\boldsymbol{x})\, d\boldsymbol{x} < \infty.$$

These assumptions are sufficient.[4]

**S4.1.4 Theorem** *Let $g\colon \Theta \times X \to \boldsymbol{R}$ satisfy Assumptions S4.1.1, S4.1.2, and S4.1.3. Then the function $G\colon \Theta \to \boldsymbol{R}$ defined by*

$$G(\theta) = \int_X g(\theta; \boldsymbol{x})\, d\boldsymbol{x}$$

*is differentiable on $\Theta$ and*

$$G'(\theta) = \int_X D_1 g(\theta; \boldsymbol{x})\, d\boldsymbol{x}.$$

*Proof*: For a proof, see Theorem 24.5 of Aliprantis and Burkinshaw [1, pp. 193–194] *and* the remarks following it. ∎

Assumption S4.1.3 is awkward to use, so a stronger condition that implies it is useful. The next theorem may be easier to apply. It requires $X$ to be closed and bounded, and $g$ to have a jointly continuous partial derivative. It is similar to Theorem 8.11.2 in Dieudonné [2, p. 177].

**S4.1.5 Theorem** *Let $X$ be a closed and bounded rectangle in $\boldsymbol{R}^{\mathrm{n}}$, let $\Theta$ be an open interval in $\boldsymbol{R}$, and let $g\colon \Theta \times X \to \boldsymbol{R}$ be jointly continuous. Assume that the partial derivative $D_1 g(\theta; \boldsymbol{x})$ is jointly continuous on $\Theta \times X$. Then the function $G\colon \Theta \to \boldsymbol{R}$ defined by*

$$G(\theta) = \int_X g(\theta; \boldsymbol{x})\, d\boldsymbol{x}$$

*is differentiable on $\Theta$ and*

$$G'(\theta) = \int_X D_1 g(\theta; \boldsymbol{x})\, d\boldsymbol{x}.$$

*Proof*: Since $g$ and $D_1 g$ are jointly continuous, they are bounded on every set of the form $[\theta_0 - \varepsilon, \theta_0 + \varepsilon] \times X$, and so integrable. Therefore the Hypotheses of Theorem S4.1.4 are satisfied. ∎

## S4.2 An illustrative (counter)example

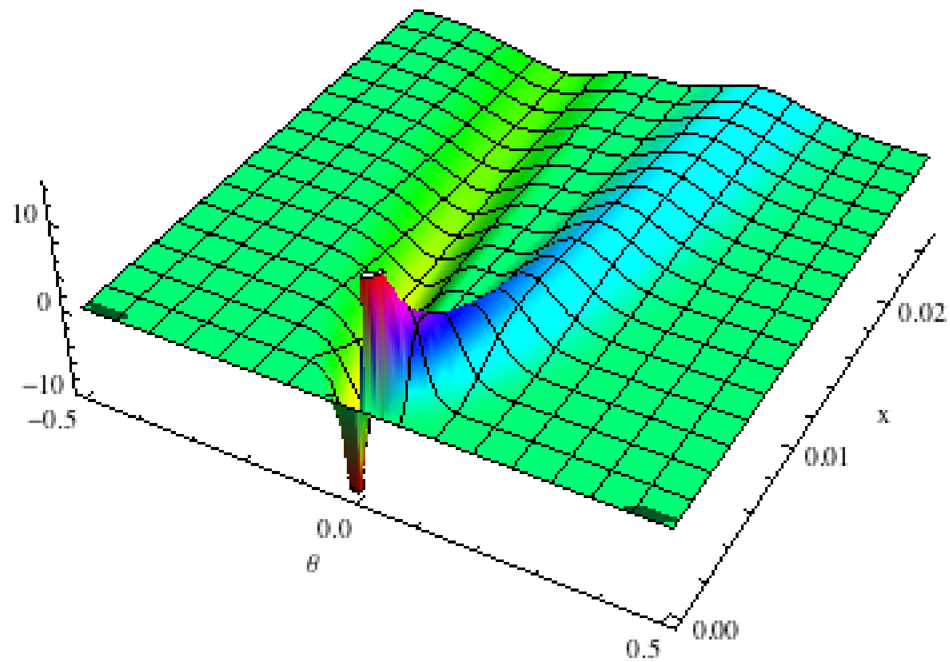
The following example shows what can go wrong when the uniform local integrability Assumption S4.1.3 is violated. You can find it in Gelbaum and Olmsted [3, Example 9.15, p. 123], but similar examples are well-known.

**S4.2.1 Example** Let $\Theta = \boldsymbol{R}$ and let $X = [0, 1]$. Define $g\colon \Theta \times X \to \boldsymbol{R}$ via

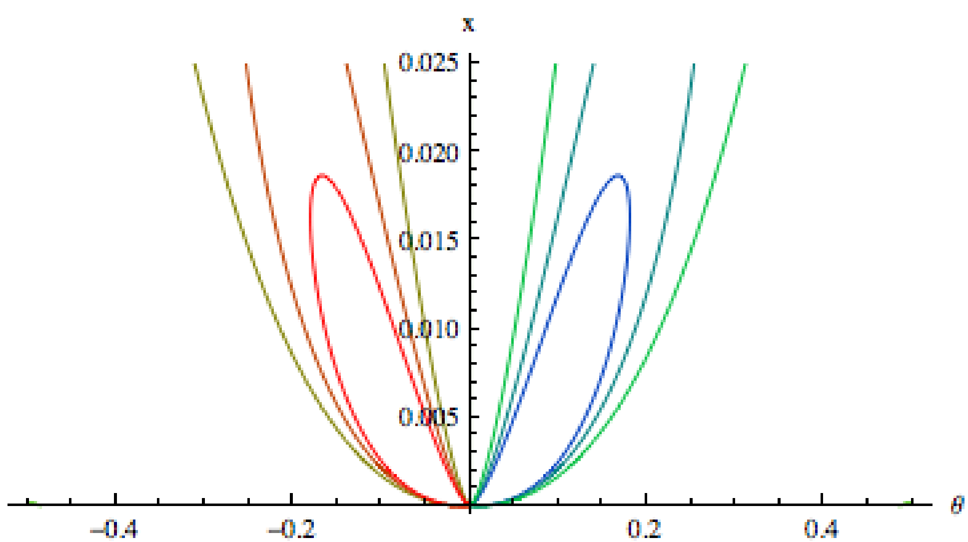
$$g(\theta, x) = \begin{cases} \dfrac{\theta^3}{x^2} e^{-\theta^2/x} & x > 0, \\ 0 & x = 0. \end{cases}$$

The function $g$ is plotted in Figure 17.2.

[4] We can make weaker assumptions, see, for instance, Aliprantis and Burkinshaw [1, pp. 193–194]. The problem is that the weaker assumptions involve measure-theoretic concepts that are beyond the scope of this course.




Surface of graph.




Contours.

Figure 17.2. Plots of $g(\theta;x) = \dfrac{\theta^3}{x^2}e^{-\theta^2/x}$, $(x > 0)$.

(Notice that for fixed $x$, the function $\theta \mapsto g(\theta, x)$ is continuous at each $\theta$; and for each fixed $\theta$, the function $x \mapsto g(\theta, x)$ is continuous at each $x$, including $x = 0$. (This is because the exponential term goes to zero much faster than polynomial term goes to zero as $x \to 0$.) The function $g$ is not jointly continuous though: Along the curve $x = \theta^2$ we have $g(\theta, x) = e^{-1}/\theta$, which diverges to $\infty$ as $\theta \downarrow 0$ and diverges to $-\infty$ as $\theta \uparrow 0$.)

Define

$$\begin{aligned} G(\theta) &= \int_0^1 g(\theta; x)\, dx \\ &= \theta^3 \int_0^1 \frac{1}{x^2} e^{-\theta^2/x}\, dx. \end{aligned}$$

Consulting a table of integrals if necessary, we find the indefinite integral $\int \frac{1}{x^2} e^{-a/x}\, dx = e^{-a/x}/a$. Thus, letting $a = \theta^2$ we have

$$\boxed{G(\theta) = \theta e^{-\theta^2} \qquad (\theta \in \boldsymbol{R}).}$$

The function $G$ is plotted in Figure 17.3. Differentiation yields

$$\boxed{G'(\theta) = (1 - 2\theta^2) e^{-\theta^2} \qquad (\theta \in \boldsymbol{R}).}$$

Note that $G'(0) = (1 - 2 \cdot 0^2) e^{-0^2} = 1$.

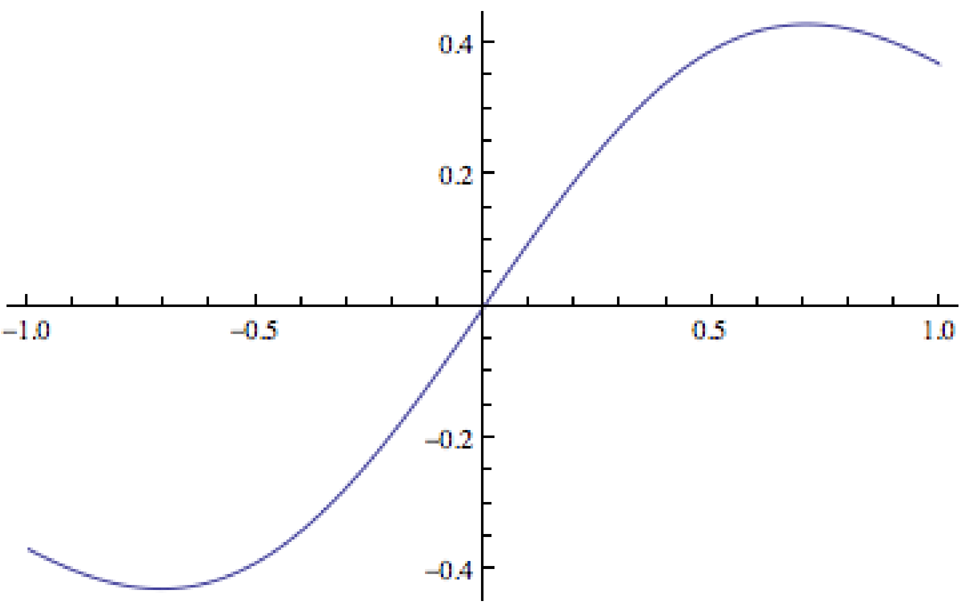



Figure 17.3. $G(\theta)$.

Now let's compute $D_1 g(\theta; x)$: For $x = 0$, $g(\theta; x) = 0$ for all $\theta$, so $D_1 g(\theta; 0) = 0$. For $x > 0$, we have

$$\begin{aligned} D_1 g(\theta; x) &= \frac{3\theta^2}{x^2} e^{-\theta^2/x} + \frac{\theta^3}{x^2} e^{-\theta^2/x} (-2\theta/x) \\ &= \Big( \frac{3\theta^2}{x^2} - \frac{2\theta^4}{x^3} \Big) e^{-\theta^2/x}. \end{aligned}$$

So

$$D_1 g(\theta;x) = \begin{cases} \left(\frac{3\theta^2}{x^2} - \frac{2\theta^4}{x^3}\right)e^{-\theta^2/x} & x > 0 \\ 0 & x = 0. \end{cases}$$

(Note that for fixed $\theta$, the limit of $D_1 g(\theta;x)$ as $x \downarrow 0$ is zero, so for each fixed $\theta$, $D_1 g(\theta;x)$ is continuous in $x$. Also, for each fixed $x$, $D_1 g(\theta;x)$ is continuous in $\theta$. But again, along the curve $x = \theta^2$, we have $D_1 g(\theta;x) = \left(3\theta^{-2} - 2\theta^{-2}\right)e^{-1} = -e^{-1}/\theta^2$ which diverges to $\infty$ as $\theta \to 0$. Thus $D_1 g(\theta;x)$ is not jointly continuous at $(0,0)$. See Figures 17.4 and 17.5.)

Define the integral

$$I(\theta) = \int_0^1 D_1 g(\theta;x)\,dx = \int_0^1 \left(\frac{3\theta^2}{x^2} - \frac{2\theta^4}{x^3}\right)e^{-\theta^2/x}\,dx.$$

It satisfies $I(0) = 0$.

At $\theta = 0$, we have

$$G'(0) = (1 - 2\cdot 0^2)e^{-0^2} = 1 \neq 0 = I(0) = \int_0^1 D_1 g(0;x)\,dx,$$

So the conclusion of Theorem S4.1.4 fails.

The failure of the theorem is confined to the case $\theta = 0$. For $\theta > 0$, $I(\theta)$ can be computed as

$$\begin{aligned} I(\theta) = \int_0^1 D_1 g(\theta;x)\,dx &= \int_0^1 \left(\frac{3\theta^2}{x^2} - \frac{2\theta^4}{x^3}\right)e^{-\theta^2/x}\,dx \\ &= 3\theta^2 \int_0^1 \frac{1}{x^2}e^{-\theta^2/x}\,dx - 2\theta^4 \int_0^1 \frac{1}{x^3}e^{-\theta^2/x}\,dx \\ &= 3\theta^2 \left[\frac{e^{-\theta^2/x}}{\theta^2}\Big|_{x=0}^{x=1}\right] - 2\theta^4 \left[e^{-\theta^2/x}\left(\frac{1}{\theta^4} + \frac{1}{\theta^2 x}\right)\Big|_{x=0}^{x=1}\right] \\ &= (1 - 2\theta^2)e^{-\theta^2}. \end{aligned}$$

For $\theta > 0$, we have

$$G'(\theta) = (1 - 2\theta^2)e^{-\theta^2} = I(\theta) = \int_0^1 D_1 g(\theta;x)\,dx.$$

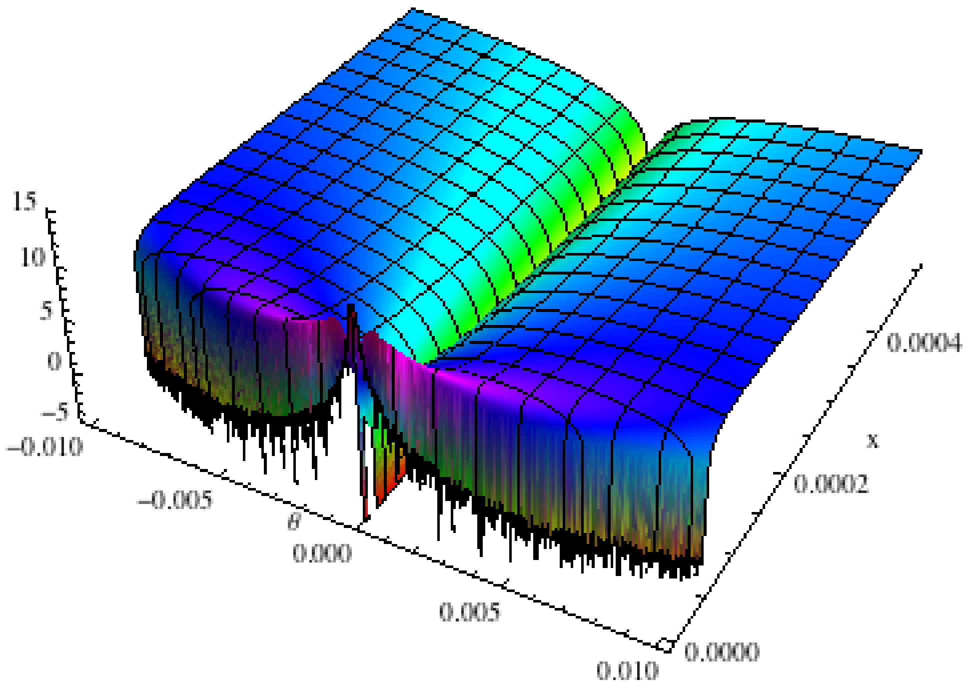
So the only problem is right at $\theta = 0$.

Let's check Assumption S4.1.3 for $\theta_0 = 0$. We need to find an $\varepsilon > 0$ so that $|\theta| < \varepsilon$ implies $|D_1 g(\theta;x)| \leqslant h(x)$ where $\int_0^1 h(x)\,dx < \infty$. Now for $x > 0$,

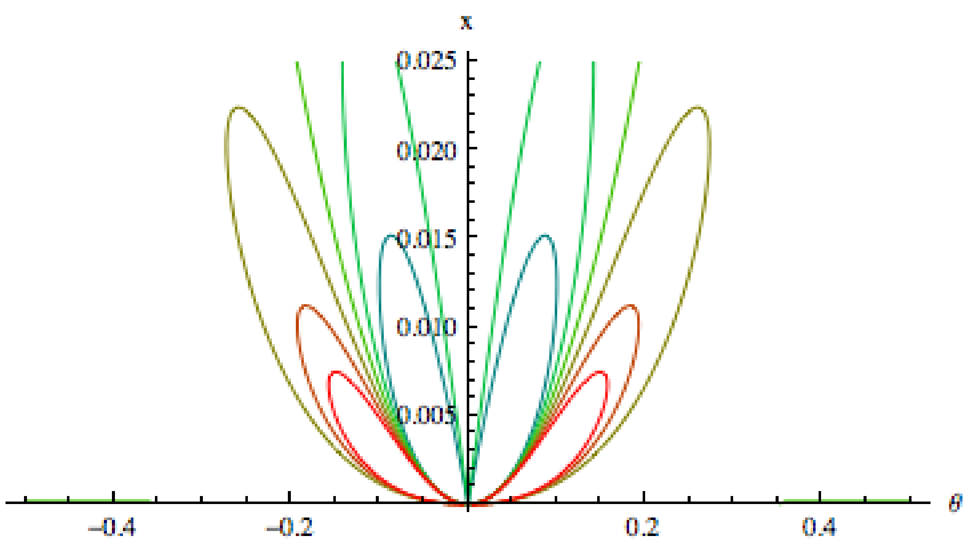
$$D_1 g(\theta;x) = \left(\frac{3\theta^2}{x^2} - \frac{2\theta^4}{x^3}\right)e^{-\theta^2/x}.$$

Looking at points of the form $x = \theta^2$, we see that $h(x)$ must satisfy

$$h(x) \geqslant D_1 g(\sqrt{x};x) = \left(\frac{3}{x} - \frac{2}{x}\right)e^{-1} = e^{-1}/x.$$




Surface of the log graph.




Contours.

Figure 17.4. Plots of the log of $D_1 g(\theta; x) = \left(\frac{3\theta^2}{x^2} - \frac{2\theta^4}{x^3}\right) e^{-\theta^2/x}$.

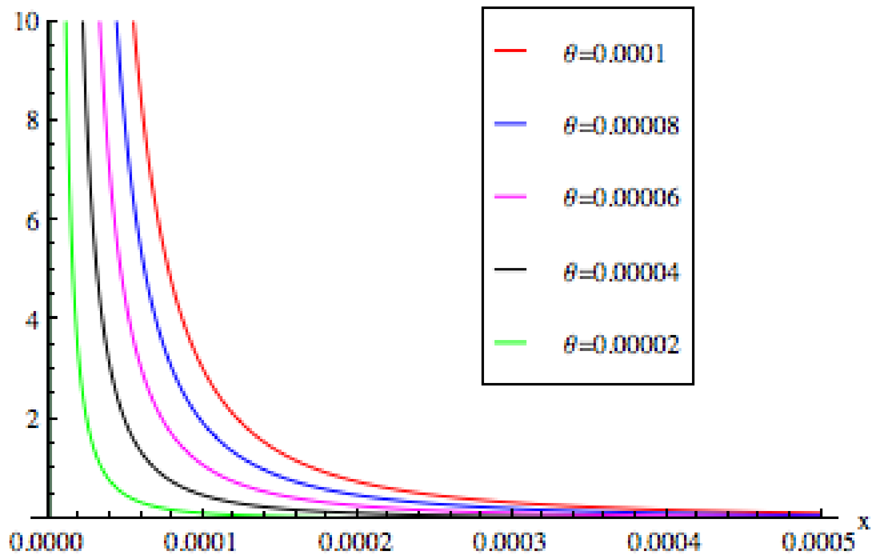



Figure 17.5. Cross-sections of $D_1 g(\theta; x) = \left(\frac{3\theta^2}{x^2} - \frac{2\theta^4}{x^3}\right) e^{-\theta^2/x}$.

See Figure 17.5.

So for any $\varepsilon > 0$ for all $\theta < \sqrt{\varepsilon}$, we have $x \leqslant \theta$ implies $h(x) \geqslant e^{-1}/x$. Thus

$$\int_0^1 h(x)\,dx \geqslant \int_0^{\sqrt{\varepsilon}} h(x)\,dx \geqslant e^{-1} \int_0^{\sqrt{\varepsilon}} \frac{1}{x}\,dx = \infty.$$

Thus Assumption S4.1.3 is violated by this example, and the conclusion of the theorem fails. □

## Bibliography

[1] C. D. Aliprantis and O. Burkinshaw. 1998. *Principles of real analysis*, 3d. ed. San Diego: Academic Press.

[2] J. Dieudonné. 1969. *Foundations of modern analysis*. Number 10-I in Pure and Applied Mathematics. New York: Academic Press. Volume 1 of Treatise on Analysis.

[3] B. R. Gelbaum and J. M. Olmsted. 2003. *Counterexamples in analysis*. Minneola, NY: Dover. Slightly corrected reprint of the 1965 printing of the work originally published in 1964 by Holden–Day in San Francisco.