Notes on numerical integration of functions

Tomás A. Arias

Cornell University

Department of Physics

January 26, 2004

Contents

1 Introduction **2**

2 Theory of regular-interval integration **2**
2.1 Rules for narrow intervals 2
2.2 Rules for extended intervals 5
2.3 Error of the extended rules 5

3 Digit-shift test **7**

4 Richardson Extrapolation **7**

5 Change of variables **9**

List of Figures

1 Decomposition of the integral $\int_a^b f(x)\,dx$ into N regular intervals 3
2 Integration of interval I_i. 4
3 Linear behavior of the trapezoid rule when viewed as a function of h^2. 8
4 Mapping of regular sampling of unit interval under change of variables: (a) linear map, (b) map to semi-infinite interval, (c) map to semi-infinite interval with concentration of points at origin. 10
5 Integrand of $\int_0^\infty \frac{1}{1+r^2}\,dr$: (a) direct integration, (b) under change of variables in Eq. (21). . . 11
6 Integrand of $\int_0^\infty \frac{e^{-r}}{\sqrt{r}}\,dr$: (a) direct integration, (b) under change of variables in Eq. (22). . . . 11

List of Tables

1 Results of trapezoid and Simpson's integration of $\int_0^1 (1/\ln 2)/(1+x)\,dx$ for N intervals 7

1 Introduction

As we have seen in the notes on density functional theory, once the orbitals $\psi_i(\vec{x})$ have been determined, the final steps in determining the total energy E involve calculating various integrals. Because we will determine the $\psi_i(\vec{x})$ numerically, we must calculate these integrals numerically as well.

Generically, numerical integration techniques take the form of a weighted sum of sample values,

$$\int f(x)\,dx \approx \sum_i w_i f(x_i), \tag{1}$$

where the w_i are the weights, the x_i are the sample points and the $f(x_i)$ are the sample values. Note that the integral may be multidimensional, in which case the sample points x are vectors.

Different techniques for choosing the sample points categorize the various techniques. These include choosing the sample points x_i

1. at regular intervals (trapezoid, Simpson's and higher-order rules);
2. fixed in number at locations chosen to produce the highest possible accuracy (Gaussian quadrature);
3. according to the behavior of $f(x_i)$ (adaptive integration);
4. at random (Monte Carlo integration).

Given the sample points, the mathematics of integration then determines the appropriate weights so that (1) becomes a good approximation.

Each of the above approaches has strengths and weaknesses. The regular-interval approaches (1) are best when the sample values $f(x_i)$ come from a regular process, such as experimental measurements or another numerical calculation. For high accuracy under conditions where the evaluation of $f(x)$ is expensive but we have control over the selection of the x_i, Gaussian quadrature (2) comes to the fore. Both (1) and (2) presume that the integrand $f(x)$ varies relatively uniformly throughout the integration domain. In cases where $f(x)$ varies inhomogeneously, but in a predictable way, the combination of (1) or (2) with the elegant approach of a change of variables (Section 5) can be very powerful. Otherwise, and adaptive approaches (3) become necessary. Finally, Monte Carlo (4) methods are the only choice for integrals in very large dimensional spaces, such as thermodynamic averages or quantum mechanical path integrals. When multiple approaches apply, simplicity should always be the deciding factor.

For us, the $\psi_i(\vec{x})$ will come from a numerical process producing data on regular intervals. The $\psi_i(\vec{x})$ do not vary uniformly, but have strong variations near the nuclei where the Coulomb energy $-Z/r$ is singular. This behavior is quite predictable, and the simplest viable approach is therefore the combination of a regular interval approach (1) with a change of variables, which is the subject of these notes.

2 Theory of regular-interval integration

2.1 Rules for narrow intervals

The regular interval approaches all derive from the trapezoid rule through the powerful concept of Richardson extrapolation (Section 4). We therefore need consider in detail only the theory of the trapezoid rule.

Figure 1 illustrates the basic concept underlying regular-interval approaches, decomposition of the desired integral

$$I \equiv \int_a^b f(x)\,dx, \tag{2}$$

into a sum of integrals over N regular intervals of width $h \equiv (b-a)/N$

$$I \;=\; \sum_{i=1}^{N} \int_{x_{i-1}}^{x_i} f(x)\,dx, \tag{3}$$

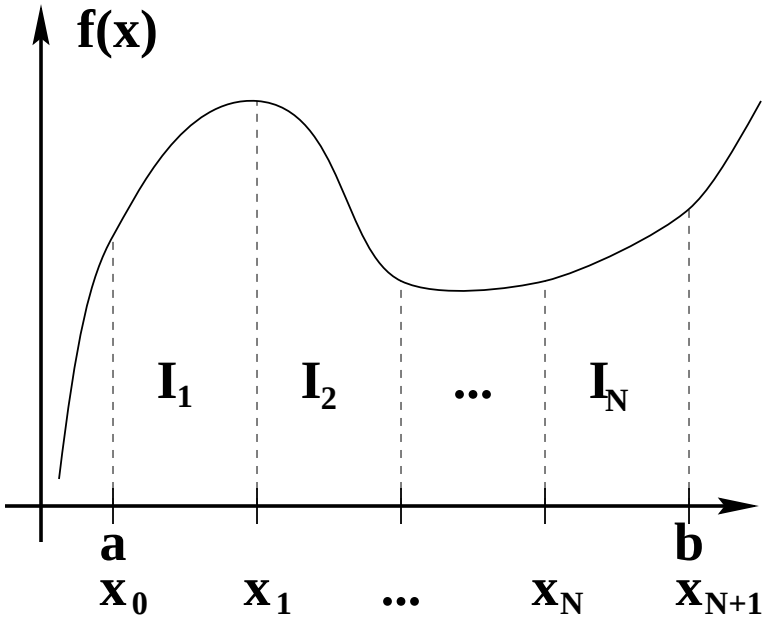


Figure 1: Decomposition of the integral $\int_a^b f(x)\,dx$ into N regular intervals

$$\equiv \sum_{i=1}^{N} I_i \tag{4}$$

where $x_i \equiv a + ih$ for $i = 0 \ldots N$. The rationale for this analysis is that the narrowness of the intervals allows the development of expansions for the I_i. For the trapezoid rule, the x_i become $N+1$ sample points for the integration.

Narrowness of the intervals i ensures that Taylor series about the center points $x_i^{(m)} \equiv (x_{i-1} + x_i)/2$ well represent $f(x)$ over the corresponding interval,

$$\begin{aligned} f(x) &= \sum_n \frac{1}{n!} f_m^{[n]} u^n \\ &= f_m + f'_m u + \frac{1}{2} f''_m u^2 + \frac{1}{6} f'''_m u^3 + \frac{1}{24} f^{iv} u^4 + \ldots, \end{aligned} \tag{5}$$

where $u \equiv x - x_i^{(m)}$, and we define the value and derivatives of the function at the midpoint as $f_m^{[n]} \equiv f^{[n]}(x_i^{(m)})$. The integral over interval i is then

$$\begin{aligned} I_i &= \int_{-h/2}^{h/2} \left(\sum_n \frac{1}{n!} f_m^{[n]} u^n \right) du \\ &= \sum_{n \text{ even}} \frac{1}{n!} f_m^{[n]} \frac{1}{n+1} 2 \left(\frac{h}{2} \right)^{n+1} \\ &= \sum_{n \text{ even}} \frac{1}{2^n (n+1)!} f_m^{[n]} h^{n+1} \\ &= h \left(f_m + \frac{1}{24} f''_m h^2 + \frac{1}{1920} f^{iv}_m h^4 + \ldots \right), \end{aligned} \tag{6}$$

where all of the odd n terms integrate to zero by symmetry.

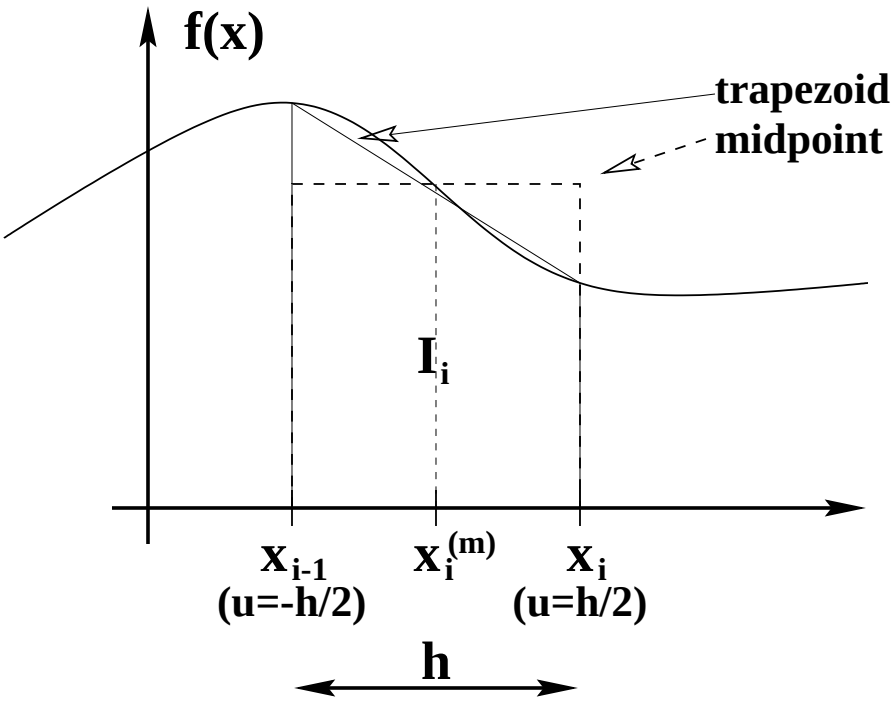


Figure 2: Integration of interval I_i.

From (6), the lowest-order approximation to I_i is

$$\begin{aligned} M_i &\equiv f_m h, \\ &= f\left(\frac{x_{i-1}+x_i}{2}\right) h, \end{aligned} \tag{7}$$

the so-called midpoint rule. (See Figure 2.) The associated error is second order,

$$M_i - I_i = h\left(-\frac{1}{24} f''_m h^2 - \frac{1}{1920} f^{iv}_m h^4 + \mathcal{O}(h^6)\right), \tag{8}$$

where $\mathcal{O}(h^6)$ denotes that the remaining terms are of the order h^6 or higher.

If the x_i are our sample points, we have no direct knowledge of the f_m, however. Working with the sample points nearest to $x^{(m)}$, $x_{i-1} = x_i^{(m)} - h/2$ and $x_i = x_i^{(m)} + h/2$, we do have knowledge of the values

$$\begin{aligned} f(x_{i-1}) &= \sum_n \frac{1}{n!} f^{[n]}_m \left(-\frac{h}{2}\right)^n \\ &= \sum_n \frac{(-1)^n}{2^n n!} f^{[n]}_m h^n \\ &= f_m - \frac{1}{2} f'_m h + \frac{1}{8} f''_m h^2 - \frac{1}{48} f'''_m h^3 + \frac{1}{384} f^{iv}_m h^4, \\ f(x_i) &= \sum_n \frac{1}{2^n n!} f^{[n]}_m h^n \\ &= f_m + \frac{1}{2} f'_m h + \frac{1}{8} f''_m h^2 + \frac{1}{48} f'''_m h^3 + \frac{1}{384} f^{iv}_m h^4, \end{aligned} \tag{9}$$

where we have used the series (5). Multiplying the average of these two values with h gives the *trapezoid rule* for approximating I_i,

$$\begin{aligned} T_i &\equiv \frac{f(x_{i-1}) + f(x_i)}{2} h \\ &= h \sum_{n \text{ even}} \frac{1}{2^n n!} f_m^{[n]} h^n \\ &= h \left(f_m + \frac{1}{8} f_m'' h^2 + \frac{1}{384} f_m^{iv} h^4 + \ldots \right). \end{aligned} \tag{10}$$

This also has a second-order error,

$$T_i - I_i = h \left(\frac{1}{12} f_m'' h^2 + \frac{1}{480} f_m^{iv} h^4 . \mathcal{O}(h^6) \right). \tag{11}$$

2.2 Rules for extended intervals

Once given a rule for computing integrals I_i over narrow intervals, Eq. (3) prescribes a method for numerical calculation of integrals over extended regions: for the midpoint rule,

$$\begin{aligned} M &\equiv \sum_{i=1}^{N} M_i \\ &= \sum f(x_i^{(m)}) h, \\ &= \sum f\left(\frac{x_{i-1} + x_i}{2} \right) h, \end{aligned} \tag{12}$$

and for the trapezoid rule,

$$\begin{aligned} T &\equiv \sum_{i=1}^{N} T_i \\ &= \sum_{i=1}^{N} \frac{f(x_{i-1}) + f(x_i)}{2} h \\ &= h \begin{pmatrix} \frac{f_0}{2} & + & \frac{f_1}{2} & & & & & & \\ & + & \frac{f_1}{2} & + & \frac{f_2}{2} & & & & \\ & & & + & \frac{f_2}{2} & + & \cdots & & \\ & & & & & & \cdots & + & \frac{f_{N-1}}{2} \\ & & & & & & & + & \frac{f_{N-1}}{2} \quad + \quad \frac{f_N}{2} \end{pmatrix} \\ &= h \left(\frac{1}{2} f_0 + f_1 + f_2 + \ldots + f_{N-1} + \frac{1}{2} f_N \right), \end{aligned} \tag{13}$$

(14)

where $f_i \equiv f(x_i)$.

2.3 Error of the extended rules

Using (8), we have for the midpoint rule that

$$M - I = \sum_{i=1}^{N} M_i - \sum_{i=1}^{N} I_i$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} (M_i - I_i) \\
&= \sum_{i=1}^{N} h \left(-\frac{1}{24} f''(x_i^{(m)}) h^2 + \mathcal{O}(h^4) \right) \\
&= -\frac{1}{24} h^2 \left(\sum_{i=1}^{N} f''(x_i^{(m)}) h \right) + \mathcal{O}(h^4).
\end{aligned}
$$

As $h \to 0$, the term in brackets approaches a constant (namely, $\int_a^b f''(x)\,dx = f'(b) - f'(a)$). Thus, the leading order error in the midpoint rule is second order in h. Moreover, the term in brackets is, in fact, the midpoint rule for computing $\int_a^b f''(x)\,dx$! Thus, we may complete the analysis and compute precisely the numerical prefactor on h^2,

$$
\begin{aligned}
M - I &= -\frac{1}{24} h^2 \left(f'(b) - f'(a) + \mathcal{O}(h^2) \right) + \mathcal{O}(h^4) \\
&= -\frac{1}{24} \left(f'(b) - f'(a) \right) h^2 + \mathcal{O}(h^4). \qquad (15)
\end{aligned}
$$

The trapezoid rule therefore errs in the second order with a prefactor proportional to $f'(a) - f'(b)$! If, by coincidence or design, $f'(a) = f'(b)$, the midpoint rule becomes correct to fourth order!

Working in the same way from (11), we have for the trapezoid rule,

$$
\begin{aligned}
T - I &= \sum_{i=1}^{N} (T_i - I_i) \\
&= \sum_{i=1}^{N} h \left(\frac{1}{12} f''(x_i^{(m)}) h^2 + \frac{1}{480} f^{iv}(x_i^{(m)}) h^4 + \mathcal{O}(h^6) \right) \\
&= \frac{1}{12} h^2 \sum_{i=1}^{N} f''(x_i^{(m)}) h + \frac{1}{480} h^4 \sum_{i=1}^{N} f^{iv}(x_i^{(m)}) h + \mathcal{O}(h^6) \\
&= \frac{1}{12} h^2 \left(\int_a^b f''(x)\,dx - \frac{1}{24} \left(f'''(b) - f'''(a) \right) h \right) + \frac{1}{480} h^4 \left(\int_a^b f^{iv}(x)\,dx + \mathcal{O}(h^2) \right) + \mathcal{O}(h^6) \\
&= \frac{1}{12} h^2 \left(f'(b) - f'(a) \right) - \frac{1}{720} h^4 \left(f'''(b) - f'''(a) \right) + \mathcal{O}(h^6),
\end{aligned}
$$

where we have twice used the result (15). The trapezoid rule thus also becomes fourth-order if $f'(a) = f'(b)$, and sixth-order if also $f'''(a) = f'''(b)$.

Because the error series for both M_i and T_i involve only even terms with numerical prefactors times the respective derivatives of $f(x)$, we see that the full expansion for the error of the trapezoid rule has the form

$$
T - I = \sum_{n \text{ even}} C_n \left(f^{[n-1]}(b) - f^{[n-1]}(a) \right) h^n,
$$

for some set of numerical coefficients C_n[1]. Thus, if we could arrange for all of the derivatives of $f(\ldots)$ to be equal at the endpoints a and b, then the trapezoid rule will converge faster than h to any power — exponentially! This is not so artificial as first appears: a very common case is the numerical integration of periodic functions.

[1] By considering the effects of the trapezoid rule on $\int_0^\infty e^{-sx}\,dx$, you should be able to convince yourself that C_n is equal to the numerical prefactor of the h^n term in the power series expansion for $h\left(1/2 + 1/(e^{sh} - 1)\right)$.

N	h	Trapezoid ($T(h)$)	Trapezoid (w/ bug)	Richardson Extrap ($\frac{4}{3}T(h/2) - \frac{1}{3}T(h)$)	Simpson ($S(h/2)$)
10	0.1000	1.0009 0056 2871	1.0360 6765 6056	1.0000 0028 0034	1.0000 0028 0034
100	0.0100	1.0000 0901 6731	1.0036 0673 7630	1.0000 0000 0028	1.0000 0000 0028
1000	0.0010	1.0000 0009 0168	1.0003 6067 3760	1.0000 0000 0000	1.0000 0000 0000
10000	0.0001	1.0000 0000 0901	1.0000 3606 7376	1.0000 0000 0000	1.000 000 000 000

Table 1: Results of trapezoid and Simpson's integration of $\int_0^1 (1/\ln 2)/(1+x)\,dx$ for N intervals

3 Digit-shift test

To illustrate the ideas of Section 2.3, the second column of Table 1 displays the result of the trapezoid rule (13) for the integral

$$\int_0^1 \frac{1/\ln 2}{1+x}\,dx = 1,$$

as a function of the number of intervals N. Displayed in this way the data for the trapezoid rule exhibit an interesting pattern: each increase in the number of intervals N by a factor of ten shifts the pattern of digits ("901") representing the error two places to the right. This is the visual representation of our mathematical analysis: the shifting to the right by two places corresponds precisely to quadratic scaling of the error with h and the pattern of shifting digits represents precisely the prefactor on h^2 in (16),

$$\begin{aligned}
\frac{1}{12}\left(f'(b) - f'(a)\right) &= \frac{1}{12}\left(-\frac{1/\ln 2}{(1+x)^2}\right|_0^1 \\
&= \frac{1}{12\ln 2}\left(1 - \frac{1}{4}\right) \\
&= \frac{3}{48\ln 2} \\
&\approx 0.090168.
\end{aligned}$$

Displaying the results of a numerical calculation in this way is a very powerful tool for detecting bugs and correcting errors. The h^2 convergence of the trapezoid rule is quite delicate and even small programming errors will destroy it. The third column of the table, for instance, shows the typical behavior when the factor of $1/2$ is forgotten at the end points and/or when dividing by $N-1$ instead of N. A routine with either bug indeed converges to the correct result, but the convergence is only linear in h and such subtle bugs are easily detected.

4 Richardson Extrapolation

Plotting $T(h)$ as a function of h^2, Figure 3 shows that the behavior $T(h) = I + \alpha h^2$ is evident for even the modest values $N = 2, 4, 8$. This plot suggests a very powerful method of finding much better estimates for the integral $I \equiv T(0)$: simply *extrapolate* the data in the plot to $h = 0$. In general, we will not know *a priori* the analytic derivatives $f'(a)$ and $f'(b)$ and thus the slope α of the approach to the origin in the plot. But, with two values of $T(h)$ we can determine a very good estimate. Algebraically, we simply solve

$$\begin{aligned}
T(h) &\approx T(0) + \alpha h^2 \\
T(h/2) &\approx T(0) + \alpha (h/2)^2,
\end{aligned}$$

to find the value

$$T(0) \approx \frac{4}{3}T(h/2) - \frac{1}{3}T(h). \tag{16}$$

The fourth column of Table 1 shows that the resulting extrapolation formula (16) indeed converges much more quickly than $T(h)$, as h^4 in fact.

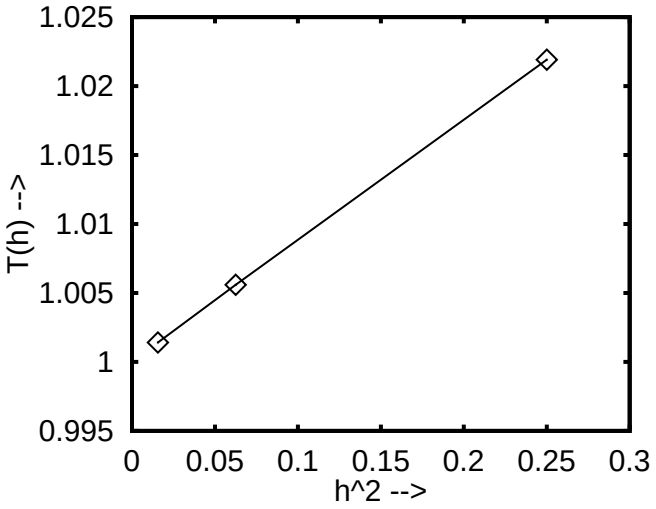


Figure 3: Linear behavior of the trapezoid rule when viewed as a function of h^2.

The extrapolation (16) translates directly into an uniform-interval method. From the definition of the trapezoid rule (13),

$$\begin{aligned} -\frac{1}{3}T(h) &= -\frac{1}{3}h\left(\frac{1}{2}f_0 + f_1 + f_2 + \ldots + f_{N-1} + \frac{1}{2}f_N\right) \\ \frac{4}{3}T(h/2) &= \frac{4}{3}\frac{h}{2}\left(\frac{1}{2}f_0 + f_{1/2} + f_1 + \ldots + f_{(N-1)/2} + \frac{1}{2}f_N\right), \end{aligned}$$

where $f_{i-1/2} \equiv f(x_{i-1/2})$ with $x_{i-1/2} \equiv a + (i - 1/2)h$ denoting sample values at the intermediate points which $T(h/2)$ uses but $T(h)$ does not. Summing the above two formulae then gives an explicit interval method,

$$\begin{aligned} S(h/2) &\equiv \frac{4}{3}T(h/2) - \frac{1}{3}T(h) \\ &= \frac{h}{6}\left(f_0 + 4f_{1/2} + 2f_1 + \ldots + 4f_{(N-1)/2} + f_N\right), \end{aligned} \tag{17}$$

$$\tag{18}$$

which is none other than the familiar Simpson's rule! For this rule, we denote the fact that the spacing between sample points x_i is now $h/2$ and not h by writing $S(h/2)$. The final column of Table 1 confirms numerically that Simpson's rule evaluated in the traditional way is precisely equivalent to the extrapolation (16).

Understanding this relationship allows us to rapidly determine the error in Simpson's from the error in the trapezoid rule (16):

$$\begin{aligned} S(h/2) - I &= \frac{4}{3}T(h/2) - \frac{1}{3}T(h) - I \\ &= \frac{4}{3}\left(T(h/2) - I\right) - \frac{1}{3}\left(T(h) - I\right) \\ &= \frac{4}{3}\left(\frac{1}{12}\left(\frac{h}{2}\right)^2\left(f'(b) - f'(a)\right) - \frac{1}{720}\left(\frac{h}{2}\right)^4\left(f'''(b) - f'''(a)\right) + \mathcal{O}(h^6)\right) \\ &\quad -\frac{1}{3}\left(\frac{1}{12}h^2\left(f'(b) - f'(a)\right) - \frac{1}{720}h^4\left(f'''(b) - f'''(a)\right) + \mathcal{O}(h^6)\right) \\ &= \frac{1}{2880}h^4\left(f'''(b) - f'''(a)\right) + \mathcal{O}(h^6), \end{aligned} \tag{19}$$

where the quadratic terms have cancelled by the construction which led to (16), leaving only the quartic term. The patterns of shifting digits in Table 1 again confirm our computation of the prefactor,

$$\begin{aligned} \frac{1}{2880}\left(f'''(b) - f'''(a)\right) &= \frac{1}{2880\ln 2} 6\left(1 - \frac{1}{16}\right) = \frac{90}{46080\ln 2} \\ &= 0.002818\ldots \end{aligned}$$

One may proceed in this manner to methods of arbitrarily high order. For instance, extrapolating $T(h)$ to $h^2 = 0$ as a quadratic function using the information which $T(h)$, $T(h/2)$ and $T(h/4)$ provide results in the sixth-order approach known as Bode's rule.

The idea of extracting an exact analytic result by extrapolating several calculations at finite values of a parameter to the result at zero value is an extremely powerful tool in scientific computing, where it is known as Richardson extrapolation. (The same idea occurs in statistical mechanics where it is often known as finite-size scaling.)

5 Change of variables

Change of variables is an extremely powerful method for performing integrals not only analytically but also numerically. A change of variables can considerably improve the accuracy of regular-interval techniques for functions with rapid variations in particular regions of the integration domain and can allow one to perform integrals which would otherwise be impossible, such as for functions with integrable singularities or integrals over infinite domains.

Change of variables begins with the familiar analytic formula,

$$\int_0^1 f(r(x))\frac{dr(x)}{dx}\,dx = \int_{a=r(0)}^{b=r(1)} f(r)\,dr, \tag{20}$$

where $r(x)$ is any function which maps the unit interval $x \in [0,1]$ to the interval $r \in [a,b]$. The most trivial example of such a map is the linear map

$$\begin{aligned} r(x) &= a + (b-a)x \\ \frac{dr(x)}{dx} &= b - a \end{aligned}$$

which transforms integration over any interval to integration over the unit interval. Figure 4(a) illustrates such a map along with the distribution of resulting sample points $r(x_i)$.

Maps which transform integration over an infinite interval to integration over a finite interval represent a far more important example. The map

$$\begin{aligned} r(x) &= \frac{1}{1-x} - 1 = \frac{x}{1-x} \\ \frac{dr(x)}{dx} &= \frac{1}{(1-x)^2}, \end{aligned} \tag{21}$$

for instance, makes possible the evaluation of integrals over the interval $[0,\infty]$ with the techniques discussed above. As $x \to 0$, $r(x) \to x$, and this map spaces the sample points $r(x_i)$ regularly near the origin. However, As $x \to 1^-$, however, the sample points eventually tend toward infinity. (See Figure 4(b).) Under this map, the integral

$$\int_0^\infty f(r)\,dr = \int_0^\infty \frac{1}{1+r^2}\,dr = \frac{\pi}{2},$$

becomes

$$\int_0^1 \frac{1}{1+\left(\frac{x}{1-x}\right)^2}\frac{1}{(1-x)^2}\,dx = \int_0^1 \frac{1}{(1-x)^2+x^2}\,dx,$$

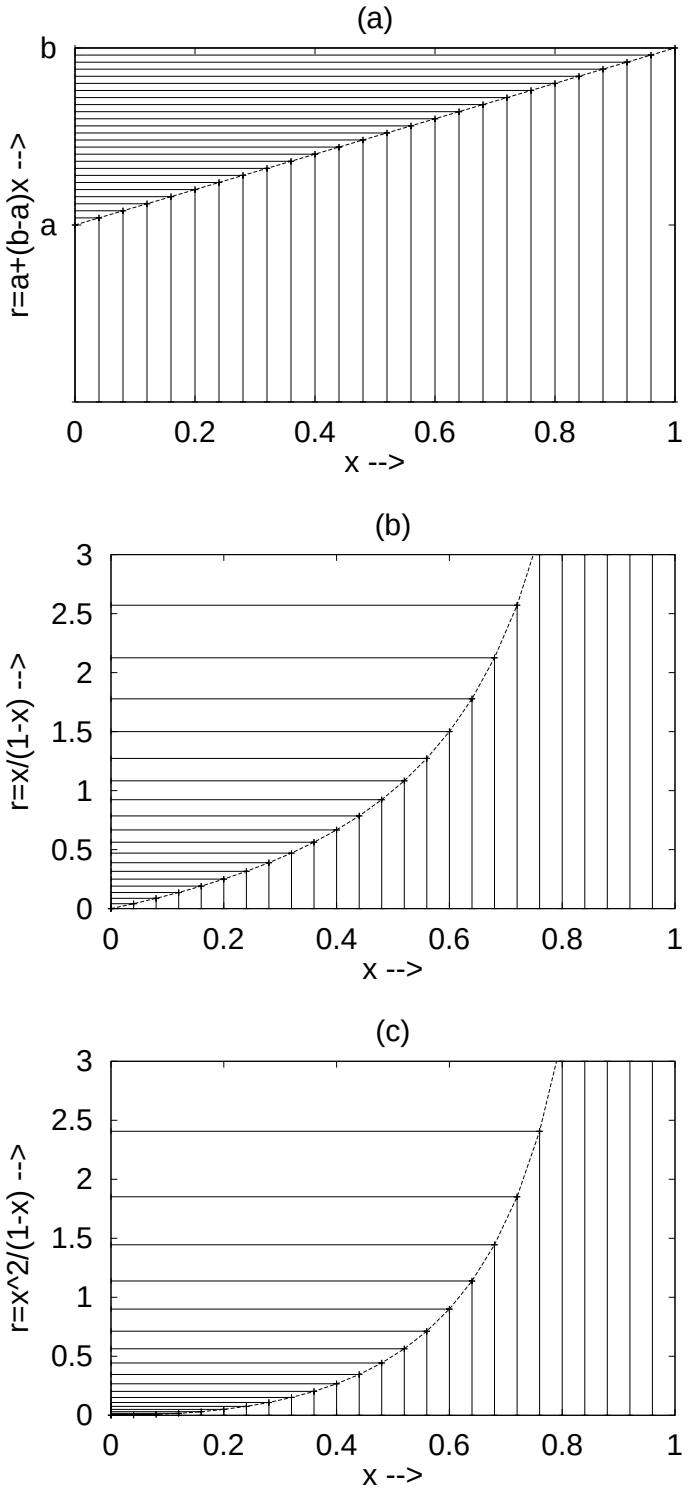


Figure 4: Mapping of regular sampling of unit interval under change of variables: (a) linear map, (b) map to semi-infinite interval, (c) map to semi-infinite interval with concentration of points at origin.

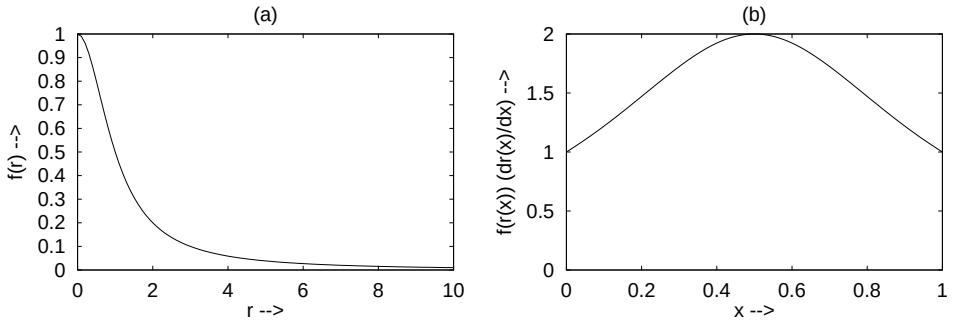


Figure 5: Integrand of $\int_0^\infty \frac{1}{1+r^2}\,dr$: (a) direct integration, (b) under change of variables in Eq. (21).

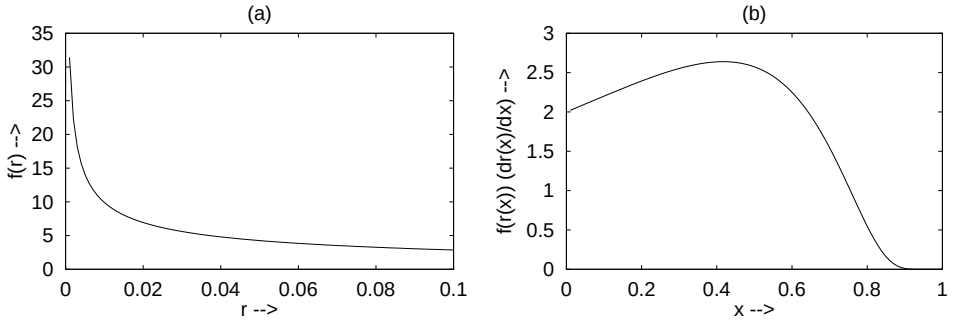


Figure 6: Integrand of $\int_0^\infty \frac{e^{-r}}{\sqrt{r}}\,dr$: (a) direct integration, (b) under change of variables in Eq. (22).

which Figure 5 illustrates to be much more manageable. In practice, such changes of variables require care to ensure that the proper limiting values for $f(r(x))(dr/dx)$ are used at the endpoints where the change of variables factor $dr/dx = 1/(1-x)^2$ becomes singular.

Finally, as an example of a problem with an integrable singularity consider the integral

$$\int_0^\infty \frac{e^{-r}}{\sqrt{r}} = \sqrt{\pi}.$$

As $r \to 0$, the integrand approaches $1/\sqrt{r}$ and diverges toward infinity. (See Figure 6(a).) Heuristically, the solution is to concentrate sample points $r(x_i)$ near the singularity at the origin. A map which both extends the range of integration to infinity and concentrates points near the origin is

$$\begin{aligned} r(x) &= \frac{1}{1-x} - 1 - x = \frac{x^2}{1-x} \\ \frac{dr(x)}{dx} &= \frac{1}{(1-x)^2} - 1, \end{aligned} \tag{22}$$

(See Figure 4(c).) Figure 6(b) shows that the resulting integrand $f(r(x))(dr(x)/dx)$ is now well behaved and quite suited to regular interval techniques.