22 Numerical Integration

Most integrals cannot be computed analytically. Some of the most important numerical integration algorithms are inseparably connencted to the theory of orthogonal polynomials. Also discussed are the effectiveness of the simple trapezoidal rule and high-dimensional integrals.

Key words: Gauss schemes, IMT formula, DE formula, quasi-Monte Carlo method, Monte Carlo method

Summary:
(1) Roughly speaking, Gauss formulas are versatile and useful. Probably, up to 4 or 5-tuple integrals, direct use of the scheme may be practical. (→**22A.3**, **22A.5**, **22A.6**).
(2) However, if a very accurate integration is needed, variable transformation schemes should be used, esp., the DE formula (→**22B.2**).
(3) If the integration is over a moderately high ($\sim$ 10) dimensional region, then quasi-Monte Carlo method **22C.5** should be considered first with the conditioning of the function according to **22C.2**. I fhte dimension is higher, then currently no better versatile method than the Monte Carlo method is known**22C.6**.

22.A Gauss Formulas

22A.1 Numerical integration. Simple numerical integration methods as the trapezoidal rule or Simpson's rule has the following general structure

$$\int_{-1}^{1} f(x)dx \simeq \sum_{v=1}^{N} C_v f(\frac{v}{N}) \quad \text{(the general Newton-Cotes formula).} \tag{22.1}$$

We have N freedom to choose C_v. Hence, it is possible to choose them so that the formula is exact for $f(x) = 1, x, \ldots, x^{N-1}$ ((cf. Weierstrass' theorem ß**17.3**). Gauss pointed out that there is no necessity to choose equidistant points v/N to sample the function values. See the following example.

22A.2 Simple demonstration. We choose $N = 2$:

$$\int_{-1}^{1} f(x)dx \sim C_1 f(x_1) + C_2 f(x_2) \tag{22.2}$$

We choose C_i and x_i so that the formula is exact for $f = 1, x, x^2$ and x^3. We have four formulas

$$\begin{aligned} 1 &: 2 = C_1 + C_2, \\ x &: 0 = C_1 x_1 + C_2 x_2, \\ x^2 &: 2/3 = C_1 x_1^2 + C_2 x_2^2, \\ x^3 &: 0 = C_1 x_1^3 + C_2 x_2^3. \end{aligned}$$

From these equations, we solve as

$$C_1 = C_2 = 1,$$
$$x_1 = -x_2 = 1/\sqrt{3}.$$

Therefore, the $N = 2$ *Gauss-formula* (G2) is

$$\int_{-1}^{1} f(x)dx \simeq f\left(\frac{1}{\sqrt{3}}\right) + f\left(-\frac{1}{\sqrt{3}}\right). \tag{22.3}$$

If we need the integration

$$I = \int_a^b \phi(u)du, \tag{22.4}$$

introduce the variable x running from -1 to 1 such that

$$u = \frac{1}{2}[(b-a)x + a + b] \tag{22.5}$$

and

$$I = \frac{1}{2}(b-a)\int_{-1}^{1} \phi([(b-a)x + a + b]/2)dx. \tag{22.6}$$

Examples for (22.3) are given as[316]

	$\int_0^{\pi/2} \sin x dx$	$\int_0^1 \sqrt{x}dx$	$\int_0^1 x^{3/2}dx$	$\int_0^1 \frac{x}{e^x-1}dx$	$\int_0^1 f^*(x)dx$
exact	1	2/3	0.4	0.77751164...	0.306853...
G2	0.99848...	0.6738...	0.3987...	0.77750464...	0.2261...

Here $f^*(x) = 1/(x+2)$ for $x \in [0, e-2]$, $f^*(x) = 0$ for $x \in [e-2, 1]$.

[316]From P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, 1975); not updated but still useful.

As we see, for smooth functions the method is amazingly powerful. If we choose the 4 point formula for $I = \int_0^{\pi/2} \sin x dx$, $I = 1.000000$, correct to six decimal places. (The Simpson rule (→**22A.8**) with 64 points produces 0.99999983).

Exercise.
(1) Compute the following integral analytically:

$$\int_{-1}^{1} dx (x^2 - 1) e^{-x^2/2}. \tag{22.7}$$

Prescribe a method to compute this numerically with the aid of (only) G2 with the relative error of 10^{-5}.
(2) Construct the $N = 2$ Gauss formula for the integral of range $[-1, 1]$ with the weight $e^{-|x|}$. Apply it to $\cos x$ and compare the result with the ordinary Gauss-Legendre formula with $N = 2$ applied to $e^{-|x|}\cos x$ on $[-1, 1]$.
(3) Compute

$$\int_0^{\pi/2} \cos x \, sgn(\pi/4 - x) dx \tag{22.8}$$

to the relative accuracy of 10^{-4} using only G2. In this case if G2 is naively used for the whole inteval, the error is about 20%.

22A.3 Fundamental theorem of Gauss quadrature. Let $w(x)$ be a weight function for the interval $[a, b]$. Then, there exist real numbers $x_1, \ldots, x_N$ and $C_1, \ldots, C_N$ with the following properties
(i) $a < x_1 < x_2 < \cdots < x_N < b$,
(ii) $C_k > 0$ for $k = 1, 2, \cdots, N$,
(iii)

$$\int_a^b f(x) w(x) dx = \sum_{k=1}^{N} C_k f(x_k) \tag{22.9}$$

is exact for every polynomial $f(x)$ of degree not more than $2N - 1$. □
Actually, $x_1 \ldots, x_N$ are the zeros of p_N, the N-th member of the orthogonal polynomial family on $[a, b]$ with the weight $w(x)$ (→**21A.2**), and

$$C_k = \int_a^b \frac{p_N(x) w(x) dx}{p_N'(x)(x - x_k)} \quad (k = 1, \ldots, N).$$

□
For example, for $\int_{-1}^{1} f(x) dx$, $p_N(x) = \sqrt{\frac{2N+1}{2}} P_N(x)$ (→**21A.5**) so that the scheme is called the Gauss-Legendre formula.
[Demo] We demonstrate the theorem for $L_2([-1, 1])$, the most important case. Let f be an m-th order polynomial, and the desired integration formula is given by

$$\int_{-1}^{1} f(x) dx = \sum_{k=1}^{N} C_k f(x_k), \tag{22.10}$$

as in (iii). Here the fact (→**21A.11**) that the zeros of orthogonal polynomials are all in its domain has been fully utilized. Notice that f can be uniquely decomposed as

$$f = P_n Q + R, \tag{22.11}$$

where P_n is the n-th order Legendre polynomial, and R is a polynomial of order less than n. Since the order of Q is $m-n$, if $m-n \le n-1$ (i.e., $m \le 2n-1$), then P_n is orthogonal to Q (→**21A.3(1)**). Hence, for $m \le 2n-1$, we conclude

$$\int_{-1}^{1} f(x)dx = \int_{-1}^{1} R(x)dx. \tag{22.12}$$

According to our formula (22.10), we have

$$\int_{-1}^{1} f(x)dx = \sum_{k=1}^{N} C_k P_n(x_k)Q(x_k) + \sum_{k=1}^{N} C_k R(x_k). \tag{22.13}$$

Therefore, we immediately see that if we can choose x_k to be the zeros of P_n, then the first term on RHS vanishes. That is, (22.12) is true for our formula under construction. For this to be true, we need to set $n = N$ (→**21A.11**) and $m = 2N-1$. We have fixed the sampling point locations. If we can choose C_k so that (22.12) holds exactly for all the $N-1$ order polynomials, then we can integrate all the polynomials up to the order $2N-1$ exactly by our integration formula. Therfore, the remaining task is to determine C_k so that

$$\int_{-1}^{1} R(x)dx = \sum_{k=1}^{N} C_k R(x_k) \tag{22.14}$$

is exact for any choice of $N-1$ order polynomial R. Notice that generally we can write

$$R(x) = \sum_{k=1}^{N} R(x_k) l_k(x), \tag{22.15}$$

where[317]

$$l_k(x) = \prod_{i \ne k}^{n} \left(\frac{x - x_i}{x_k - x_i} \right). \tag{22.16}$$

Hence, the following choice solves our problem:

$$C_k = \int_{-1}^{1} l_k(x)dx. \tag{22.17}$$

Since $l_k(x)(x - x_k)$ is proportional to P_N (all the zeros are common!),

$$C_k = \int_{-1}^{1} \frac{P_N(x)}{(x-x_k)P_N'(x_k)}dx = \frac{2}{NP_{N-1}(x_k)P_N'(x_k)} = \frac{2(1-x_k^2)}{[NP_{N-1}(x_k)]^2}. \tag{22.18}$$

[317]This is the standard *Lagrange interpolation formula.*

Exercise.
Demonstrate the formula for the weight of the Gauss-Legendre formula:

$$C_k = \frac{2(1-x_k^2)}{[NP_{N-1}(x_k)]^2}. \tag{22.19}$$

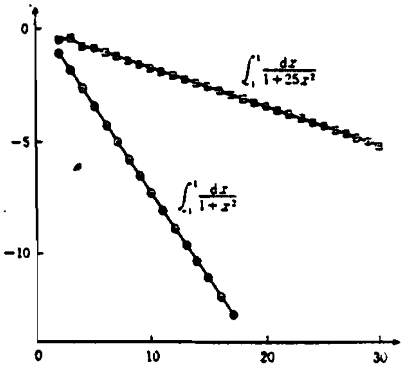
[Hint.

22A.4 Error estimate of Gauss formulas.
(1) If f is at least $2N$ times continuously differentiable (i.e., in C^{2N}), then the integration (on $[-1,1]$) error is bounded by

$$|\text{error}| \le \frac{2^{2N+1}(N!)^4}{(2N+1)((2N)!)^3} \max_{x\in[-1,1]} |f^{(2N)}(x)|. \tag{22.20}$$

(2) If f is holomorphic ($\to$**5.4**) in $\Omega \equiv \{z \,|\, |z+1| + |z-1| = \rho + \rho^{-1}\}$ for $\rho > 1$, then

$$|\text{error}| \le \frac{\pi(\rho+\rho^{-1})}{\rho^{2N+1}} \max_{z\in\Omega} |f(z)|. \tag{22.21}$$



(from 杉原·室田)

Exercise.
Calculate the following three integrals:

$$(1) \int_{-1}^{1} e^{-x^2}dx, \ (2) \int_{-1}^{1} \sin|x|dx, \ (3) \int_{-1}^{1} \cos x \, sgn(x^2 - 1/2)dx \tag{22.22}$$

with the aid of the Gauss-Legendre formulas for $N = 2, 4$, and 8 and discuss the results. (The necessary table is on p916 of Abramowitz and Stegun).

22A.5 How to get the weights. Abscissa and weight factors are tabulated in, e.g., Abramowitz-Stegun, *Handbook of Mathematical Functions* (Dover, 1972), but it is recommended to compute them to avoid any transcription mistakes.

22A.6 Many dimension. We can of course extend the formula for many dimensional cases. [See Davis & Rabinowitz Chapter 5]. For example, a singular integral like

$$\int_{-1}^{1}\int_{-1}^{1} dxdy \frac{1}{1-xy}$$

can be accurately calculated <u>without</u> any special considerations on the singularities.

22A.7 Integral equation solver. The Gauss method may be the best general numerical method to solve integral equations.

22A.8 Trapezoidal vs. Simpson rule[318] Let

$$I_e = 2h\left\{\sum_{r=1}^{n-1} f(a+2rh) + \frac{1}{2}[f(a)+f(a+2nh)]\right\}, \quad (22.23)$$

$$I_o = 2h\left\{\sum_{r=1}^{n-1} f(a+(2r+1)h)\right\}. \quad (22.24)$$

To compute the following integral

$$I = \int_a^{a+2nh} f(x)dx, \quad (22.25)$$

the trapezoidal rule uses

$$I \simeq \frac{1}{2}(I_e + I_o), \quad (22.26)$$

and the Simpson rule uses

$$I \simeq \frac{1}{3}(I_e + 2I_o). \quad (22.27)$$

Usually, it is believed that the Simpson rule is superior to the trapezoidal rule. However, this is not always the case. If

$$I = \int_a^b f(x)dx = \int_{a+h}^{b+h} f(x)dx, \quad (22.28)$$

where h is the increment of integration, then the trapezoidal rule is superior to the Simpson rule. If f vanishes or becomes very small (like $\exp(-x^2)$) outside the domain sufficiently inside $[a,b]$, or if f is a periodic function and $[a,b]$ is a period, then (22.28) hold. [See **22A.9** for the computation of Fourier coefficients.] The purpose of the modification in the Simpson rule is to eliminate the end effect of the integration range. This is why the trapezoidal rule can be better if there is no end effect. Therefore, the Simpson rule is better than the trapezoidal rule, when (22.28) does not hold.

22A.9 Discrete Fourier transform I. Let

$$a_n = \frac{1}{N}\sum_{k=0}^{2N} X_k \cos\left(\frac{nk\pi}{N}\right), \quad (22.29)$$

$$b_n = \frac{1}{N}\sum_{k=0}^{2N} X_k \sin\left(\frac{nk\pi}{N}\right). \quad (22.30)$$

[318] This section is based on an essay by H. Takahashi, 'Superposition in numerical integration,' Sugaku Seminar, March 1971.

Then,

$$X_k = \frac{1}{2}(a_0 + a_N \cos k\pi)) \sum_{n=1}^{N-1} \left\{ a_n \cos\left(\frac{mnk\pi}{N}\right) + b_n \sin\left(\frac{mnk\pi}{N}\right) \right\}, \tag{22.31}$$

if $X_k = f(k\pi/N)$, then (22.30) is obtained from the standard formulas for Fourier coefficients through 'approximating' the integrals with the aid of the trapezoidal rule. However, notice that the formulas are exact. This is an example of the merit of the trapezoidal rule for periodic functions.

22A.10 Discrete Fourier transform II. Let $\boldsymbol{X} \equiv \{X_n\}_{n=0}^{N-1}$ be a sequence of complex numbers, and

$$e(x) \equiv \exp(-2\pi i x). \tag{22.32}$$

The following sequence $\hat{\boldsymbol{X}} \equiv \{X^n\}$ is called the *discrete Fourier transform* of $\boldsymbol{X}$:

$$X^k = \sum_{n=0}^{N-1} e\left(\frac{kn}{N}\right) X_n. \tag{22.33}$$

Its inverse transformation is given by

$$X_n = \frac{1}{N} \sum_{k=0}^{N-1} e\left(\frac{-kn}{N}\right) X^k. \tag{22.34}$$

Cf. **32B.12**.

22.B Variable Transformation Schemes

22B.1 Functions of double exponential decay. If f is an analytic function, then the trapezoidal rule gives an excellent result for the integral over $\boldsymbol{R}$. This seems to be a well known fact. If the integrand decays double exponentially, i.e., for some positive constants A, B and C

$$|f| \sim A \exp(-B \exp(Cx)) \tag{22.35}$$

The error of the trapezoidal rule truncated at N

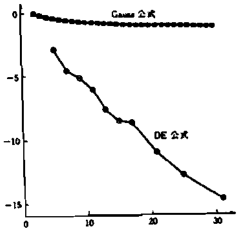
$$T_h = h \sum_{k=-N}^{N} f(kh) \tag{22.36}$$

for the integral of f from $-\infty$ to $+\infty$ is given by

$$|T_h - I| \leq const. \|f\| \exp(-\tilde{C}N/\ln N) \tag{22.37}$$

for some positive $\tilde{C}$. This means that if N is doubled, then the number of the significant digits doubles.

22B.2 Double exponential (DE) formula. The DE formula was proposed by Takahashi[319] and Mori in 1974, and is regarded the most effective integration formula currently. The essence is to change the independent variable so that the function decays double-exponentially. For example, for the integral of an analytic function f on $[-1, 1]$



from 杉原, 室田

$$x = \phi(t) \equiv \tanh\left(\frac{\pi}{2}\sinh t\right) \tag{22.38}$$

and the DE formula reads[320]

$$\int_{-1}^{1} f(x)dx \simeq h \sum_{k=-N}^{N} f(\phi(hk))\phi'(hk). \tag{22.39}$$

However, the DE formula is not effective for the integrals of Fourier transformation type.

Discussion.

The DE formula is powerful even for an integrand with end singularities:

$$I \equiv \int_{-1}^{+1} dx(1-x^2)^{-1/2} = \pi \tag{22.40}$$

If we use the Gauss-Legendre formula to this, the error is never less than 10^{-2} for $n \leq 30$. The DE formula with 5 terms is already with only less then 1% error. With 10 points, the error is about 10^{-6}. With $n = 30$ the error is about 10^{-15}. The improvement is roughly exponential $10^{-n/2}$. [This is in conformity with the theoretical error estimate.]

22B.3 Numerical estimate of Fourier transform. For

$$\int_0^{\infty} f(x) \sin\left(\frac{\pi(x-\alpha)}{\lambda}\right) dx \tag{22.41}$$

---

[319]This is the same person of the 'Takahashi gas', proving that there is no phase transition in 1-space with short range interactions. He is the most creative statistical physicist Japan has ever produced when he was young, but later became the leader of computer research in Japan, saying physics was his hobby.

[320]H Takahashi and M Mori, Publ. RIMS **9**, 721 (1974).

the following transformation is effective:[321]

$$\psi(t) = \frac{t}{1 - \exp(-2\pi \sinh t)}. \tag{22.42}$$

The formula reads

$$\int_0^\infty f(x) \sin\left(\frac{\pi(x-\alpha)}{\lambda}\right) dx \simeq \lambda \sum_{k=-N}^{N} g\left(\frac{\lambda}{h}\psi\left(\frac{h(k\lambda+\alpha)}{\lambda}\right)\right)\psi'\left(\frac{h(k\lambda+\alpha)}{\lambda}\right), \tag{22.43}$$

where $g(x) = f(x)\sin[\pi(x-\alpha)/\lambda]$.

22.C Multidimensional Integrals

22C.1 Overview. An immediate idea is to use the one dimensional formulas repeatedly (direct product scheme). Other interesting methods are the Monte Carlo or quasi-Monte Carlo methods. These latter methods are characterized by the error estimate which is independent of the spatial dimensionality but dependent only on the number of sampling points. Here we discuss only two methods for very large dimensions. The quasi Monte Carlo method is becoming increasingly important, because the error improves as $1/N$ instead of $1/\sqrt{N}$. However, there seems to be no versatile general scheme applicable to all the cases. This is a very active field of research esp., in relation to finance.

22C.2 Polynomial variable transformation: recommended preconditioning. Let p be an integer not less than 2. If a function $f(\{x_i\})$ has continuous partial derivatives

$$\frac{\partial^{j_1+\cdots+j_s} f(x_1,\cdots,x_s)}{\partial x_1^{j_1}\cdots\partial x_s^{j_s}} \tag{22.44}$$

for all $j_1,\cdots,j_s \in \{0,1,\cdots,p\}$, then we can use the following transformation

$$x_i = \phi(y_i) \equiv \frac{(2p+1)!}{(p!)^2}\int_0^{y_i} u^p(1-u)^p du \tag{22.45}$$

to convert the integrand f to

$$f(\phi(y_1),\cdots,\phi(y_s))\phi'(y_1)\cdots\phi'(y_s) \tag{22.46}$$

[321] T. Ooura and M. Mori, J. Comp. Appl. Math. **38**, 353-360 (1991).

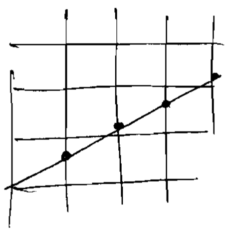
whose multidimensional Fourier coefficients have the following estimate:

$$c_k \leq const \times |k|^{-p}. \tag{22.47}$$

With this smoothness condition (→**17.12**), many integration formulas become more effective than without the transformation. Thus usually, it is recommended to transform the integrand with the aid of this transformation prior to application of integration schemes.

22C.3 Weyl's equidistribution theorem. If α is irrational, then for any $0 \leq a \leq b \leq 1$ we have

$$\frac{1}{N}\#\{n \mid \{n\alpha\} \in [a,b], n \in \{1,2,\cdots,N\}\} \to |b-a|. \tag{22.48}$$



Here $\{a\} = a - [a]$ is the fractional part of a, and $\#A$ is the number of members (the cardinality) of the set A. We will not give any proof for this,[322] but this should be intuitively clear, if the reader imagines a particle geodesically moving (i.e., going straight) on the 2-torus, and $[0,1]$ is the coordinate of its section (the so-called Poincarè section in the theory of dynamical systems). A multidimensional version should not be hard to formulate and understand in a similar fashion. Thus we get

22C.4 Theorem [Weyl]. Let $1, \alpha_1, \cdots, \alpha_s$ be rationally independent.[323] Then,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} f(\{k\alpha_1\},\cdots\{k\alpha_s\}) = \int_{[0,1]^s} f(\{x\})d\{x\}. \tag{22.49}$$

22C.5 Improved Haselgrove method.[324]

$$\int_{[0,1]^s} f(\{x\})d\{x\} \simeq \frac{1}{N}\sum_{k=1}^{N} w_q(k/N)f(\{k\alpha_1\},\cdots\{k\alpha_s\}), \tag{22.50}$$

where

$$w_q(x) = \frac{(2q+1)!}{(q!)^2}x^q(1-x)^q. \tag{22.51}$$

The representative irrational numbers $\alpha_1, \cdots, \alpha_s$ are chosen (semi-empirically) as

[322] See Section 3 of Körner.

[323] That is, there are no integers $p_0, p_1, \cdots, p_s$ (not all of them are simultaneously equal to 0) such that $p_0 + \sum p_k\alpha_k = 0$

[324] M. Sugihara and K. Murota, Math. Computation **39**, 549-554 (1982).

(1) If $s+3$ is a prime, then $\alpha_j = 2\cos(2j\pi/(2s+3))$,
(2) Otherwise, $\alpha_j = 2^{j/(s+1)}$.
w_q is introduced to reduce the error further. A detailed error estimate is available, but the main features of the error is that it is bounded by the number proportional to N^{-q}.

22C.6 Monte Carlo method. To compute

$$I \equiv \int_{[0,1]^s} f(x_1,\cdots,x_s)dx_1\cdots dx_s \tag{22.52}$$

the Monte Carlo method randomly and uniformly samples points in the cube $[0,1]^s$ as $y_1, y_2, \cdots$ and claim

$$S_N = \frac{1}{N}\sum_{k=1}^{N} f(y_k) \to I. \tag{22.53}$$

The principle should be understandable from the random analogue of **22C.3**.

Its error can be estimated with the aid of Chebychev's inequality[325] as

$$\text{Probability}(|I - S_N| \geq 2/\sqrt{\epsilon N}) \leq \epsilon \tag{22.54}$$

for f such that $|f| \leq 1$.
For example, if $N = 10^6$, then with probability 99% we can get the answer with 2% relative error independent of the dimension of the space! However, the accuracy improves only as $N^{-1/2}$.

Exercise.
(1) We wish to compute

$$\int_{-1}^{1}\cdots\int_{-1}^{1} e^{-(x_1+x_2+\cdots+x_N)^2/N}dx_1\cdots dx_N \tag{22.55}$$

with the aid of the Monte Carlo method. How many samples do we need conservatively to obtain the integral with 5% relative error with probability 99.9%?
(2) We wish to compute the following integral by the Monte Carlo method:

$$I = \int_D dx_1\cdots dx_{100} r(1-r), \tag{22.56}$$

where $r = \sqrt{\sum_{i=1}^{100} x_i^4}/5$, and the domain D is the 100 dimensional hypercube $[0,1]\times\cdots\times[0,1]$. How many sample points are (conservatively) needed, if we wish to get I with the error less than 2% with the probability more than 99.5%?

[325] $a^2\text{Probability}(|x| \geq a) \leq \langle x^2\rangle$, which can be derived easily from the obvious inequality $x^2 \geq a^2\Theta(|x| \leq a)$.

(3) **Generalization of the Chebychev inequality**. Let f be a positive function,[326] and $\varphi_A \equiv \inf_{x \in A} \varphi(x)$. Then,

$$\varphi_A \text{Probability}(X \in A) \le \langle \varphi \rangle. \tag{22.57}$$

The inequality we have used is a special case with $\varphi = x^2$.

[326] Measurable w.r.t. the probability measure under consideration (→**19a**).