PARTIAL SOLUTION TO
LAST ISSUE'S HOMEWORK ASSIGNMENT

MULTIDIMENSIONAL INTEGRATION: PARTITION AND CONQUER

By Dianne P. O'Leary

IN THE LAST ISSUE, WE INVESTIGATED VARIOUS MONTE CARLO INTEGRATION METHODS ON A SAMPLE PROBLEM—FINDING THE AREA OF A QUARTER CIRCLE—AND ON A PARTITION FUNCTION.

PROBLEM 1.

Suppose we want to estimate the area of a quarter circle with radius r:

$$I = \int_0^r \sqrt{r^2 - x^2}\,dx = \int_0^r f(x)dx.$$

Let $r = 0.5$ and use two methods:

- Method 1, testing whether randomly generated points in the rectangle $[0, r] \times [0, r]$ are inside or outside the quarter circle Ω and multiplying the fraction inside by r^2, the area of the rectangle.
- Method 2, computing the average value of $f(x)$ and multiplying by r, the length of $\Omega = [0, r]$.

Compare the quality of the two estimates for 10, 100, 1,000, 10,000, and 100,000 points by measuring the error and the convergence rate.

How many function evaluations does your favorite integration routine use to get an estimate of comparable quality?

Answer:

Sample code is given on the *CiSE* Web site (www.computer.org/cise/homework). Method 2 gives somewhat better results because it averages the function values themselves rather than just using them to decide whether a point is inside or outside the region. Three-digit accuracy is achieved for 100,000 points in Method 1 and for 1,000 and 100,000 points for Method 2. The convergence rate for Method 1 is consistent with $1/\sqrt{n}$, because the product of the error with $\sqrt{n}$ is approximately constant for large n, but for Method 2, the results are somewhat more variable. Matlab's function `quad` uses 13 function evaluations to get three-digit accuracy.

Clearly, for low-dimensional integration of smooth functions, Monte Carlo methods are not the methods of choice. Their value becomes apparent only when the dimension d is large so that methods like `quad` would be forced to use a lot of function evaluations.

PROBLEM 2.

Compute new estimates of the integral of Problem 1 using Method 3, importance sampling, taking 10 samples of $f(x)$ to determine the function $p(x)$. Compare with your previous results.

Answer:

Sample code is available on the *CiSE* Web site. Importance sampling produces better estimates at lower cost: see the answer to Problem 4 for detailed results.

PROBLEM 3.

Generate 500 pseudorandom points $\mathbf{v}_p$ in $\Re^2$ and 500 quasirandom points $\mathbf{w}_p$ in $\Re^6$ ($p = 1, \ldots, 500$). Plot the pseudorandom points. Then for the quasirandom points, make a plot of the first two coordinates, a separate plot of the third and fourth coordinates, and a final plot of the fifth and sixth coordinates. Discuss the desirability of using each of these four choices for "random" points.

Answer:

Figure A shows the results. The pseudorandom points from Matlab's `rand` are designed to have good statistical properties, but they leave large gaps in space. The quasirandom points are both more predictable and more evenly distributed: they tend to lie on diagonal lines, with longer strings as the coordinate number increases. Other algorithms for generating quasirandom points avoid this defect.

PROBLEM 4.

Compute new estimates of the integral of Problem 1 using quasirandom numbers in Method 2 instead of pseudorandom numbers. Compare the results.

Authorized licensed use limited to: to IEEExplore provided by Virginia Tech Libraries. Downloaded on October 24, 2008 at 09:43 from IEEE Xplore. Restrictions apply.

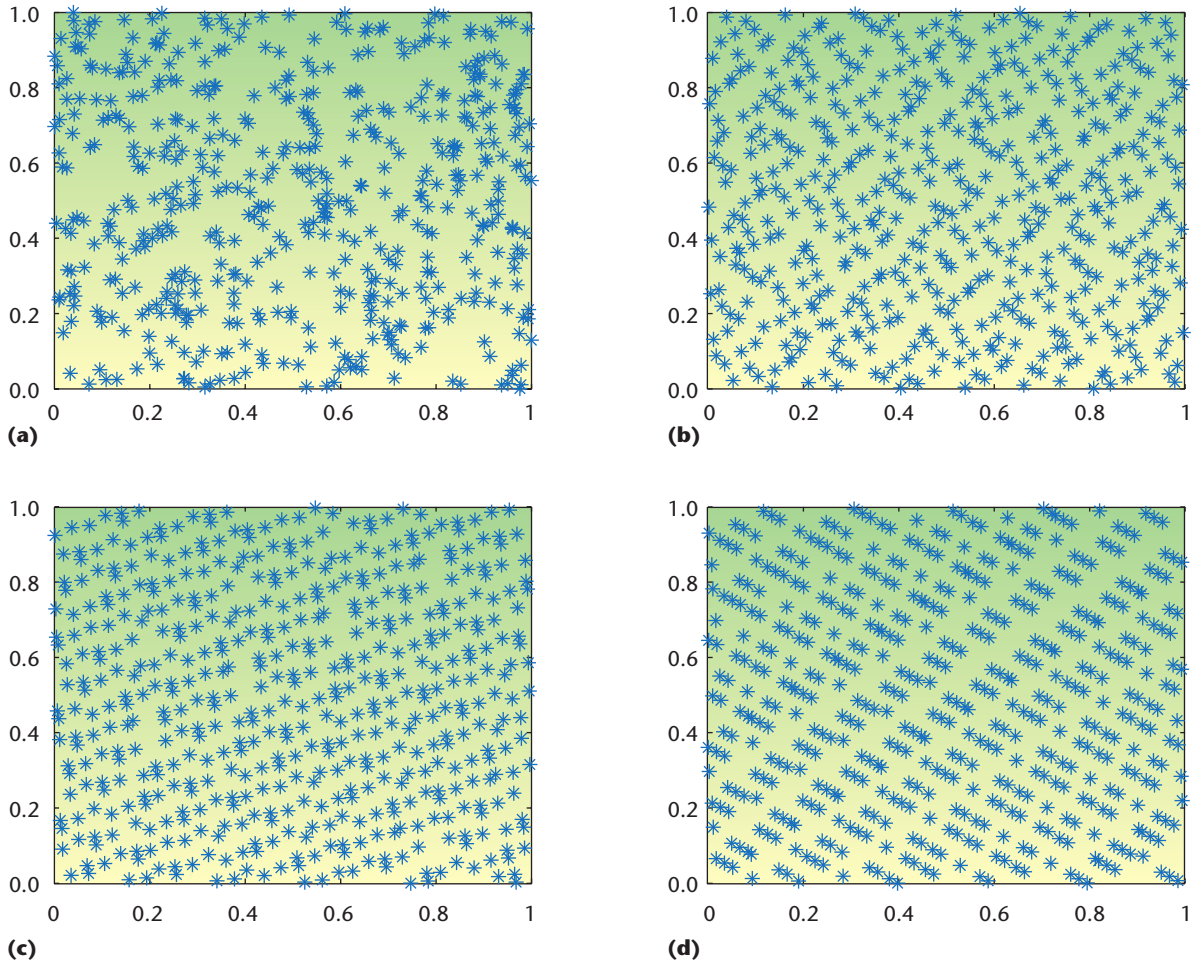


Figure A. The answer to Problem 3. (a) 500 pseudorandom points, and (b) the first two quasirandom coordinates, (c) the second two, and (d) the third two.

Answer:

Table 1 gives the absolute value of the errors in the estimates from each of the four methods. We can determine the convergence rates from the slope of a straight line fit to the logs of each set of errors. We get the best results with Method 4, using quasirandom numbers in Method 2. Method 3, importance sampling, was also quite good.

Consider a chain of particles in which each particle interacts with its one or two neighbors on the chain. For this set of particles, the partition function is

$$Z_d(L) = \int_{-\infty}^{\infty} \rho(a, a, L)da, \quad (1)$$

where $L = 1/(\alpha T)$, T is the temperature, α is the Boltzmann constant, d is the number of particles, and ρ is

$$\rho_d(a, b, L) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(a, x_1, L)g(x_1, x_2, L) \ldots g(x_d, b, L)dx_1dx_2 \ldots dx_d,$$

where

$$g(x, y, L) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{1}{2\delta}(x-y)^2 - \frac{1}{2}\delta\left(V(x)+V(y)\right)\right),$$

with $\delta = L/(d+1)$.

PROBLEM 5.

a. Let $L = 1$ and $k = 1.38$ Angstroms^2 g/sec^2 K and consider the harmonic oscillator potential $V(x)$. Determine finite integration limits for $a, x_1, \ldots, x_d$ so that the partition function $Z_d(L)$ in Equation 1 can be approximated to three-digit accuracy. (Do this by bounding the neglected part of the integral.)

b. Use your favorite one-dimensional integration routine to estimate the partition function $Z_d(L)$. When you need a

Authorized licensed use limited to: to IEEExplore provided by Virginia Tech Libraries. Downloaded on October 24, 2008 at 09:43 from IEEE Xplore. Restrictions apply.

Table 1. Absolute value of estimated errors.

n	Method 1	Method 2	Method 3	Method 4
10	3.65e-03	1.11e-02	4.67e-03	1.50e-02
100	1.35e-03	3.38e-03	1.02e-03	2.49e-03
1,000	2.85e-03	2.38e-04	1.22e-05	3.00e-04
10,000	1.57e-03	1.14e-03	1.75e-04	4.10e-05
100,000	4.97e-04	1.72e-04	1.44e-05	5.14e-06

function value $\rho(a, a, L)$, use Monte Carlo integration to obtain it. Try n = 100, 1,000, 10,000, 100,000, 1,000,000 (if possible), and d = 1, 2, 4, 8, 16.

c. Repeat the experiment using quasirandom points in the Monte Carlo integration. What can you say about the accuracy and convergence rate of your estimates?

Answer:

There are several ways to get bounds. The one used here is a variation on Mei Huang's ideas.

For (a), let $x_0 = x_{d+1} = a$ and let $\mathbf{x} = [x_0, \ldots, x_d]^T$ and $\boldsymbol{P}\mathbf{x} = [x_1, \ldots, x_{d+1}]^T$. Note that

$$g(a, x_1, L)g(x_1, x_2, L) \ldots g(x_d, a, L)$$

$$= \left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\frac{1}{2\delta}\sum_{i=1}^{d+1}(x_i - x_{i-1})^2 - \frac{\alpha\delta}{4}\left(\sum_{i=1}^{d+1}x_i^2 + \sum_{i=0}^{d}x_i^2\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\frac{1}{2\delta}\left(2\|\mathbf{x}\|^2 - 2\mathbf{x}^T\boldsymbol{P}\mathbf{x}\right) - \frac{\alpha\delta}{2}\|\mathbf{x}\|^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\left(\frac{1}{\delta} + \frac{\alpha\delta}{2}\right)\|\mathbf{x}\|^2 + \frac{1}{\delta}\mathbf{x}^T\boldsymbol{P}\mathbf{x}\right).$$

Now, because $\|\mathbf{x}\| = \|\boldsymbol{P}\mathbf{x}\|$ and because $\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\ \|\mathbf{y}\| \cos(\theta)$, where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$, we see that $\mathbf{x}^T\boldsymbol{P}\mathbf{x} \leq \mathbf{x}^T\mathbf{x}$, so

$$g(a, x_1, L)g(x_1, x_2, L) \ldots g(x_d, a, L)$$

$$\leq \left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\left(\frac{1}{\delta} + \frac{\alpha\delta}{2}\right)\|\mathbf{x}\|^2 + \frac{1}{\delta}\mathbf{x}^T\mathbf{x}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\frac{\alpha\delta}{2}\|\mathbf{x}\|^2\right).$$

Now if we integrate the function $g(a, x_1, L)\, g(x_1, x_2, L) \ldots g(x_d, a, L)$ over the region $\Omega = [-b, b]^{d+1} \subset \mathcal{R}^{d+1}$ instead of $\mathcal{R}^{d+1}$, the error is bounded by

$$\int_{\mathcal{R}^{d+1}-\Omega} g(a, x_1, L)g(x_1, x_2, L) \ldots g(x_d, a, L)d\mathbf{x}$$

$$\leq \int_{\mathcal{R}^{d+1}-\Omega}\left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\frac{\alpha\delta}{2}\|\mathbf{x}\|^2\right)d\mathbf{x}$$

$$= \int_{\mathcal{R}^{d+1}}\left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\frac{\alpha\delta}{2}\|\mathbf{x}\|^2\right)d\mathbf{x} - \int_{\Omega}\left(\frac{1}{\sqrt{2\pi\delta}}\right)^{d+1} \exp\left(-\frac{\alpha\delta}{2}\|\mathbf{x}\|^2\right)d\mathbf{x}$$

$$= \left(\frac{1}{\sqrt{2\pi\delta}}\int_{-\infty}^{\infty}\exp\left(-\frac{\alpha\delta}{2}x^2\right)dx\right)^{d+1} - \left(\frac{1}{\sqrt{2\pi\delta}}\int_{-b}^{b}\exp\left(-\frac{\alpha\delta}{2}x^2\right)dx\right)^{d+1}.$$

We can now use the fact that these integrals are closely related to the integral for the normal distribution:

$$F(x) = \frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{x}\exp\left(-t^2/(2\sigma^2)\right)dt,$$

with $\sigma^2 = 1/(\alpha\delta)$, so we see that

$$\int_{\mathcal{R}^{d+1}-\Omega} g(a, x_1, L)g(x_1, x_2, L) \ldots g(x_d, a, L)\, d\mathbf{x} \leq (\delta\sqrt{\alpha})^{d+1}(1 - (F(b)) - F(-b))^{d+1}) \equiv e(\delta, \alpha, d, b).$$

Using tables or computer functions of the normal distribution function, we can compute this bound.

For (b) and (c), visit the *CiSE* Web site for a sample program:

- Getting a good solution to this problem via Monte Carlo integration is difficult. It's very easy to underestimate the answer if you spend a lot of function evaluations on regions that have function values close to zero.
- For efficiency and to avoid intermediate underflow, the program should use an expression like `exp(a+b)` instead of `exp(a)*exp(b)`.

Authorized licensed use limited to: to IEEExplore provided by Virginia Tech Libraries. Downloaded on October 24, 2008 at 09:43 from IEEE Xplore. Restrictions apply.

Importance sampling, using a normal distribution, would be a better approach to this problem, but the best approach uses the partition-and-conquer method discussed in the last issue.

Acknowledgments

I'm grateful to Mei Huang for her work on Problem 5.

Dianne P. O'Leary is a professor of computer science and a faculty member in the Institute for Advanced Computer Studies and the Applied Mathematics Program at the University of Maryland. She received a BS in mathematics from Purdue University and a PhD in computer science from Stanford. She is a member of SIAM, ACM, and AWM. Contact her at oleary@cs.umd.edu; www.cs.umd.edu/users/oleary/.

ADVERTISER / PRODUCT INDEX
JANUARY/FEBRUARY 2005

Advertiser/Product	Page Number
D.E. Shaw & Company	**Cover 2**
Physics Today	**71**
SIAM	**Cover 4**
Springer	17

Boldface denotes advertisements in this issue.

Advertising Personnel

Marion Delaney
IEEE Media, Advertising Director
Phone: +1 212 419 7766
Fax: +1 212 419 7589
Email: md.ieeemedia@ieee.org

Marian Anderson
Advertising Coordinator
Phone: +1 714 821 8380
Fax: +1 714 821 4010
Email: manderson@computer.org

Sandy Brown
IEEE Computer Society,
Business Development Manager
Phone: +1 714 821 8380
Fax: +1 714 821 4010
Email: sb.ieeemedia@ieee.org

Advertising Sales Representatives

Mid Atlantic (product/recruitment)
Dawn Becker
Phone: +1 732 772 0160
Fax: +1 732 772 0161
Email: db.ieeemedia@ieee.org

New England (product)
Jody Estabrook
Phone: +1 978 244 0192
Fax: +1 978 244 0103
Email: je.ieeemedia@ieee.org

New England (recruitment)
Robert Zwick
Phone: +1 212 419 7765
Fax: +1 212 419 7570
Email: r.zwick@ieee.org

Connecticut (product)
Stan Greenfield
Phone: +1 203 938 2418
Fax: +1 203 938 3211
Email: greenco@optonline.net

Midwest (product)
Dave Jones
Phone: +1 708 442 5633
Fax: +1 708 442 7620
Email: dj.ieeemedia@ieee.org

Will Hamilton
Phone: +1 269 381 2156
Fax: +1 269 381 2556
Email: wh.ieeemedia@ieee.org

Joe DiNardo
Phone: +1 440 248 2456
Fax: +1 440 248 2594
Email: jd.ieeemedia@ieee.org

Southeast (product/recruitment)
Bob Doran
Phone: +1 770 587 9421
Fax: +1 770 587 9501
Email: bd.ieeemedia@ieee.org

Midwest/Southwest (recruitment)
Darcy Giovingo
Phone: +1 847 498-4520
Fax: +1 847 498-5911
Email: dg.ieeemedia@ieee.org

Southwest (product)
Josh Mayer
Phone: +1 972 423 5507
Fax: +1 972 423 6858
Email: josh.mayer@wageneckassociates.com

Northwest (product)
Peter D. Scott
Phone: +1 415 421-7950
Fax: +1 415 398-4156
Email: peterd@pscottassoc.com

Southern CA (product)
Marshall Rubin
Phone: +1 818 888 2407
Fax: +1 818 888 4907
Email: mr.ieeemedia@ieee.org

Northwest/Southern CA (recruitment)
Tim Matteson
Phone: +1 310 836 4064
Fax: +1 310 836 4067
Email: tm.ieeemedia@ieee.org

Japan
Sandy Brown
Phone: +1 714 821 8380
Fax: +1 714 821 4010
Email: sbrown@computer.org

Europe (product/recruitment)
Hilary Turnbull
Phone: +44 1875 825700
Fax: +44 1875 825701
Email: impress@impressmedia.com

Authorized licensed use limited to: to IEEExplore provided by Virginia Tech Libraries. Downloaded on October 24, 2008 at 09:43 from IEEE Xplore. Restrictions apply.