Global Optimization of Smooth Univariate Functions: Average Case Analysis

James M. Calvin

Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102

We consider the problem of approximating the global minimum of a twice continuously differentiable function than may have many local minima. In the average-case setting under Gaussian measures, we give bounds on the attainable convergence rates of global optimization algorithms. The bounds show that the superlinear convergence of some iterative methods for approximating local minima can not be attained by global optimization algorithms in the average case.

The continuing search for lattice rules: exploiting structured matrices.

Ronald Cools

Dept. of Computer Science, Katholieke Universiteit Leuven, Belgium

Lattice rules are a type of integration rules designed for periodic multivariate functions on a unit cube. The construction of efficient lattice rules for higher dimensions is an area of ongoing research.

At the previous FoCM meeting (Oxford, 1999) we reported the results of a large scale computer search for optimal lattices in 3 and 4 dimensions. We searched a well defined set of lattices which probably, but not certainly, contains the optimal lattices. Therefore we called our results K-optimal lattice rules. We managed to reach degree 24 in 4 dimensions, but that required a very expensive search. The fundamental problem with this search is the fact that the amount of time required grows as δ^{s^2} where δ is the degree and s is the dimension. Details of this work are described in [1].

In [2] a subset of the above class of lattices generated by a particular type of structure matrices was investigated and it turned out that many 3 and 4-dimensional K-optimal lattices belong to that class. This is a big achievement because searching in this class requires an amount of time 'only' proportional to δ^s. This talk describes ongoing research that extends this approach to 5 and 6 dimensions. The rules we obtained so far are more efficient, i.e. they use less lattice points, than previously known results of the same degree. Furthermore, searching in 6 dimensions takes only a fraction of the time we needed previously in our 4 dimensional search, for the same δ.

References

[1] R. Cools and J.N. Lyness. Three- and four-dimensional K-optimal lattice rules of moderate trigonometric degree. *Math. Comp.*, 70(236):1549–1567, 2001.

[2] J. Lyness and T. Sørevik. Four-Dimensional Lattice Rules Generated by Skew-circulant Matrices. *Math. Comp.*, To appear.

Asymptotic Equidistribution on $d \geq 2$ dimensional spheres

Steven B Damelin

Department of Mathematics, Georgia Southern University,

Post Box 8093, Statesboro, GA 30460

We study discrepancy and numerical integration of continuous functions on $d \geq 2$ dimensional spheres for points that are asymptotically equidistributed. This relates to recent work of the author with Grabner (Graz) and Saff (Vanderbilt) and Gotz (Eichstätt).

Quantum Integration in Sobolev Spaces

Stefan Heinrich

Fachbereich Informatik, Universität Kaiserslautern, D-67653 Kaiserslautern, Germany, email: heinrich@informatik.uni-kl.de

Some results presented in my plenary talk will be discussed in more detail. We are concerned with quantum algorithms for integrating multivariate functions from Sobolev spaces. These algorithms will be constucted from mean computation algorithms for finite sequences in discrete L_p spaces. We present, in particular, a general technique of connecting mean computation with integration, which is an analogon of Maiorov's discretization technique for approximation. Our technique applies to the quantum as well as to the classical settings. We discuss some consequences.

An Algorithm-Driven Approach to Error Analysis for Multidimensional Integration

Fred J. Hickernell

Department of Mathematics, Hong Kong Baptist University
Kowloon Tong, Hong Kong SAR, China

For the problem of multidimensional integration there are several questions that one can ask, such as: what is the asymptotic convergence rate of a particular algorithm? is this convergence rate optimal among all algorithms? is the convergence rate essentially independent of the dimension? In studying these questions one typically specifies a space of integrands, and then the answers depend on the space chosen. One might call this a function space-driven approach This talk proposes an alternative algorithm-driven approach. One first specifies the algorithm, and then asks what is the largest set of integrands with a desired property. The results obtained by this approach are compared to results from the function space-driven approach.

Density Functions for Average Case Inference

Mark A. Kon

Boston University, Boston, MA 02215, USA

We formulate a theory of density functions for continuous measures on infinite dimensional spaces. We show that maximum likelihood methodologies in statistics can be translated into infinite dimensional average case settings using these notions. Infinite dimensional measures can usefully be replaced by density functions or distributions on the space for the purpose of estimation. This allows use of maximum likelihood methods in the average case setting as well as in infinite dimensional nonparametric statistical models.

Worst Case Complexity of Multivariate Feynman-Kac Path Integration

Marek Kwas[a] (speaker) and Youming Li[b]

[a] Department of Computer Science, Columbia University
New York, NY 10027
Institute of Applied Mathematics and Mechanics,
University of Warsaw, Poland

[b] Department of Computer Science, University of Kentucky
Lexington, KY 40506

We study the multivariate Feynman-Kac path integration problem. This problem was studied in [1] for the univariate case. We describe an algorithm based on uniform approximation, instead of the L_2-approximation used in [1]. Similarly to [1], our algorithm requires extensive precomputing. We also present bounds on the complexity of our problem. The lower bound is provided by the complexity of a certain integration problem, and the upper bound by the complexity of the uniform approximation problem. The algorithm presented in this paper is almost optimal for the classes of functions for which uniform approximation and integration have roughly the same complexities.

References

[1] L. Plaskota, G. W. Wasilkowski, H. Woźniakowski, A New Algorithm and Worst Case Complexity for Feynman–Kac Path Integration, *J. Comp. Phys.*, **164**, 335-353 (2000)

When can sparse grid methods be efficiently used?

Youming Li

Computer Science Department
University of Kentucky
Lexington, KY 40506

Sparse grid methods are useful tools for constructing high dimensional algorithms based upon univariate algorithms. Their efficient applicability has mainly been known for problems defined over Hilbert spaces. We show that they are efficiently applicable to tensor product problems for linear functionals over Banach spaces, as well as for general operators in certain important cases. The result holds for the worst case setting.

Quasi-Monte Carlo integration over $\mathbf{R}^d$

Peter Mathé[a], and Gang Wei[b]

[a]Peter Mathé, Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, D - 10117 Berlin, Germany

[b]Gang Wei, Department of Mathematics, Hong Kong Baptist University Kowloon, Hong Kong.

We show that integrals $\int f(x)\rho(x)dx$ over R^d with a probability weight function ρ can be evaluated for a wide class of functions f using a QMC algorithm based on a proper decomposition of the domain $\mathbf{R}^d$ and arranging low discrepancy points over a series of hierarchical hypercubes. Thus we propose to decompose $\mathbf{R}^d = \bigcup_{j=0}^{m+1} I_j$, where

$$\begin{aligned} Q_j &:= \left\{x \in \mathbf{R}^d,\ |x_l| < 2^j,\ l = 1, \dots, d\right\}, j = 0, 1, \dots, m, \\ I_j &:= Q_j \setminus Q_{j-1},\ \ j = 1, \dots, m, \end{aligned}$$

with $I_0 = Q_0,\ \ I_{m+1} = Q_m^c$, and rewrite $\int f(x)\rho(x)dx = \sum_{j=0}^{m+1} \int_{I_j} f(x)\rho(x)dx$.

If ρ is bounded and smooth, and f has bounded variation and is integrable, then it can be shown that each $f(x)\rho(x)$ has bounded variation over I_j. The above decomposition can be used to base quadrature formulas by replacing the integrals by quadrature formulas using low discrepancy points, which results in

$$S_{m,\mathbf{n}}(f, \rho) := \sum_{j=0}^{m} \frac{q_j}{n_j} \sum_{i=1}^{n_j} f(y_{ij})\rho(y_{ij})\chi_{I_j}(y_{ij}),$$

where $m+1$ is any number of cubes and $\mathbf{n} := (n_0, \dots, n_m)$, where each n_j is the number of points used within cube Q_j, $j = 0, \dots, m$, and the χ_{I_j} are the indicator functions. Furthermore, $q_j :=$ Volume$(Q_j) = 2^{(j+1)d}$ and

$$\begin{aligned} y_{ij} &= -2^j 1 + 2^{(j+1)} * x_{ij},\ \ 0 \le j \le m,\ \ 1 \le i \le n_j, \\ P_j &= \{x_{ij} : 1 \le i \le n_j\} \subset [0,1)^d,\ \ 0 \le j \le m. \end{aligned}$$

For certain classes of power/exponential decaying weights ρ the algorithm is shown to be of optimal order.

The complexity of strong approximation of sde's

Thomas Müller-Gronbach

TU Darmstadt
Fachbereich Mathematik
Schloßgartenstraße 7
D-64289 Darmstadt

Please don't use any nonstandard symbols or fonts. If please let us know: remember, our secretarial deal with a large number of abstracts. In the majority of cases stochastic differential equations can not be solved explicitly so that numerical methods must be used for the approximation of the solution. Strong approximation means pathwise approximation of the solution, either globally, say, on the unit interval, or at finitely many points. First, we survey classical results, which focus on upper bounds for the error at the discretization points. Then we turn to the following natural question:

> Assume that N functionals of the driving Brownian motion may be evaluated, e.g., evaluations of the driving Brownian motion at N points in the unit interval. Which functionals should be chosen and how should the data be used in order to obtain the best possible approximation to the solution?

We study this question subject to the class of functionals and the error criterion that is used. More specifically, we provide answers in an asymptotic sense, i.e., as $N \to \infty$, by establishing sharp upper and lower error bounds together with the corresponding asymptotic constants. In particular, it turns out that choosing the discretization in an adaptive way is necessary in order to achieve asymptotic optimaliy.

Tractability of Approximation for Weighted Korobov Spaces on Classical and Quantum Computers

Erich Novak[a], Ian H. Sloan[b], and Henryk Woźniakowski[c]

[a]Mathematisches Institut, Universität Jena, Ernst-Abbe-Platz 4, 07740 Jena, Germany, email: novak@mathematik.uni-jena.de

[b]School of Mathematics, University of New South Wales, Sydney 2052, Australia, email: i.sloan@unsw.edu.au

[c]Department of Computer Science, Columbia University, New York, NY 10027, USA, and Institute of Applied Mathematics and Mechanics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland, email: henryk@cs.columbia.edu

We study the approximation problem (or problem of optimal recovery in the L_2-norm) for weighted Korobov spaces with smoothness parameter α. The weights γ_j of the Korobov spaces moderate the behavior of periodic functions with respect to successive variables. The non-negative smoothness parameter α measures the decay of Fourier coefficients. For $\alpha = 0$, the Korobov space is the L_2 space, whereas for positive α, the Korobov space is a space of periodic functions with some smoothness and the approximation problem corresponds to a compact operator. The periodic functions are defined on $[0,1]^d$ and our main interest is when the dimension d varies and may be large. We consider algorithms using two different classes of information. The first class Λ^{all} consists of arbitrary linear functionals. The second class Λ^{std} consists of only function values and this class is more realistic in practical computations.

We want to know when the approximation problem is tractable. Tractability means that there exists an algorithm whose error is at most ε and whose information cost is bounded by a polynomial in the dimension d and in ε^{-1}. Strong tractability means that the bound does not depend on d and is polynomial in ε^{-1}. In this paper we consider the worst case, randomized and quantum settings.

In the worst case setting, we apply known results to prove that strong tractability and tractability in the class $\Lambda^{\rm all}$ are equivalent. This holds iff $\alpha > 0$ and the sum-exponent s_γ of weights is finite, where $s_\gamma = \inf\{ s > 0 : \sum_{j=1}^{\infty} \gamma^s < \infty \}$.

In the worst case setting for the class $\Lambda^{\rm std}$ we must assume that $\alpha > 1$ to guarantee that functionals from $\Lambda^{\rm std}$ are continuous. The notions of strong tractability and tractability are not equivalent. In particular, strong tractability holds iff $\alpha > 1$ and $\sum_{j=1}^{\infty} \gamma < \infty$.

In the randomized setting, it is known that randomization does not help over the worst case setting in the class $\Lambda^{\rm all}$. For the class $\Lambda^{\rm std}$, we prove that strong tractability and tractability are equivalent and this holds under the same assumption as for the class $\Lambda^{\rm all}$ in the worst case setting, that is, iff $\alpha > 0$ and $s_\gamma < \infty$.

In the quantum setting, we consider only upper bounds for the class $\Lambda^{\rm std}$ with $\alpha > 1$. We prove that $s_\gamma < \infty$ implies strong tractability. Hence for $s_\gamma > 1$, the randomized and quantum settings both break worst case intractability of approximation for the class $\Lambda^{\rm std}$.

We also indicate cost bounds on algorithms with error at most ε. The speedup of the quantum setting over the randomized setting is of order $\varepsilon^{-(1+p/2)}$. We stress that p can be arbitrarily large, and in this case the speedup is huge.

Adaptive model selection based on indirect observations blurred by deterministic noise

Sergei V. Pereverzev[a]

[a]Ukrainian Acad. of Sci., Institute of Mathematics, Tereschenkivska Str.3, 01601 Kiev, Ukraine

To estimate an unknown function x from some Hilbert space X one usually uses a sequence of finite-dimensional models (linear subspaces) and some model selection criterion. The unknown function may or may not be in one of the models. In the latter case one thinks of the models as providing an approximation, and approximation error is considered as the loss caused by the model. We study the case when one wishes to estimate unknown x from indirectly observed data near $y = Ax$, where A is some compact linear operator acting in a Hilbert space Y and describing the way the observations are indirect. In practice indirect observations are observed usually in the presence of some noise, so that we observe y_δ such that $\|y - y_\delta\| \le \delta$. The presence of the noise in the data imposes the lower bound on the loss. For example, if $x \in Range(\varphi(A^*A))$, where φ is non-negative operator monotone function with $\varphi(0) = 0$, then it is known that the minimal worst-case loss has the order

$$\Delta_\delta(\varphi) \asymp \delta \sqrt{\psi^{-1}(\frac{1}{\delta^2})},$$

where $\psi(u) = u/\varphi^2(1/u)$. The main accomplishment of the present work is the criterion that adaptively selects the model causing the loss of minimal order and having order optimal description length (dimension of the linear subspace). If $X = X^0, Y = Y^0$ are the spaces from appropriate Hilbert scales $\{X^\lambda\}, \{Y^\lambda\}$ scaled in the same manner as classical

Sobolev scale, and A, A^* act along these scales with step a as isomorphisms between pairs $(X^{\lambda-a}, Y^{\lambda}), (Y^{\lambda-a}, X^{\lambda})$, then the optimal order of description length is

$$n_\delta(\varphi) \asymp [\psi^{-1}(\frac{1}{\delta^2})]^{1/2a}.$$

Average case complexity of weighted approximation and integration on $\mathbb{R}^d$

Leszek Plaskota

Warsaw University, Poland

We study the average case complexity, comp(ε), of weighted L_2-approximation and weighted integration of functions defined on $\mathbb{R}^d$. Approximations (quadratures) are constructed based on finitely many samples. We focus on how the complexity depends on the weight ρ and on the prior Gaussian distribution μ of functions. Our results show the following. (i) The approximation problem is well-defined for any weight. For integration to be well defined, ρ has to decrease sufficiently fast. (ii) For approximation, comp(ε) can be infinite or it can increase arbitrarily fast as $\varepsilon \to 0$. If the integration problem is well-defined then comp(ε) increases slower than ε^{-2}. (iii) If the weight ρ passes a specific level of decay, depending on μ, then, for both problems, comp(ε) is in order the same as for ρ being the characteristic function of the unit ball.

For approximation the proofs are constructive. For integration general upper bounds are obtained using Monte Carlo arguments.

This is a joint work with Klaus Ritter of Technische Universitaet Darmstadt and Grzegorz W. Wasilkowski of The University of Kentucky.

Complexity and Algorithms for Fixed Points with l_∞ norm

S. Shellman, K. Sikorski

School of Computing, University of Utah, Sale Lake City, UT 84112

We present an outline and numerical tests of a BEDFix (bisection-envelope deep-cuts fixed point) algorithm for approximationg fixed points of bivariate functions transforming unit cube into itself. The functions are assumed to be nonexpanding with respect to the infinity norm, i.e., the Lipschitz constant $q = 1$. The worst case number of function evaluations needed to compute an ε-residual solution x: $||f(x) - x|| \leq \varepsilon$ is bounded by $2log(1/\varepsilon) + 1$. The average number of function evaluations is about 30 percent of the upper bound, as observed on about 2^{26} numerical tests. The algorithm also computes an ε-absolute solution z: $||z - \alpha|| \leq \varepsilon$, for the contracting case $q < 1$, where $f(\alpha) = \alpha$ is the actual fixed point. For q close to 1 the BEDFix is much more efficient than the Simple Iteration algorithm. We also show that problem with the absolute error criterion and $q > 1$ has infinite worst case complexity. This means that there exist no algorithms based on function evaluations that solve this problem in the worst case. We briefly outline possible generalizations of BEDFix for arbitrary dimension d.

Constructive Approaches to Quasi-Monte Carlo methods

Ian H. Sloan

School of Mathematics, University of New South Wales, Sydney 2052, Australia, email: i.sloan@unsw.edu.au

Quasi-Monte Carlo (or QMC) methods are methods for approximating the s-dimensional integral

$$I(f) = \int_{[0,1]^s} f(\mathbf{x}) d\mathbf{x}$$

by an equal weight quadrature rule of the form

$$Q(f) = \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_i),$$

with the points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ chosen deterministically, rather than randomly as in the classical Monte Carlo method.

Recently greater interest in QMC methods has been spurred by the apparent success of some methods of this type for financial models involving integrals over hundreds of dimensions.

This talk will review the principles behind QMC methods, with special attention to the recent algorithmic construction of shifted lattice rules, in which the construction proceeds from one dimension to the next. Also of interest are randomised versions of QMC methods, which have some of the characteristics of the classical Monte Carlo method. One such method is a family of randomly shifted lattice rules, which have been shown to achieve the optimal tate of convergence in the setting of the weighted Sobolev spaces introduced in 1998 by Sloan and Woźniakowski.

Sparseness of support vector machines

Ingo Steinwart[a]

Friedrich-Schiller-Universität
07743 Jena, Germany
steinwart@minet.uni-jena.de

Classification is one of the major problems considered in learning theory. Besides the question wether a given algorithm can learn from samples it is also important to know how many samples are needed to guarantee a certain accuracy and how many samples are actually used to evaluate the decisions of the classifier. The latter is also known as the sparseness of the algorithm.

In this talk we first give a brief introduction to the standard model for classification tasks. We then discuss how many samples are needed to guarantee a certain accuracy. In particular we consider some negative results. The major part of this talk is devoted to support vector machines which are of rapidly increasing interest. After a brief description of these algorithms we first present recent results on their learning ability. We then discuss the sparseness of support vector machines. In particular we show that in typical situations they are less sparse than they are usually assumed to be.

On Derandomization of Randomized $(0, s)$-Sequences

Shu Tezuka[a]

[a]IBM Tokyo Research Laboratory,
1623-14, Shimotsuruma, Yamato,
Kanagawa, Japan 242-8502
e-mail: tezuka@jp.ibm.com

In this talk, we discuss two classes of digital $(0, s)$-sequences, generalized Faure sequences and new Faure sequences, in the light of i-binomial property. The i-binomial property is known to play a crucial role when we analyze lower bounds for the two-dimensional discrepancy of Faure sequences. We give a necessary and sufficient condition for generator matrices of each of the two classes to have i-binomial property.

On Tractability of Weighted Integration over Bounded and Unbounded Regions in $\mathbb{R}^s$

Fred J. Hickernell[a], Ian H. Sloan[b], and Grzegorz W. Wasilkowski[c]

[a]Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[b] School of Mathematics, University of New South Wales, Sydney 2052, Australia

[c]Department of Computer Science, University of Kentucky, 773 Anderson Hall, Lexington, KY 40506-0046, USA

We prove that for the space of functions with mixed first derivatives bounded in L_1 norm, the weighted integration problem over bounded or unbounded regions is equivalent to the corresponding classical integration problem over the unit cube. This correspondence yields tractability of the general weighted integration problem.

Quasi-Monte Carlo Sampling for Computing the Trace of a Function of a Matrix

Jessie Wong Mei Ning

Department of Mathematics, Hong Kong Baptist University
Kowloon Tong, Hong Kong SAR, China

In the past ten years, there has been some research on finding the trace of the inverse and the determinant of a matrix. For example Bai, Fahey and Golub solve the above problems by Monte Carlo method [BFG96]. This presentation explores the effectiveness of using quasi-Monte Carlo (QMC) samplings on an n-dimensional finite grid domain to solve these problems. The comparisons to Monte Carlo (MC) sampling on this domain will be made. The various types of error analyzes including worst case, random case and average case will also be discussed. Besides, we include the control variate method in order to improve the accuracy and convergence rate of the MC and QMC samplings. Some numerical results for estimating the trace of the inverse and the determinant of a matrix are presented.

References:

[BFG96] Z. Bai, M. Fahey, G. Golub, *Some Large Scale Matrix Computation Problems*, J. Comput. Appl. Math., 74, No 1-2, pp.71–89, 1996.

Surface Approximation is Sometimes Easier Than Surface Integration

Arthur G. Werschulz[a] and Henryk Woźniakowski[b]

[a]Department of Computer and Information Sciences, Fordham University, and Department of Computer Science, Columbia University

[b]Department of Computer Science, Columbia University, and Institute of Applied Mathematics, University of Warsaw

The approximation and integration problems consist of finding an approximation to a function f or its integral over some fixed domain Σ. For the classical version of these problems, we have partial information about the functions f and complete information about the domain Σ; for example, Σ might be a cube or ball in R^d. When this holds, it is generally the case that integration is not harder than approximation; moreover, integration can be much easier than approximation. What happens if we have partial information about Σ? In this talk we study the surface approximation and surface integration problems, in which $\Sigma = \Sigma_g$ depends on a function g. We assume that the functions f are r times continuously differentiable scalar functions of l variables, and the functions g are s times continuously differentiable injective functions of d variables with l components. The class of surfaces is generated as images of cubes or balls, or as oriented cellulated regions. Error for the surface approximation problem is measured in the L_q sense. These problems are well defined provided that $l \geq l$, $r \geq 0$, and $s \geq 1$. Information consists of function evaluations of f and g.

We show that the worst case ε-complexity of surface approximation is proportional to $\varepsilon^{-1/\mu}$ with $\mu = \min\{r, s\}/d$. For $s \geq 2$, we show that the worst case ε-complexity of surface integration is proportional to $\varepsilon^{-1/\nu}$ with

$$\nu = \min\left\{\frac{r}{d}, \frac{s - \delta_{s,1}(1 - \delta_{d,l})}{\min\{d, l-1\}}\right\},$$

where $\delta_{\cdot,\cdot}$ is the Kronecker delta function. (This bound also holds for several sub-cases with $s = 1$; we conjecture that it holds for all $l \geq l$, $r \geq 0$, and $s \geq 1$.)

Using these results, we determine when surface approximation is easier than, as easy as, or harder than surface integration; all three possibilities can occur. In particular, if $l > d$ and $r = s = 1$, then $\mu = 1/d$ and $\nu = 0$, so that surface integration is unsolvable and surface approximation is solvable; this is an extreme case for which surface approximation is easier than surface integration.