Fred J. Hickernell, Peter Kritzer (Eds.)
Multivariate Algorithms and Information-Based Complexity

Radon Series on Computational and Applied Mathematics

Volume 27

Multivariate Algorithms and Information-Based Complexity

Edited by
Fred J. Hickernell and Peter Kritzer

DE GRUYTER

Editors
Prof. Dr. Fred J. Hickernell
Illinois Institute of Technology
Department of Applied Mathematics
Center for Interdisciplinary Scientific
3105 Dearborn St.
Chicago 60616
USA
peter.kritzer@oeaw.ac.at

PD Dr. Peter Kritzer
Austrian Academy of Sciences
Altenbergerstraße 69
4040 Linz
Austria
peter.kritzer@oeaw.ac.at

ISBN 978-3-11-063311-5
e-ISBN (PDF) 978-3-11-063546-1
e-ISBN (EPUB) 978-3-11-063315-3
ISSN 1865-3707

Library of Congress Control Number: 2020934321

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Typesetting: VTeX UAB, Lithuania
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Preface: Multivariate algorithms and information-based complexity

The authors of this book include several of the invited speakers in the workshop *Multivariate Algorithms and Information-Based Complexity*, which was part of the RICAM Special Semester on *Multivariate Algorithms and their Foundations in Number Theory* in the fall of 2018. The special semester consisted of four larger and two smaller workshops on various topics ranging from Pseudo-Randomness and Discrepancy Theory to Information-Based Complexity and Uncertainty Quantification. This book arises from the second workshop, which took place at the Johann Radon Institute for Computational and Applied Mathematics (RICAM) of the Austrian Academy of Sciences in Linz, Austria, on November 5–9, 2018.

Multivariate continuous problems occur in a multitude of practical applications, such as physics, finance, computer graphics, and chemistry. The number of variables involved, d, can be in the hundreds or thousands. The information complexity of a given problem is the minimal number of information operations required by the best algorithm to solve the problem for a prescribed set of inputs within a certain error threshold, ε. Typical examples of information operations are function values and linear functionals. The field of information-based complexity (IBC), founded by Traub and Wozniakowski in the 1980s, analyzes the information complexity for multivariate problems and determines how it depends on d and ε. A crucial question is under which circumstances one can avoid a curse of dimensionality, namely, exponential growth of the information complexity with d. This book addresses the analysis of multivariate (continuous) problems, especially from the IBC viewpoint.

The problems discussed by the authors reflect the breadth of current inquiry under the umbrella of multivariate algorithms and IBC. The chapter entitled"The control variate integration algorithm for multivariate functions defined at scattered data points" studies a method of approximating the integral of a multivariate function, in which one uses the exact integral of a control variate based on a least-squares multivariate quasiinterpolant. Numerical examples demonstrate that such an algorithm can overcome the curse of dimensionality for multivariate least-squares fits. The second chapter, titled "An adaptive random bit multilevel algorithm for SDEs", considers the approximations of expectations for functionals applied to the solutions of stochastic differential equations by employing Monte Carlo methods based on random bits instead of random numbers. An adaptive random bit multilevel algorithm is provided and compared numerically to other methods. The chapter "RBF-based penalized least-squares approximation of noisy scattered data on the sphere" deals with the approximation of noisy scattered data on the 2-dimensional unit sphere. In particular, global and local penalized least-squares approximation based on radial basis functions (RBFs) are explored. The authors of the fourth chapter in this book, "On the power of random information", consider a problem from the core of IBC theory,

https://doi.org/10.1515/9783110635461-201

namely the quality of random information in approximation and integration problems. Random information is compared to optimal information for such problems. It is shown that in some cases random information can be almost optimal, whereas in other cases it may yield much worse results than optimal information. The chapter "Optimality criteria for probabilistic numerical methods" lies in the field of Bayesian decision theory. To be more precise, the authors study an optimality criterion from Bayesian experimental design and consider its implied optimal information in the numerical context. Furthermore, this chapter compares this information to information commonly used in average-case-optimal numerical methods. The sixth chapter of the book, "ε-Superposition and truncation dimensions, and multivariate decomposition method for ∞-variate linear problems" deals with linear problems on weighted normed function spaces, the elements of which depend on infinitely many variables. The author focuses on the ε-truncation and ε-superposition dimension, as well as multivariate decomposition methods, which are means to reduce the number of variables in the problem to a relatively small finite number. The anchored and other types of decompositions are used. The chapter "Adaptive approximation for multivariate linear problems with inputs lying in a cone" completes the book and considers adaptive approximation algorithms for functions lying in particular subsets of certain function spaces. In contrast to much of the IBC literature, it is not assumed that the functions to be approximated lie in a ball of a certain radius, but instead in a cone. It is shown that for such problems it is an advantage to use adaptive rather than nonadaptive algorithms for function approximation.

All chapters were reviewed by renowoned experts. We wish to thank these anonymous referees for their precious help.

We would like to express our gratitude towards all speakers of the workshop for giving excellent talks on their respective fields of expertise and for contributing to the success of the workshop:

- Paul Constantine (University of Colorado),
- Martin Ehler (University of Vienna),
- Mario Hefter (TU Kaiserslautern),
- Stefan Heinrich (TU Kaiserslautern),
- Kerstin Hesse (Paderborn University),
- Fred J. Hickernell (Illinois Institute of Technology),
- James (Mac) Hyman (Tulane University),
- Thomas Müller-Gronbach (University of Passau),
- Erich Novak (University of Jena),
- Chris Oates (Newcastle University and Alan Turing Institute),
- Houman Owhadi (California Institute of Technology),
- Sergei Pereverzyev (RICAM), Leszek Plaskota (University of Warsaw),
- Paweł Przybyłowicz (AGH University of Science and Technology Kraków),
- Klaus Ritter (TU Kaiserslautern),
- Ingo Steinwart (University of Stuttgart),

- Michaela Szölgyenyi (University of Klagenfurt),
- Greg W. Wasilkowski (University of Kentucky),
- Henryk Woźniakowski (Columbia University and University of Warsaw),
- Henry Wynn (London School of Economics),
- Larisa Yaroslavtseva (University of Passau).

We also would like to thank all those who participated in the workshop. Annette Weihs and Wolfgang Forsthuber provided valuable administrative and technical support. Further information on the RICAM Special Semester "Multivariate Algorithms and their Foundations in Number Theory" can be found at

https://www.ricam.oeaw.ac.at/specsem/specsem2018/

The financial support of the Johann Radon Institute for Computational and Applied Mathematics (RICAM) of the Austrian Academy of Sciences is gratefully acknowledged.

Fred J. Hickernell and Peter Kritzer,
Chicago and Linz, March 2020.

Contents

Jeremy Dewar, Mu Tian, and James M. Hyman

1 The control variate integration algorithm for multivariate functions defined at scattered data points

Abstract: We describe numerical studies to demonstrate the accuracy and efficiency of approximating the integral of a function over a prescribed multidimensional region using the exact integral of a control variate based on a least-squares multivariate polynomial quasiinterpolant. This control variate integration approach is shown to be effective in approximating the integral of smooth functions, where only a few function values are available. We compare these methods with traditional Monte Carlo quadrature on independent and identically distributed random and low-discrepancy (quasi-Monte Carlo) points. We verify that the control variate integration methods converge at the same rate as Monte Carlo methods, but have a much smaller error constant. We provide numerical examples to demonstrate that sparse quasiinterpolation methods are a practical approach to reduce the curse of dimensionality for high-dimensional multivariate least-squares fits.

Keywords: Quadrature, cubature, multivariate, integration, low discrepancy sampling, quasi-Monte Carlo

MSC 2010: 62H12, 65D30, 65D05

1.1 Introduction

Estimating the integrals of high-dimensional multivariate functions arises in estimating the marginal likelihood in Bayesian analysis of problems ranging from quantifying the uncertainty in pricing financial options [3, 15, 22] to the solution of stochastic partial differential equations with random coefficients, such as the flow of fluids through porous media [4, 18]. There are accurate and robust numerical methods to integrate a function over a low-dimensional domain. Most of these methods are either built on a lattice grid or a particular set of deterministic points, such as Gaussian or Smolyak

Acknowledgement: The authors thank Lin Li for his feedback on these simulations and many helpful discussions about the CVI and future extensions of the method. We are grateful to Chris Oates for his many helpful suggestions to improve the paper. This research was partially supported by the funding from Tulane University Evelyn and John G. Phillips Distinguished Chair in the Mathematical Sciences.

Jeremy Dewar, James M. Hyman, Department of Mathematics, Tulane University, New Orleans, LA 70118, USA, e-mails: jdewar@tulane.edu, mhyman@tulane.edu
Mu Tian, Facebook, 1101 Dexter Ave N, Seattle, WA 98109, USA, e-mail: mtian@fb.com

https://doi.org/10.1515/9783110635461-001

quadrature points, and are based on the exact integration of an interpolating function through the sample points.

As the dimension of the problem increases, traditional numerical integration algorithms [6] suffer from the curse of dimensionality, and require the number of samples to increase exponentially. Integration formulas built on sparse grids [2, 21] partially overcame the curse of dimensionality. However, even these approaches suffer from the curse of dimensionality above eight to ten dimensions. Monte Carlo (MC) integration is based on independent identically distributed (IID) pseudo-random sampling and the mean of the samples converges, with probability 1, to the desired integral by the strong law of large numbers. The MC has a probabilistic convergence rate of $O_p(1/\sqrt{N})$, independent of the dimension, as the number of samples, N, increases.

The quasi-Monte Carlo (QMC) method is based on nonrandom low-discrepancy sampling (LDS) and converges faster, $O((\log N)^{(d-1)}/N)$, where d is the dimension [3, 9]. Recent improvements in the generating functions for low-discrepancy sampling has improved the asymptotic convergence rates of QMC methods [2, 8, 19]. Unfortunately, the errors of the best QMC methods can still be unacceptably large when there are few samples.

For higher-dimensional situations when there are few samples, $f(\mathbf{x}_i)$, available, we fit a least-squares quasiinterpolant, $\hat{f}(\mathbf{x})$, through the data that minimizes the variance of the residuals, $r_i = \hat{f}(\mathbf{x}_i) - f(\mathbf{x}_i)$. This least squares fit is a *quasiinterpolant*, since it is not exact at the data points and is a control variate estimating the underlying function. We then approximate the integral of $f(x)$ by the exact integral of $\hat{f}(\mathbf{x})$ and an estimate for the integral of the residual. We demonstrate that this *control-variate integraton*, (CVI), algorithm is an effective approach [12] for smooth functions, where there are few samples, and the effective dimension is relatively small. This is often the case when additional function evaluations (samples), $f(\mathbf{x})$, are not available or expensive, such as with large-scale simulations.

The MC and QMC methods approximate the integral of the underlying function by the mean of the samples. This approximation is equivalent to fitting the data with a least squares constant and integrating this constant. That is, traditional Monte Carlo integration can be viewed as a CVI using a least-squares constant function. We observed that the higher-degree quasiinterpolants could reduce this error by a factor of 10, 100, or more.

Let $\mathbf{x} \in R^d$ and the function $f(\mathbf{x}) : R^d \to R^1$ be integrable and bounded. The integral of $f(\mathbf{x})$ on the d-dimensional unit hypercube $\Omega = [0,1]^d$ is

$$\int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} = \int_0^1 \cdots \int_0^1 f(x_1, x_2, \ldots, x_d) dx_1 \cdots dx_d \,. \tag{1.1}$$

All integrations in this paper will be over the domain $[0,1]^d$, and the domain will be dropped from the integral notation from here on. When the integral is over a different region, we will assume that the problem can be mapped to the unit cube through

a change of variables. Note that creating a good mapping can be a nontrivial problem [16]. Even in bounded regions with large aspect ratios, the mapping can convert a function that is equally smooth in all directions to one that can vary much more in some directions than others.

A traditional approach for approximating the integral is by weighted sum [7],

$$I_f \overset{\text{def}}{=} \int f(\mathbf{x})d\mathbf{x} \approx \sum_{i=1}^{N} w_i f(\mathbf{x}^i) \overset{\text{def}}{=} \hat{I}_f \,, \tag{1.2}$$

where w_i is the integration weight associated with the sample point $f(\mathbf{x}^i)$, and $\mathbf{x}^i \in [0,1]^d$ for $i = 1,\ldots,N$ samples. If the samples are approximately uniformly distributed in $[0,1]^d$, then the Monte Carlo methods use the sample mean, $w_i = 1/N$, to approximate the integral. Most classical integration methods define the weights so that the quadrature is exact for a class of functions interpolating the data.

The asymptotic integration error estimates for the MC method is

$$|\hat{I}_f - I_f| = O_p\left(\frac{1}{\sqrt{N}}\sigma(f)\right), \tag{1.3}$$

where $\sigma(f) = [\int (f(\mathbf{x}) - I_f)^2 dx]^{1/2}$. The corresponding asymptotic integration error estimate for QMC method is

$$|\hat{I}_f - I_f| = O\left(\frac{(\log N)^{(d-1)}}{N} V[f]\right). \tag{1.4}$$

Here, $V[f]$ is the variation in f, $V[f] = \int |\frac{\partial^d f}{\partial_{x_1}\cdots\partial_{x_d}}| dx_1 \cdots dx_d + \sum_{i=1}^{N} V[f(\mathbf{x}^i)]$ [20]. Note that the QMC error is not a tight bound, and in practice, the error often converges faster. Also, although the QMC method converges faster, the error can be biased, unlike the MC method [3].

In the CVI, we first fit the sample with a least-squares quasiinterpolant, $\hat{f}(\mathbf{x})$. We then use this function as a control variate and estimate the integral by the exact integral of $\hat{f}$, plus an approximation of the residual $r(x) = f(x) - \hat{f}(x)$,

$$\int f(\mathbf{x})dx = \int \hat{f}(\mathbf{x})dx + \int f(\mathbf{x}) - \hat{f}(\mathbf{x})dx = \int_{\text{exact}} \hat{f}(\mathbf{x})dx + \int r(\mathbf{x})dx \,. \tag{1.5}$$

The error is now determined by the integral of the residual and asymptotically decreases at the same rate as the original Monte Carlo approximation. This is to be expected since the error estimates are the product of the sample discrepancy based on the distribution and number of samples, which are unchanged, and the variation of the function being integrated, which is independent of the number of samples.

Since the integral of the control variate, $\int \hat{f}$, is known exactly, the only error is the integral of the residual $\int r(x)dx$. That is, the CVI integration error estimates are

the same as the MC and QMC error estimates when applied to integrating the residual $r(\mathbf{x})$, as they were for integrating $f(\mathbf{x})$. The residual is in the null space of the basis functions. Therefore, the expected value is zero, $E(r) = 0$, if the quasiinterpolant is exact for constant functions. If we assume that the integral of the residual is zero, then the Monte Carlo methods can be viewed as a CVI, where the quasiinterpolant is the constant function.

The CVI is also called a regression estimator, and is typically used with low- degree approximating functions [5, 12, 14, 17]. Our numerical study focuses on using higher-degree polynomial control variates and analyzing the magnitude of the error (i. e., the constant in front of that convergence rate) for both a fixed number of samples N, and as the number of samples increases. We also analyze the complexity of the algorithm as the dimension d increases.

1.2 Quasiinterpolation algorithm

We represent our control variates as a sum of orthogonal basis functions, $f(\mathbf{x}) = \sum_k \sum_j \beta_{j,k}\psi_{j,k}(\mathbf{x})$, with total degree K. The multivariate basis functions, $\psi_{j,k}(\mathbf{x}) : R^d \to R^1$ of degree $k \leq K$, are decomposed into the product univariate basis functions, $\{\phi_{k_i^j}(x_i)\}_{j=1}^{t_k^d} : R^1 \to R^1$,

$$\psi_{j,k}(\mathbf{x}) = \prod_{i=1}^{d} \phi_{k_i^j}(x_i)\,. \tag{1.6}$$

Here, $\mathbf{x} = [x_1, x_2, \ldots, x_d]^T \in R^d$ and k_i^j is the highest degree of x_i in the univariate basis function $\phi_{k_i^j}(x_i)$, where j ranges from 1 to $t_K^d = \binom{K+d-1}{d-1}$ and is constrained by $\sum_{i=1}^d k_i^j \leq K$. In this paper, we use the Legendre polynomial basis functions shifted to the interval $[0,1]$: $\phi_0(x) = 1$, $\phi_1(x) = 2x - 1$, $\phi_2(x) = 6x^2 - 6x + 1$, ….

A polynomial of total degree K has $M = \sum_{l=1}^K t_l^d = \binom{K+d}{d}$ basis functions. We order and relabel β_l and $\psi_l(\mathbf{x})$ lexicographically as a function of j and k, and express our control variate quasiinterpolant as

$$\hat{f}(\mathbf{x}) = \sum_{l=1}^{M} \hat{\beta}_l \psi_l(\mathbf{x}). \tag{1.7}$$

High-degree polynomials in high-dimensions are quickly constrained by the curse of dimensionality (Table 1.1). For the solution to be unique and the coefficients identifiable, then, there must be more independent samples than coefficients in the quasiinterpolating function. For even moderate degree polynomials in moderate dimensional problems, the number of terms rapidly increases.

Table 1.1: Table of the number of terms in multivariate polynomial with total degree K. The mixed variable terms in multivariate polynomials create an explosion in the number of terms as a function of the degree and dimension. For example, a 5^{th} degree polynomial in 20 dimensions has 53,130 terms. We use sparse quasiinterpolation algorithms to reduce the number of terms in the quasiinterpolant.

K / d =	1	2	3	4	5	10	20
0	1	1	1	1	1	1	1
1	2	3	4	5	6	11	21
2	3	6	10	15	21	66	231
3	4	10	20	35	56	286	1,771
4	5	15	35	70	126	1,001	10,626
5	6	21	56	126	252	3,003	53,130
6	7	28	84	210	462	8,008	230,230
10	11	66	286	1,001	3,003	184,756	30,045,015

The coefficients $\hat{\beta}_l$ are defined by solving the linear ordinary least-squares problem

$$\hat{\boldsymbol{\beta}} = \underset{\beta_1,\beta_2,\ldots\beta_M}{\operatorname{argmin}} \sum_i \left(\hat{f}(\mathbf{x}^i) - f(\mathbf{x}^i)\right)^2 = \underset{\beta_1,\beta_2,\ldots\beta_M}{\operatorname{argmin}} \|A\boldsymbol{\beta} - \mathbf{f}\|_2^2 \ . \tag{1.8}$$

Here, $\mathbf{x}^i$ is the i^{th} sample value, and the $(i,j)^{\text{th}}$ element of the design matrix A is defined by $a_{i,j} = \psi_j(\mathbf{x}^i)$.

One of the challenges in least squares approximations is the variance-bias trade-off. If the expansion has too many high-degree terms, the variance increases and the function will be over-fit. If there are too few terms, then the function will be underfit, and the residual errors will be biased and larger than necessary. We use an l_1 penalty function (LASSO) to regularize the least-squares fit and pick out the most needed terms using sparse subset selection. That is, instead of solving the standard least-squares problem (1.8), we add an l_1 penalty term and solve

$$\hat{\boldsymbol{\beta}} = \underset{\beta_1,\beta_2,\ldots\beta_M}{\operatorname{argmin}} \left\{ \frac{1}{2}\|A\boldsymbol{\beta} - \mathbf{f}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \right\} \ . \tag{1.9}$$

We solve the system using cyclic coordinate descent algorithm [10, 13], and the optimal λ is determined by a cross-validation test. If the function $f(\mathbf{x})$ varies in some directions more than others, then this sparse subset selection approach extracts the appropriate basis functions based on the effective dimension of the active subspaces.

We used all the samples when fitting the data and chose λ by the simple method of increasing λ until just below, where there was a significant increase in the norm of the residuals, $\|\hat{y} - y\|$. We will be comparing this approach with more sophisticated cross-validation data-splitting methods in the future.

An advantage of using sparse subset selection is that the quadrature only has to be evaluated on the remaining sparse set basis functions with nonzero coefficients. We also considered using a simpler l_2 regularization for the least squares fit. This results in solving an easier (linear) problem for the control variate coefficients, but requires evaluating the integral of all of the basis functions. It is not clear which of these approaches will be the most effective, and we are currently comparing them on both smooth and nonsmooth test problems.

1.3 Numerical examples

We compare the methods on five- and six-dimensional smooth functions by generating LDS and IID samples form uniform distributions of size $500 \leq N \leq 4000$, and quasiinterpolating polynomials with total degrees $K = 0, 1, \ldots 5$. We then compare the methods a nonsmooth wedge function in six dimensions. We analyze the distributions of errors from an ensemble of 300 runs, each with distinct samples. The IID samples are generated using the MATLAB pseudorandom generator, *rand*, and the LDS are generated using the Niederreiter or Sobol sequences.

For the test problems, we calculate the exact error, the convergence rate, the constant factor in the error estimate, and an approximate integration error using bootstrapping. We compute the error distributions over a large ensemble of simulations to avoid misleading results, where any particular sample set might have unusually large or small errors.

For each case in the ensemble of 300 runs, the testing procedure is to

1. generate the N sample locations $\{\mathbf{x}^i\}_1^N$,
2. evaluate $f(\mathbf{x}^i)$,
3. fit a (possibly regularized) least-squares polynomial $\hat{f}(\mathbf{x})$ of degree K through the sample,
4. define the approximate integral as $\int \hat{f}(\mathbf{x}) d\mathbf{x}$,
5. estimate the error using bootstrapping (resampling with replacement) [13], and finally
6. evaluate the true integration error for the CVI, $e = |\int (\hat{f}(\mathbf{x}) - f(\mathbf{x})) dx|$.

After each run, the true integration error is compared with the bootstrapping approximation of the standard error.

We compared the convergence rate of the CVI method on smooth and discontinuous functions defined on IID and LDSs, for N ranging from a few hundred to a few thousand, and polynomial quasiinterpolants of degree less than 5.