Pergamon

Nonlinear Analysis, Theory, Methods & Applications, Vol. 30, No. 3, pp. 1439–1446, 1997
Proc. 2nd World Congress of Nonlinear Analysts

0362-546X/97 $17.00 + 0.00

PII: S0362-546X(96)00134-4

HIGH DIMENSIONAL NUMERICAL PROBLEMS

ERICH NOVAK

Mathematisches Institut, Universität Erlangen-Nürnberg
Bismarckstr. 1 1/2, 91054 Erlangen, Germany, novak@mi.uni-erlangen.de

Key words and phrases: high dimensional problems, curse of dimension, integration, global optimization, Smolyak's construction, Boolean methods, sparse grids, optimal error bounds, partially separable problems, complexity of numerical problems, information-based complexity

1. INTRODUCTION

High dimensional numerical problems are ubiquitous although one tries to avoid them because of their intrinsic difficulty. They are present, for example, in statistical mechanics, see [23] for an introduction. Another very significant problem is that of computing the native 3-dimensional conformation (folded state) of a globular protein given its amino acid sequence by the global minimization of high dimensional nonconvex energy functions, see [16]. Yet another important example is the fast valuation of financial derivatives, see [17]. Many applications even require approximate values of path integrals, i.e., integrals over infinite dimensional spaces, see [25].

There are limits of practical computability. These limits are often related to the high dimension of the problem. There is a "curse of dimension" which means that the minimal cost of computing an approximation is exponential in the dimension of the problem. Hence, many high dimensional numerical problems are practically impossible to solve.

Important is not only the number of variables, i.e., the dimension, but also the smoothness and the structure of the class of functions. To a certain degree, a high dimension can be compensated by enough smoothness and/or structure. As a rule, such a compensation is not possible by means of classical methods. Hence we have to invent new methods. By smoothness we mean that the underlying functions have certain derivatives whose norms are not too large. By structure we mean certain geometrical properties, like convexity. Later we will discuss partially separable problems, another structural property. We will see that this property is strongly related to certain nonclassical smoothness classes of Nikolskii, Korobov, and Babenko.

We want to expand the limits of practical computability by using, in an optimal way, the a priori knowledge about the smoothness and the structure of the problem.

It is a major goal of scientific computing to expand traditional limits of practical computability. To reach this goal, one should use the art and the techniques of computer science, such as parallel processing. In addition, however, it is very important to optimize the underlying algorithms with respect to their cost and accuracy. Optimal numerical methods can only be realized if the given information concerning the analytical model is fully used. This analysis-based view can exhibit the fundamental limits of practical computability. It is remarkable that these limits were already discussed by Kolmogorov, Bakhvalov, Nikolskii and Babenko, almost 40 years ago. Now these limits are systematically studied in the field of "information-based complexity", see [7], [11-13], [21], and [24-29].

2. THE CURSE OF DIMENSION

The curse of dimension occurs for many numerical problems. Assume that we want to compute a value $S(f)$ of a (linear or nonlinear) operator

$$S : X \to G$$

between two normed spaces. Often X is a space of functions f and we consider discretizations of S of the form

$$S_n(f) = \varphi(f(x_1), \ldots, f(x_n)). \tag{1}$$

As the error criterion we often take the worst case error

$$e(S_n) = \sup_{f \in F} \|S(f) - S_n(f)\|$$

over the unit ball F of X. The cost of computing $S_n(f)$ is at least n and often proportional to n. For simplicity we use here the number n of function evaluations as a measure for the cost.

We define the optimal error bounds

$$e_n(S) = \inf_{S_n} e(S_n)$$

and consider different operators S on the standard smoothness classes

$$C_d^k = \{f : [0,1]^d \to \mathbf{R} \mid \|f^{(\alpha)}\|_\infty \le 1,\ |\alpha| \le k\},$$

where $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbf{N}_0^d$ denotes a partial derivative with the order $|\alpha| = \alpha_1 + \ldots + \alpha_d$. The optimal error bounds

$$e_n(S) \asymp n^{-k/d} \tag{2}$$

are well known for many problems, see [1], [7], [12], [21], and [26]. The bound (2) holds, for example, for the problem of numerical integration with

$$S(f) = \int_{[0,1]^d} f(x)\, dx$$

or for the problem of global optimization,

$$S(f) = \inf_{x \in [0,1]^d} f(x).$$

It also holds for approximation (recovery) of the function and for elliptic boundary value problems with smooth data.

Suppose that we want to compute an approximation S_n with error at most ε for a given positive and usually small ε. Then (2) yields

$$\text{cost} = n \asymp \left(\frac{1}{\varepsilon}\right)^{d/k}.$$

The curse of dimension means that the minimal cost increases exponentially with the dimension d. Hence, the curse of dimension is present for many problems defined for the class C_d^k.

The proof of (2) consists of two parts. First we must prove a lower bound that is true for arbitrary methods of the form (1). Second we must prove that there is a sequence $(S_n)_n$ of methods such that

$$e(S_n) \leq c_{d,k} \cdot n^{-k/d} \tag{3}$$

for positive numbers $c_{d,k}$. The upper bound (3) can be achieved for methods that use the function values from a regular grid. For the problem of numerical integration, for example, product formulas with the Gaussian method lead to the optimal order $n^{-k/d}$ for every k and every d.

Using the function values from a grid is a very classical idea that leads to good methods if d is small, say $d < 6$. Product rules cannot be applied, however, if d is large. Consider, for example, the case $d = 40$ and $k = 80$. Then the optimal order of convergence is n^{-2}, see (2). It is not clear, however, how to get practical methods since the smallest product method already needs 2^{40} points which is too large. Therefore the optimal order of product methods is only a theoretical result and we need new methods that realize (almost) the optimal order of convergence and can be used with a relatively small number of function values.

Known methods cannot fully use the existing smoothness if d is large, at least not for reasonable small values of n. The situation is better in the periodic case, see [18] and [20]. Using periodization techniques we also have methods for the nonperiodic case but it seems to be difficult to use optimally the existing smoothness in this way. Often Monte Carlo methods and/or number theoretic methods are used in high dimensions. Such methods can be applied for relatively small n but again they can not fully use the smoothness properties of f. We will see in Section 4 that a construction of Smolyak from the year 1963 is very powerful in this situation.

The error estimate (2) is a "negative" result if $d \gg k$. In this case every method converges slowly and we should look for "smaller" classes F that allow smaller errors and still contain many interesting functions.

It should be stressed that not all high dimensional problems are intractable. In the case of ode's $y'(x) = f(y(x))$ with $y : [0,1] \to \mathbf{R}^d$ the optimal order is

$$e_n(S) \asymp n^{-k},$$

independently of d, see [8] and [9]. Also the problem of convex optimization behaves nicely, see [11] and [21].

3. PARTIALLY SEPARABLE PROBLEMS AND NONCLASSICAL FUNCTION SPACES

Many high dimensional problems have a particular structure; they are partially separable. This means that f may depend on a huge number of variables $x_1, \ldots, x_d$ but can be written as a sum of functions that only depend on few variables. A simple case would be

$$f(x) = \sum_{1 \leq i,j \leq d} f_{i,j}(x_i, x_j), \tag{4}$$

where f is a sum of functions that only depend on two variables. At first glance it is not clear how we can use property (4) if we want to compute the integral of f. Assume that $f_{i,j} \in C^2$ for each $f_{i,j}$ in the sum (4). Then there exist, in addition to $f \in C^2$, many higher mixed derivatives, in particular

$$\frac{\partial^d f}{\partial x_1 \ldots \partial x_d}.$$

The function f is in a class of functions with bounded mixed derivatives. Such classes were already studied by Nikolskii, Korobov, and Babenko in the sixties.

We consider the scale

$$F_d^k = \{f : [0,1]^d \to \mathbf{R} \mid \|f^{(\alpha)}\|_\infty \le 1 \text{ where } \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_d],\ \alpha_i \le k\}.$$

Observe that $f \in F_d^k$ has a partial derivative of the order $k \cdot d$ but the derivative

$$\frac{\partial^{k+1} f}{\partial x_1^{k+1}}$$

may not exist. Hence the class F_d^k is not isotropic.

Because of their tight connection to partially separable functions we want to study the classes F_d^k, as well as the more traditional classes C_d^k. A given "smooth" function f usually is an element of several different classes F_d^k and/or C_d^k, the exact degree of smoothness often is not known to us. Therefore it is important to have a "universal" method that is optimal (or almost optimal) for many different classes. Hence the following problem seems to be of great practical importance.

Problem: Construct optimal (or almost optimal) algorithms for C_d^k and F_d^k. In particular, find such algorithms

- simultaneously for all classes C_d^k and F_d^k;
- for relatively small n.

This problem can be solved with the help of Smolyak's construction. The resulting method uses the full smoothness of the underlying function, but it is not necessary to know the degree of smoothness in advance.

4. THE CONSTRUCTION OF SMOLYAK

Special cases of this method were invented independently for different applications and are known under different names, such as hyperbolic cross points methods, Boolean methods, discrete blending methods, and sparse grid methods, see [3], [6], [14], [15], [19], [24], and [30].

We assume that for $d = 1$ a sequence of formulas

$$U^i(f) = \sum_{j=1}^{m_i} f(x_j^i) \cdot a_j^i$$

is given. In the case of numerical integration the a_j^i are just numbers, in the case of interpolation (or reconstruction) the a_j^i are functions. The method U^i uses m_i function values and we assume that U^{i+1} has smaller error than U^i and $m_{i+1} > m_i$. We may assume, for example, that

$$m_{i+1} \approx 2\, m_i. \tag{5}$$

Define then, for $d > 1$, the tensor product formulas

$$(U^{i_1} \otimes \cdots \otimes U^{i_d})(f) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} f(x_{j_1}^{i_1}, \ldots, x_{j_d}^{i_d}) \cdot (a_{j_1}^{i_1} \otimes \cdots \otimes a_{j_d}^{i_d}).$$

If the a_j^i are numbers then $\otimes$ just denotes the ordinary product of real numbers. In the case of functions we have

$$(a_1 \otimes a_2)(x_1, x_2) = a_1(x_1) \cdot a_2(x_2).$$

A tensor product formula clearly needs

$$m_{i_1} \cdot m_{i_2} \cdot \ldots \cdot m_{i_d}$$

function values. It is known that all tensor product formulas $S_n = U^{i_1} \otimes \cdots \otimes U^{i_d}$ lead to relatively large errors for the classes F_d^k, the optimal order of such methods is

$$e(S_n) \asymp n^{-k/d}.$$

The Smolyak formulas $A(q,d)$ are clever linear combinations of tensor product formulas such that:

- only tensor products with a relatively small number of knots are used;
- the linear combination is chosen in such a way that an interpolation property for $d = 1$ is preserved for $d > 1$.

The Smolyak formulas are defined by

$$A(q,d) = \sum_{q-d+1 \le |\vec{i}| \le q} (-1)^{q-|\vec{i}|} \cdot \binom{d-1}{q-|\vec{i}|} \cdot (U^{i_1} \otimes \cdots \otimes U^{i_d}),$$

where $q \ge d$.

We now consider the error $e(A(q,d))$ of $A(q,d)$ on F_d^k. Assume (5) and let for $d = 1$,

$$e(U^i) \le c_{k,1} \cdot m_i^{-k}. \tag{6}$$

It is well known that the exponent k in (6) is optimal.

Theorem ([19] and [24]): For $d > 1$ and F_d^k there exist positive numbers $c_{k,d}$ such that

$$e(A(q,d)) \le c_{k,d} \cdot n^{-k} \cdot (\log n)^{(d-1)\cdot(k+1)},$$

where $n = n(q,d)$ is the number of knots used by $A(q,d)$.

It follows that $A(q,d)$ is almost optimal for F_d^k, i.e., up to logarithmic factors. More explicit error estimates are provided in [24].

5. INTEGRATION

An efficient and practical method was suggested in [15]. We use, for $d > 1$, the Smolyak construction and start, for $d = 1$, with the classical Clenshaw-Curtis formula with

$$m_1 = 1 \quad \text{and} \quad m_i = 2^{i-1} + 1 \ \text{ for } \ i > 1.$$

The Clenshaw Curtis formulas

$$U^i(f) = \sum_{j=1}^{m_i} f(x_j^i) \cdot a_j^i$$

use the knots

$$x_j^i = -\cos \frac{\pi(j-1)}{m_i - 1}$$

(and $x_1^1 = 0$). Hence we use non-equidistant knots. The weights a_j^i are defined in such a way that U^i is exact for all (univariate) polynomials of degree m_i. The resulting method has several remarkable properties:

- we use non-equidistant knots and so obtain much smaller errors for smooth functions;
- we obtain almost optimal quadrature formulas, simultaneously for each class F_d^k and C_d^k;
- no periodization is needed;
- the number of knots increases slowly:

$$n(d,d) = m_1^d = 1,$$

$$n(d+1,d) = 2d+1,$$

$$n(d+2,d) = 2d^2+2d+1;$$

- the resulting method has a high degree of exactness: $A(q,d)$ is exact for

$$\sum_{|\vec{i}|=q} (\mathbf{P}_{m_{i_1}} \otimes \cdots \otimes \mathbf{P}_{m_{i_d}}),$$

where $\mathbf{P}_m$ is the space of (univariate) polynomials of degree $\leq m$.

To comment on the last property, we give the following example. To integrate all polynomials of degree ≤ 5 exactly, we need

$$2d^2+2d+1$$

knots; the product Gauss rule needs 3^d knots. Hence the product Gauss rule is more efficient only for $d=1$ or $d=2$ but much less efficient for large d.

We did many numerical tests with the new method, see [15], and used the same test functions as [18], see also [4], [5], [10], [13] and [22]. The new method leads to very good results if the integrand is smooth enough and the dimension d is high enough. For small d, say $d \leq 5$, product Gauss rules often are better.

For discontinuous functions the results were poor in any dimension $d \leq 10$. We expect that number theoretic methods or Monte Carlo methods are usually better if the integrand has little smoothness.

6. GLOBAL OPTIMIZATION

The construction of Smolyak can also be used for the problem of global optimization, see [14]. The basic ideas are the following:

- use Smolyak's construction for the reconstruction of f; in this way we obtain an (almost) optimal method for all classes F_d^k and C_d^k;
- possibly use interval methods to find the minimum of the interpolating function;
- since adaptive methods are better on the average, see [2], we use an adaptive order of choosing hyperbolic cross points.

The paper [14] also contains numerical results which are promising, even if the dimension is relatively large, say $d = 50$.

REFERENCES

1. BAKHVALOV N. S., On approximate computation of integrals, Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem. 4, 3–18 (1959). [In Russian]

2. CALVIN J. M., Average performance of a class of adaptive algorithms for global optimization, preprint, Georgia Institute of Technology (1995).

3. DELVOS F.-J. & SCHEMPP W., Boolean methods in interpolation and approximation, Pitman Research Notes in Mathematics Series 230. Longman, Essex (1989).

4. GENZ A. C., A package for testing multiple integration subroutines, in: Keast, P., Fairweather, G., eds., Numerical Integration, 337–340, Kluwer, Dordrecht (1987).

5. GENZ A. C. & MALIK A. A., Algorithm 019: an adaptive algorithm for numerical integration over an n-dimensional rectangular region, J. Comp. Appl. Math. 6, 295–302 (1980).

6. GRIEBEL M., SCHNEIDER M. & ZENGER Ch., A combination technique for the solution of sparse grid problems, in: Beauwens, R., de Groen, P., eds., Iterative Methods in Linear Algebra, 263–281, Elsevier, North-Holland (1992).

7. HEINRICH S., Random approximation in numerical analysis, in Functional Analysis, K. D. Bierstedt et al., eds., Lecture Notes in Pure and Applied Mathematics 150, Marcel Dekker, New York, 123–171 (1993).

8. KACEWICZ B. Z., How to increase the order to get minimal-error algorithms for systems of ODE, Numer. Math. 45, 93–104 (1984).

9. KACEWICZ B. Z., On sequential and parallel solution of initial value problems, J. Complexity 6, 136–148 (1990).

10. KAHANER D. K., A survey of existing multidimensional quadrature routines, Contemporary Mathematics 115, 9–22 (1991).

11. NEMIROVSKI A. S., & YUDIN D. B., Problem Complexity and Method Efficiency on Optimization, John Wiley and Sons, Chicester (1983).

12. NOVAK E., Deterministic and stochastic error bounds in numerical analysis, Lecture Notes in Mathematics 1349, Springer-Verlag, Berlin (1988).

13. NOVAK E., On the power of adaption, J. Complexity, to appear.

14. NOVAK E. & RITTER K., Global optimization using hyperbolic cross points, in: State of the art in global optimization: computational methods and applications, C. A. Floudas, P. M. Pardalos (eds.), Kluwer, Dordrecht, 19–33 (1996).

15. NOVAK E. & RITTER K., High dimensional integration of smooth functions over cubes, Numer. Math., to appear.

16. PARDALOS P. M. & SHALLOWAY D. & XUE G. (eds.), Global minimization of nonconvex energy functions: molecular conformation and protein folding, DIMACS Series Vol. 23, American Math. Soc., Providence (1996).

17. PASKOV S. H. & TRAUB J. F., Faster valuation of financial derivatives, J. Portfolio Management 22, 113–120 (1995).

18. SLOAN I. H. & JOE S., Lattice Methods for Multiple Integration, Clarendon Press, Oxford (1994).

19. SMOLYAK S. A., Quadrature and interpolation formulas for tensor products of certain classes of functions, Soviet Math. Dokl. 4, 240-243 (1963).

20. TEMLYAKOV V. N., Approximation of Periodic Functions, Nova Science Publishers, New York (1994).

21. TRAUB J. F. & WASILKOWSKI G. W. & WOŹNIAKOWSKI H., Information-based complexity, Academic Press, San Diego (1988).

22. VAN DOOREN P. & DE RIDDER L., An adaptive algorithm for numerical integration over an n-dimensional cube, J. Comp. Appl. Math. 2, 207–217 (1976).

23. VESELY F. J., Computational physics, an introduction, Plenum Press, New York (1994).

24. WASILKOWSKI G. W. & WOŹNIAKOWSKI H., Explicit cost bounds of algorithms for multivariate tensor product problems, J. Complexity 11, 1–56 (1995).

25. WASILKOWSKI G. W. & WOŹNIAKOWSKI H., On tractability of path integration, J. Math. Phys. 37, 2071–2088 (1996).

26. WERSCHULZ A. G., The computational complexity of differential and integral equations, Oxford Science Publications, Oxford (1991).

27. WOŹNIAKOWSKI H., Average case complexity of linear multivariate problems, Part 1 and 2, J. Complexity 8, 337–392 (1992).

28. WOŹNIAKOWSKI H., Tractability and strong tractability of linear multivariate problems, J. Complexity 10, 96–128 (1994).

29. WOŹNIAKOWSKI H., Tractability and strong tractability of multivariate tensor product problems, in: Proc. of 2nd Gauss Symp., München 1993, J. Computing and Information 4, 1–19 (1994).

30. ZENGER Ch., Sparse grids, in: Hackbusch, W., ed., Parallel Algorithms for Partial Differential Equations, 241–251, Vieweg, Braunschweig (1991).