AVERAGE CASE COMPLEXITY
OF LINEAR MULTIVARIATE PROBLEMS

H. WOŹNIAKOWSKI

ABSTRACT. We study the average case complexity of a linear multivariate problem (LMP) defined on functions of d variables. We consider two classes of information. The first Λ^{std} consists of function values and the second Λ^{all} of all continuous linear functionals. Tractability of LMP means that the average case complexity is $O((1/\varepsilon)^p)$ with p independent of d. We prove that tractability of an LMP in Λ^{std} is equivalent to tractability in Λ^{all}, although the proof is *not* constructive. We provide a simple condition to check tractability in Λ^{all}.

We also address the optimal design problem for an LMP by using a relation to the worst case setting. We find the order of the average case complexity and optimal sample points for multivariate function approximation. The theoretical results are illustrated for the folded Wiener sheet measure.

1. INTRODUCTION

A linear multivariate problem (LMP) is defined as the approximation of a continuous linear operator on functions of d variables. Many LMP's are *intractable* in the worst case setting. That is, the worst case complexity of computing an ε-approximation is infinite or grows *exponentially* with d (see, e.g., [9]). For example, consider multivariate integration and function approximation of r times continuously differentiable functions of d variables. Then the worst case complexity is of order $(1/\varepsilon)^{d/r}$ assuming that an ε-approximation is computed using function values. Thus, if only continuity of the functions is assumed, $r = 0$, then the worst case complexity is infinite. For positive r, if d is large relative to r, then the worst case complexity is huge even for modest ε. In either case, the problem cannot be solved.

In this paper we study if tractability can be broken by replacing the worst case setting by an average case setting with a Gaussian measure on the space of functions. The average case complexity is defined as the minimal average cost of computing an approximation with average error at most ε. We consider two classes of information. The first class Λ^{std} consists of function values, and the second class Λ^{all} consists of all continuous linear functionals.

We say an LMP is tractable if the average case complexity is $O((1/\varepsilon)^p)$ with p independent of d. The smallest such p is called the exponent of the problem. Under mild assumptions, we prove that tractability in Λ^{all} is equivalent to tractability in Λ^{std} and that the difference of the exponents is at most 2. The proof of this result is *not* constructive. We provide, however, a simple condition to check tractability in Λ^{all}.

In particular, this means that multivariate integration is tractable in Λ^{std} and its exponent is at most 2. This should be contrasted with the worst case setting where, even for $d = 1$, the worst case complexity in Λ^{std} can be infinite or an arbitrary increasing function of $1/\varepsilon$ (see [14]). Of course, intractability of multivariate integration in the worst case setting can also be broken by switching to the randomized setting and using the classical Monte Carlo algorithm.

The optimal design problem of constructing sample points which achieve (or nearly achieve) the average case complexity of an LMP in Λ^{std} is a challenging problem. This problem has long been open even for multivariate integration and function approximation. In what follows, we will use the word "optimal" modulo a multiplicative constant which may depend on d but is independent of ε. Recently, the optimal design problem has been solved for multivariate integration for specific Gaussian measures (see [15] for the classical Wiener sheet measure, [5] for the the folded Wiener sheet measure, and [13] for the isotropic Wiener measure).

In this paper, we show under a mild assumption that tractability of function approximation (APP) implies tractability of other LMPs. Therefore, it is enough to address optimal sample points for APP. Optimal design for APP is analyzed by exhibiting a relation between average case and worst case errors of linear algorithms for APP. This relation reduces the study of the average case to the worst case for a different class of functions. This different class is the unit ball of a reproducing kernel Hilbert space whose kernel is given by the covariance kernel of the average case measure. Similar relations have been used in many papers for approximating continuous linear functionals; a thorough overview may be found in [11].

We illustrate the theoretical results for the folded Wiener sheet measure. In this case, an LMP is tractable and has exponent at most 2. For APP the exponents in Λ^{std} and Λ^{all} are the same. The exponent in Λ^{all} was known (see [4]), whereas the exponent in Λ^{std} was known to be at most 6 (see [3]). Tractability of APP for the folded Wiener sheet measure is in sharp contrast to intractability of APP for the isotropic Wiener measure; see [13].

Tractability of APP in the average case setting is significant, since it is known that the randomized setting does not help (see [12]). Thus, unlike multivariate integration, intractability of APP in the worst case setting cannot be broken by the randomized setting.

APP has been studied in Λ^{std} for $d = 1$ in [2, 6]. For $d \geq 1$, it was shown in [4] that the number of grid points needed to guarantee an average error ε depends exponentially on d. Of course, $O(\varepsilon^{-2-\delta})$ sample points are enough to compute an ε-approximation, $\delta > 0$. Hence, grid points are a poor choice of sample points.

In [4], the average case complexity of APP in Λ^{all} was found, and it was conjectured that the average case complexity in Λ^{std} is of the same order. We prove that this is indeed the case.

Optimal design for APP is solved by using a relation to the worst case setting in the reproducing kernel Hilbert space H. For the folded Wiener sheet measure, H is a Sobolev space of smooth *nonperiodic* functions which satisfy certain boundary conditions.

APP in the worst case setting has been studied in this Sobolev space

additionally assuming *periodicity* of functions in [7, 8] (see also [10] for $d = 2$). It was proven that *hyperbolic cross* points are optimal sample points. Hyperbolic cross points are defined as a subset of grid points whose indices satisfy a "hyperbolic" inequality. Approximation of periodic functions by trigonometric polynomials that use Fourier coefficients with these hyperbolic cross indices was first studied in [1].

For the nonperiodic case, optimal sample points for APP in the average case setting are derived from hyperbolic cross points, and the average case complexity is given by

$$\text{comp}^{\text{avg}}(\varepsilon; \text{APP}) = \Theta\left(\varepsilon^{-1/(r_{\min}+1/2)}\left(\log 1/\varepsilon\right)^{(k^*-1)(r_{\min}+1)/(r_{\min}+1/2)}\right),$$

with $r_{\min} = \min_{1 \le i \le d} r_i$, where $f^{(r_1,\ldots,r_d)}$ is continuous and where k^* denotes the number of r_i equal to $r_{\min}$. An optimal algorithm is given by a linear combination of function values at sample points derived from hyperbolic cross points.

Proofs of the results reported here can be found in [16].

2. LINEAR MULTIVARIATE PROBLEMS

A linear multivariate problem $\text{LMP} = \{\text{LMP}_d\}$ is a sequence of $\text{LMP}_d = (F, \mu, G, S, \Lambda)$. Each of F, μ, G, S, Λ may depend on d. We now define them in turn.

Let F be a separable Banach space of functions $f: D \to \mathbb{R}$, $F \subset L_2(D)$. Here, $D \subset \mathbb{R}^d$, and its Lebesgue volume $l(D)$ is in $(0, +\infty)$. We assume that all $L(f) = f(x)$ are in F^*.

The space F is equipped with a zero mean Gaussian measure μ. Let R_μ be the covariance kernel of μ, i.e., $R_\mu(t, x) = \int_F f(t) f(x) \mu(df)$ for $t, x \in D$.

Let $S: F \to G$ be a continuous linear operator, where G is a separable Hilbert space. Then $\nu = \mu S^{-1}$ is a zero mean Gaussian measure on the Hilbert space G. Its covariance operator $C_\nu = C_\nu^* \ge 0$ and has a finite trace.

Finally, Λ is either $\Lambda^{\text{all}} = F^*$ or Λ^{std} which consists of $L(f) = f(x), \forall f \in F$, for $x \in D$.

Our aim is to approximate elements $S(f)$ by $U(f)$. The latter is defined as follows. Information about f is gathered by computing a number of $L(f)$, where $L \in \Lambda$,

$$N(f) = [L_1(f), L_2(f), \ldots, L_n(f)], \quad \forall f \in F.$$

The choice of L_i and $n = n(f)$ may depend adaptively on the already computed information (see [9, Chapter 3]). Knowing $y = N(f)$, we compute $U(f) = \phi(y)$ for some $\phi: N(F) \to G$. The *average error* of U is defined as

$$e^{\text{avg}}(U) = \left(\int_F \|S(f) - U(f)\|^2 \mu(df)\right)^{1/2}.$$

To define the average cost of U, assume that each evaluation of $L(f)$, $L \in \Lambda$ and $f \in F$, costs $c = c(d) > 0$. Assume that we can perform arithmetic operations and comparisons on real numbers as well as addition of two elements from G and multiplying an element from G by a scalar; all of them with cost taken as unity. Usually $c \gg 1$.

For $U(f) = \phi(N(f))$, let $\text{cost}(N, f)$ denote the information cost of computing $y = N(f)$. Clearly, we have $\text{cost}(N, f) \geq cn(f)$. Let $n_1(f)$ denote the number of operations needed to compute $\phi(y)$ given $y = N(f)$. (It may happen that $n_1(f) = +\infty$.) The *average cost* of U is then given as

$$\text{cost}^{\text{avg}}(U) = \int_F (\text{cost}(N, f) + n_1(f)) \, \mu \, (df).$$

The *average case complexity* of LMP_d is the minimal cost of computing ε-approximations,

$$\text{comp}^{\text{avg}}(\varepsilon; \text{LMP}_d) = \inf\{\text{cost}^{\text{avg}}(U): U \text{ such that } e^{\text{avg}}(U) \leq \varepsilon\}.$$

To stress the dependence on certain parameters in $\text{comp}^{\text{avg}}(\varepsilon; \text{LMP}_d)$, we will sometimes list only those. Obviously, $\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) \leq \text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{std}})$. We show that the average case complexity functions in Λ^{all} and Λ^{std} are usually closely related.

3. TRACTABILITY OF LINEAR MULTIVARIATE PROBLEMS

An $\text{LMP} = \{\text{LMP}_d\}$ is called *tractable* if there exists $p \geq 0$ such that for all d

$$(3.1) \qquad \text{comp}^{\text{avg}}(\varepsilon; \text{LMP}_d) = O\left(c\,\varepsilon^{-p}\right).$$

The constant in the big O notation may depend on d. The infimum of the numbers p satisfying (3.1) is called the *exponent* $p^* = p^*(\text{LMP})$. To stress the role of the class Λ, we say that an LMP is *tractable in* Λ iff (3.1) holds for Λ.

In what follows, by *multivariate function approximation* we mean $\text{APP} = \text{LMP}$ with the embedding $S(f) = I_d(f) = f \in G = L_2(D)$, where the norm in $L_2(D)$ is denoted by $\| \cdot \|_d$.

We assume that for all d there exist $K_i = K_i(d)$, $i = 1, 2$, such that

$$(\text{A.1}) \qquad \|S(f)\| \leq K_1 \|f\|_d, \quad \forall f \in F,$$

$$(\text{A.2}) \qquad \|R_\mu(\cdot, \cdot)\|_{L_\infty(D)} \leq K_2.$$

Theorem 3.1. *Suppose* (A.1) *and* (A.2) *hold.*

(i) *Tractability of LMP in* Λ^{std} *is equivalent to tractability of LMP in* Λ^{all} *since*

$$\text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{all}}) = O(c\,\varepsilon^{-p(d)}) \text{ implies } \text{comp}^{\text{avg}}(\varepsilon; d, \Lambda^{\text{std}}) = O(c\,\varepsilon^{-p(d)-2}).$$

(ii) *Let* $\lambda_i(d)$ *be the ordered eigenvalues of the covariance operator of* μS^{-1}. *LMP is tractable in* Λ^{all} *iff there exists a positive number* α *such that for all* d,

$$(3.2) \qquad \sum_{i=n+1}^{+\infty} \lambda_i(d) = O(n^{-2\alpha}), \qquad \text{as } n \to +\infty.$$

The exponent of LMP is $p^* = 1/\sup\{\alpha : \alpha \text{ of } (3.2)\}$, *and* $p^* = +\infty$ *if there is no such* α.

(iii) *Tractability of APP in* Λ *with exponent* p^* *implies tractability of an LMP in* Λ *with exponent at most* p^* *provided LMP differs from APP only by the choice of* S.

We stress that the proof of Theorem 3.1 is *not* constructive. The exponents in Λ^{all} and Λ^{std} may differ by at most 2. The constant 2 is sharp. Indeed,

for the integration problem with the isotropic Wiener measure, the exponent in Λ^{std} is 2 (see [13]), and, obviously, the exponent in Λ^{all} is zero.

4. RELATION TO WORST CASE

Due to (iii) of Theorem 3.1, it is enough to analyze multivariate function approximation $\mathrm{APP} = \{\mathrm{APP}_d\}$ with $\mathrm{APP}_d = \{F, \mu, L_2(D), I_d, \Lambda^{\mathrm{std}}\}$. The *average* case errors of APP are related to *worst* case errors of the same I_d restricted to a specific subset of F. This specific subset of F is the unit ball BH_μ of a reproducing kernel Hilbert space H_μ. The space H_μ is the completion of finite-dimensional spaces of the form

$$\mathrm{span}\left(R_\mu(\cdot, x_1), R_\mu(\cdot, x_2), \ldots, R_\mu(\cdot, x_k)\right).$$

The completion is with respect to $\|\cdot\|_\mu = \langle \cdot, \cdot \rangle_\mu^{1/2}$, where $\langle R(\cdot, x), R(\cdot, t)\rangle_\mu = R(x, t)$.

Consider a linear U which uses sample points x_j. That is, we have $U(f) = \sum_{j=1}^n f(x_j)\, g_j$, where $g_j \in L_\infty(D)$. It is easy to show that

$$e^{\mathrm{avg}}(U) = e^{\mathrm{avg}}(U; \mathrm{APP}_d) = \left(\int_D \|h^*(\cdot, x)\|_\mu^2\, dx\right)^{1/2},$$

where $h^*(\cdot, x) = R_\mu(\cdot, x) - \sum_{j=1}^n g_j(x)\, R_\mu(\cdot, x) \in H_\mu$.

Consider now the same U for multivariate function approximation in the $L_\infty(D)$ norm

$$\mathrm{APP}_d^{\mathrm{wor}} = \{BH_\mu, L_\infty(D), I_d, \Lambda^{\mathrm{std}}\}$$

in the worst case setting. We now assume that H_μ is a subset of $L_\infty(D)$ and that the embedding I_d maps H_μ into $L_\infty(D)$. The worst error of U is equal to

$$e^{\mathrm{wor}}(U; \mathrm{APP}_d^{\mathrm{wor}}) = \sup\left\{\|f - U(f)\|_{L_\infty(D)} : \|f\|_\mu \le 1\right\}.$$

It is easy to show that $e^{\mathrm{wor}}(U; \mathrm{APP}_d^{\mathrm{wor}}) = \mathrm{ess\,sup}_{x\in D} \|h^*(\cdot, x)\|_\mu$, which yields

$$(4.1) \qquad e^{\mathrm{avg}}(U; \mathrm{APP}_d) \le \sqrt{l(D)}\, e^{\mathrm{wor}}(U; \mathrm{APP}_d^{\mathrm{wor}}),$$

where $l(D)$ is the Lebesgue volume of D.

5. APPLICATION FOR FOLDED WIENER SHEET MEASURES

We assume that μ is the folded Wiener sheet measure (see [4]). That is, $D = [0, 1]^d$ and F is the space of r_i times continuously differentiable functions with respect to x_i which vanish with their derivatives at points with at least one component equal to zero. The norm of F is the sup norm on $(r_1, \ldots, r_d)$ derivatives. The covariance kernel R_μ of μ is

$$R_\mu(t, x) = \prod_{j=1}^d \int_0^1 \frac{(t_j - s)_+^{r_j}}{r_j!} \frac{(x_j - s)_+^{r_j}}{r_j!}\, ds.$$

Observe that $R_\mu(t, t) \le 1$ and (A.2) holds with $K_2 \le 1$.

The space H_μ consists now of functions f of the form (see [5])

$$f(x) = \int_D \prod_{j=1}^d \frac{(x_j - t_j)_+^{r_j}}{r_j!}\, \phi(t_1, t_2, \ldots, t_d)\, dt_1\, dt_2 \cdots dt_d, \quad \forall x \in D, \ \phi \in L_2(D).$$

The inner product of H_μ is $\langle f, g \rangle_\mu = \int_D f^{(r_1, \ldots, r_d)}(t) \, g^{(r_1, \ldots, r_d)}(t) \, dt$.

Average case errors for APP_d can be bounded (see (4.1)) by analyzing the worst case of

$$APP_d^{wor} = \{BH_\mu, L_\infty(D), I_d, \Lambda^{std}\}.$$

Let W_0 be a subspace of H_μ of periodic functions for which $f^{(i_1, \ldots, i_d)}(t) = 0$ for all $i_j \leq r_j$ and all t from the boundary of D. Multivariate function approximation for the unit ball of W_0 in the worst case setting has been analyzed by Temlyakov in [7, 8]. He constructed sample points x_j and functions a_j such that for $T_n(f, x) = \sum_{j=1}^n f(x_j) a_j(x)$ we have

$$(5.1) \qquad \|f - T_n(f, \cdot)\|_{L_\infty(D)} = O(n^{-(r_{min}+1/2)} (\log n)^{(k^*-1)(r_{min}+1)}),$$

where $r_{min} = \min\{r_j : 1 \leq j \leq d\}$ and $k^* = \text{card}(\{j : r_j = r_{min}\})$.

The sample points x_j are called *hyperbolic cross points* and the functions a_j are obtained by linear combinations of the *de la Vallée-Poussin kernel*.

To extend Temlyakov's result to nonperiodic functions, define for f from BH_μ

$$g(x) = f(\vec{h}(x)), \quad \forall x \in D,$$

where $\vec{h}(x) = (h(x_1), h(x_2), \ldots, h(x_d))$ and $h(u) = 4u(1-u), \; \forall u \in [0, 1]$.

Observe that g is periodic and enjoys the same smoothness as f; that is, $g \in W_0$. There exists a constant $K = K(d, \vec{r})$ such that $\|g\|_\mu \leq K$. Define

$$U_n^*(f, t) = T_n(g, \vec{h}^{-1}(t)),$$

where $\vec{h}^{-1}(t) = (\frac{1}{2}(1 - \sqrt{1 - t_1}), \ldots, \frac{1}{2}(1 - \sqrt{1 - t_d})), \; t \in D$. We have

$$(5.2) \qquad U_n^*(f, t) = \sum_{j=1}^n f(\vec{h}(x_j)) \, a_j(\vec{h}^{-1}(t)) = \sum_{j=1}^n f(x_j^*) h_j^*(t),$$

where $x_j^* = \vec{h}(x_j)$, with a hyperbolic cross point x_j, and $h_j^*(t) = a_j(\vec{h}^{-1}(t))$.

It is possible to check that for all f from BH_μ we have

$$(5.3) \qquad \|f - U_n^*(f, \cdot)\|_{L_\infty(D)} = O(n^{-(r_{min}+1/2)} (\log n)^{(k^*-1)(r_{min}+1)}).$$

From (5.3) and (4.1) we conclude that

$$(5.4) \qquad comp^{avg}(\varepsilon; APP_d) = O(c \, \varepsilon^{-1/(r_{min}+1/2)} (\log 1/\varepsilon)^{(k^*-1)(r_{min}+1)/(r_{min}+1/2)}).$$

Clearly, $comp^{avg}(\varepsilon; APP_d)$ is bounded from below by the corresponding average case complexity in the class Λ^{all}. The latter was determined in [4]. These two average case complexity functions differ by at most a constant. Thus, the O in (5.4) can be replaced by Θ. Furthermore, the linear approximation U_n^* given by (5.2) is optimal, i.e., U_n^* computes an ε-approximation with the average cost $(c + 2)n$ which is minimal, modulo a constant, if

$$(5.5) \qquad n = O(\varepsilon^{-1/(r_{min}+1/2)} (\log 1/\varepsilon)^{(k^*-1)(r_{min}+1)/(r_{min}+1/2)}).$$

Theorem 5.1. *For APP the average case complexity functions* $comp^{avg}(\varepsilon; d, \Lambda^{std})$ *and* $comp^{avg}(\varepsilon; d, \Lambda^{all})$ *differ at most by a constant and*

$$comp^{avg}(\varepsilon; d, \Lambda^{std}) = \Theta(c \, \varepsilon^{-1/(r_{min}+1/2)} (\log 1/\varepsilon)^{(k^*-1)(r_{min}+1)/(r_{min}+1/2)}).$$

The linear U_n^* given by (5.2) which uses n sample points derived from the hyperbolic cross points with n given by (5.5) is optimal in the classes Λ^{std} and Λ^{all}.

From Theorem 5.1 we have that APP is tractable in Λ^{std} since $1/(r_{min} + 1/2) \leq 2$. The exponent of APP is the same in Λ^{all} and Λ^{std}. Since r_i may depend on d, we have

$$p^*(\Lambda^{std}) = (1/2 + \min\{r_j(d) : j = 1, 2, \ldots, d \text{ and } d = 1, 2, \ldots\})^{-1} \leq 2.$$

Obviously, any LMP which satisfies (A.1) and which is equipped with the folded Wiener sheet measure is tractable and has exponent at most $p^*(\Lambda^{std}) \leq 2$.

ACKNOWLEDGMENT

REFERENCES

1. K. I. Babenko, *On the approximation of a class of periodic functions of several variables by trigonometric polynomials*, Dokl. Akad. Nauk SSSR **132** (1960), 247–250, 982–985; English transl. in Soviet Math. Dokl. **1** (1960).

2. D. Lee, *Approximation of linear operators on a Wiener space*, Rocky Mountain J. Math. **16** (1986), 641–659.

3. A. Papageorgiou, *Average case complexity bounds for continuous problems*, Ph.D. thesis, Dept. of Computer Science, Columbia University, 1990.

4. A. Papageorgiou and G. W. Wasilkowski, *On the average complexity of multivariate problems*, J. Complexity **6** (1990), 1–23.

5. S. Paskov, *Average case complexity of multivariate integration for smooth functions*, (to appear in J. Complexity, 1993).

6. P. Speckman, *L_p approximation of autoregressive Gaussian processes*, Ph.D. thesis, Dept. of Math., UCLA, 1976.

7. V. N. Temlyakov, *Approximate recovery of periodic functions of several variables*, Math. USSR-Sb. **56** (1987), 249–261.

8. _____, Private communication, 1991.

9. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-based complexity*, Academic Press, New York, 1988.

10. G. Wahba, *Interpolating surfaces: high order convergence rates and their associated designs, with application to X-ray image reconstruction*, Dept. of Statistics, University of Wisconsin, 1978.

11. _____, *Spline models for observational data*, CBMS-NSF Regional Conf. Ser. in Appl. Math., vol. 59, SIAM, Philadelphia, PA, 1990.

12. G. W. Wasilkowski, *Randomization for continuous problems*, J. Complexity **5** (1989), 195–218.

13. _____, *Integration and approximation of multivariate functions: average case complexity with isotropic Wiener measure*, Bull. Amer. Math. Soc. (N.S.) **28** (1993) (to appear).

14. A. G. Werschulz, *Counterexamples in optimal quadratures*, Aequationes Math. **29** (1985), 183–202.

15. H. Woźniakowski, *Average case complexity of multivariate integration*, Bull. Amer. Math. Soc. (N.S.) **24** (1991), 185–194.

16. _____, *Average case complexity of linear multivariate problems, Part I: Theory, Part II: Applications*, Dept. of Computer Science, Columbia University, J. Complexity **8** (1992), 337–392.

DEPARTMENT OF COMPUTER SCIENCE, COLUMBIA UNIVERSITY AND INSTITUTE OF APPLIED MATHEMATICS, UNIVERSITY OF WARSAW
E-mail address: henryk@ground.cs.columbia.edu