Baltzer Journals October 30, 1995

Strong Tractability of Weighted Tensor Products *

Henryk Woźniakowski[1,2]

[1] *Department of Computer Science, Columbia University, New York, NY 10027, USA,*

[2] *Institute of Applied Mathematics, University of Warsaw, ul. Banacha 2, 00-097, Warsaw, Poland.*
E-mail: henryk@cs.columbia.edu

Dedicated to Ted Rivlin on the occasion of his 70th birthday

Columbia University Computer Science Department Report CUCS-021-96

We deal with approximating linear operators S_d that are defined as d weighted tensor products. We consider the worst case, average case, randomized and probabilistic settings depending on how the error of an approximation is defined. The complexity is understood as the minimal number of linear functionals which are needed to approximate S_d with error at most ε. Strong tractability means that there exist nonnegative K and p such that the complexity is bounded by $K\varepsilon^{-p}$ for all d and all $\varepsilon \le 1$. The smallest such p is called the strong exponent.

We provide necessary and sufficient conditions for weighted tensor products to be strongly tractable, and we find their strong exponent. In the worst case and randomized settings, these conditions are expressed in terms of singular values $\{\lambda_i\}$ of the problem for $d = 1$, and in terms of the weights $\{\beta_i\}$ of the problem. Strong tractability holds iff the sequences $\{\lambda_i\}$ and $\{\beta_i\}$ go to zero polynomially fast.

In the average case and probabilistic settings we consider Gaussian measures for which the traces of the covariance operators are uniformly bounded in d. We prove that strong tractability holds independently of the sequences $\{\lambda_i\}$ and $\{\beta_i\}$. In particular, if $\beta_i = 1, \forall\, i$, we get tensor product problems which are strongly tractable. The strong exponent in the average (or probabilistic) setting is always smaller than the strong exponent in the worst case setting. However, if the problem is not strongly tractable in the worst case setting, the strong exponent in the average case setting may be large.

We illustrate our analysis by the approximation of smooth periodic functions.

*The research reported here was supported in part by the National Science Foundation and the Air Force Office of Scientific Research.

Subject classification: AMS(MOS) 41A65, 41A63, 68Q25.

Keywords: multivariate linear problems, weighted tensor products, complexity, strong tractability.

1 Introduction

Linear multivariate problems are defined as the approximation of linear operators S_d that act on elements f of d variables. An approximation to $S_d(f)$ is obtained by computing a number of continuous linear functionals $L(f)$. Here, both f and $S_d(f)$ belong to Hilbert spaces. We wish to approximate S_d with error at most ε. We consider the worst case, average case, randomized and probabilistic settings, depending on how the error of an approximation is defined. The minimal number of such functionals needed to compute an approximation with error at most ε is called the complexity[1] and is denoted by $\mathrm{comp}(\varepsilon, d)$.

Strong tractability means that the complexity does not depend on d and is estimated by a polynomial in ε^{-1}, i.e., there exists two nonnegative numbers K and p such that

$$\mathrm{comp}(\varepsilon, d) \leq K\, \varepsilon^{-p}. \tag{1}$$

The smallest (or the infimum of) such p is called the strong exponent.

Strong tractability for general linear multivariate problems has been studied in [10], and for tensor product problems in [9, 11]. Strong tractability of tensor products in the worst case setting depends on the singular values $\{\lambda_i\}$ of the problem for $d = 1$. A problem is strongly tractable iff the largest singular value λ_1 is less than one, and λ_i goes to zero at least as fast as some power of i^{-1}.

The assumption $\lambda_1 < 1$ may be viewed as scaling of the problem since $\|S_1\| = \lambda_1$. Then for the d dimensional case we have $\|S_d\| = \lambda_1^d$, which goes exponentially fast to zero. Hence, if $\varepsilon > \lambda_1^d$ the problem becomes trivial since we can approximate $S_d(f)$ by zero and still the error is no more than ε. From this point of view, it seems natural to normalize the problem by taking $\|S_d\| = 1$, $\forall\, d$, which corresponds to $\lambda_1 = 1$. However, then we loose strong tractability in the worst case setting and, in fact, the complexity is roughly $d^{\gamma \log 1/\varepsilon}$ for some positive γ. Hence, the power of d goes to infinity as ε goes to zero. We finally add that scaling of linear multivariate problems and their strong tractability are interrelated with some surprising consequences, see [10].

To overcome the scaling problem and to still find strongly tractable problems even in the worst case setting, we study *weighted* tensor products in this paper.

[1] Usually, the complexity denotes the minimal cost needed to compute an approximation with error at most ε. Since we consider only linear problems over Hilbert spaces, the minimal cost is proportional to the minimal number of functionals needed to compute an approximation with error at most ε, see e.g., [6]. We choose this definition of the complexity to simplify the paper.

More precisely, we assume that

$$S_d = V_1 \otimes V_2 \otimes \cdots \otimes V_d.$$

Here the V_j's are linear operators of norm one, $\|V_j\| = 1$, such that $V_j^* V_j$ have the same eigenvectors, and their singular values are $\lambda_{1,j} = \lambda_1 = 1$ and $\{\beta_j \lambda_i\}$, $i = 2, 3, \ldots$, with $1 = \beta_1 \geq \beta_2 \geq \cdots \geq 0$. This means, that there exists a unit vector η such that for all $j = 1, 2, \ldots$ we have

$$\begin{aligned} V_j^* V_j \, \eta &= V_1^* V_1 \, \eta \\ V_j^* V_j \, \xi &= \beta_j^2 V_1^* V_1 \, \xi, \quad \forall \, \xi \text{ which is orthogonal to } \eta. \end{aligned}$$

We believe that for practical problems with large d, the successive dimensions are of different and, often, decreasing importance. This property may be modelled by the sequence $\{\beta_j\}$, and β_j serves as the weight of the jth dimensions. For small β_j the jth dimension becomes less important in approximating the operator S_d. That is why we call the problem a weighted tensor product.

We provide necessary and sufficient conditions for weighted tensor product problems to be strongly tractable, and we find their strong exponent. In the worst case setting, these conditions are expressed in terms of how fast the sequences $\{\lambda_i\}$ and $\{\beta_i\}$ go to zero. For a given nonincreasing and nonnegative sequence $\xi = \{\xi_i\}$, we say that p_ξ is the *polynomial-exponent* iff

$$p_\xi = \sup \left\{ p \geq 0 \, : \, \lim_{i \to \infty} \xi_i \, i^p = 0 \right\} \tag{2}$$

with the convention that $\sup \emptyset = 0$.

Then the weighted tensor product problem is strongly tractable in the worst case setting iff the polynomial-exponents of $\{\lambda_i\}$ and $\{\beta_i\}$ are positive, and then the strong exponent is

$$p^* = \max\{1/p_\lambda, 1/p_\beta\}.$$

It is known, see [4], that the randomized setting for approximating linear operators by arbitrary continuous functionals is closely related to the worst case setting. In particular, strong tractabilities in both settings are equivalent and the strong exponents are the same.

In the average and probabilistic settings we consider Gaussian measures. We choose the eigensystem of the covariance operator C_d of the Gaussian measure such that the importance of directions imposed by S_d is preserved. The eigenvalues of the correlation operator C_d are chosen such that the trace is finite and uniformly bounded in d. It turns out that this choice of the Gaussian measure implies strong tractability of the weighted tensor product problem in the average case setting independently of the behaviour of the sequences $\{\lambda_i\}$ and $\{\beta_i\}$. Hence, even for

$\beta_i = 1, \forall\, i$, we have strong tractability in the average case setting. In this case, we can set $V_j = V_1$, $\forall\, j$, and we have a tensor product problem.

The strong exponent in the average case setting depends on the sequences $\{\lambda_i\}$ and $\{\beta_i\}$ as well as on the eigenvalues of the correlation operators of the Gaussian measures for $d = 1, 2, \dots$. The strong exponent in the average case is always smaller than the strong exponent in the worst case. However, if the problem is not strongly tractable in the worst case than the strong exponent in the average case may be large.

In the probabilistic setting, we use known relations, see [6], with the average case setting and conclude that strong tractabilities in these two settings are equivalent, and the strong exponents are the same.

We illustrate our analysis for the approximation of smooth periodic functions. We show that this problem may be viewed as a weighted tensor product problem, and we find the strong exponents in the worst case and average case settings.

2 Weighted Tensor Product Problems

A weighted tensor product problem is defined as an approximation of a linear operator S_d which is given as a tensor product of operators V_j, $j = 1, 2, \dots, d$. More precisely, we proceed as follows. Let F_1 be a separable Hilbert space. The inner product in F_1 is denoted by $\langle \cdot, \cdot \rangle_{F_1}$. For $j = 1, 2, \dots$, consider

$$V_j : F_1 \to G_1,$$

where V_j is a continuous linear operator and G_1 is a separable Hilbert space.

Let $W_j = (V_j^* V_j)^{1/2} : F_1 \to F_1$. We scale the problem by assuming that $\|W_j\| = 1$. We also assume that W_j is compact and its eigensystem is $\{\lambda_{i,j}\}$ and $\{\eta_i\}$,

$$W_j\, \eta_i \;=\; \lambda_{i,j}\, \eta_i, \quad \langle \eta_i, \eta_j \rangle_{F_1} = \delta_{i,j},$$

where

$$\lambda_{1,j} \;=\; 1, \qquad \lambda_{i,j} \;=\; \beta_j\, \lambda_i, \quad i \;=\; 2, 3, \dots .$$

Here we assume that

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq 0, \qquad 1 = \beta_1 \geq \beta_2 \geq \cdots \geq 0.$$

This means that all W_j have the same eigenvectors η_i, whereas their eigenvalues are scaled through the sequence $\{\beta_j\}$.

We assume that $\{\eta_i\}$ forms an orthonormal basis for F_1. It is easy to see that $g_{i,j} = \lambda_{i,j}^{-1}\, V_j \eta_i$, for $\lambda_{i,j} > 0$, are orthonormal in G_1 and

$$V_j(f) \;=\; \sum_i \langle f, \eta_i \rangle_{F_1}\, \lambda_{i,j}\, g_{i,j}.$$

For $d \ge 2$, define the separable Hilbert space $F_d = F_1 \otimes \cdots \otimes F_1$ as a tensor product of F_1's. That is, F_d consists of linear combinations of d-tuples $(\eta_{i_1}, \eta_{i_2}, \dots \eta_{i_d})$, which we write for simplicity as $\eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}$. Hence, for $f \in F_d$ we have

$$f = \sum_{i_1,\dots,i_d=1}^{\infty} c_{i_1,\dots,i_d}\, \eta_{i_1}\, \eta_{i_2} \cdots \eta_{i_d} \tag{3}$$

with coefficients $c_{i_1,\dots,i_d}$ such that $\sum |c_{i_1,i_2,\dots,i_d}|^2 < +\infty$. The inner product in F_d is defined for $f = f_1 \cdots f_d$ and $h = h_1 \cdots h_d$ with $f_i, h_i \in F_1$ as

$$\langle f, h\rangle_{F_d} = \prod_{j=1}^{d} \langle f_i, h_i\rangle_{F_1}.$$

Then $\{\eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}\}$ is an orthonormal system of F_d. Note that the coefficient $c_{i_1,\dots,i_d} = \langle f, \eta_{i_1}\cdots\eta_{i_d}\rangle_{F_d}$ in (3).

Similarly we define the Hilbert space $G_d = G_1 \otimes \cdots \otimes G_1$, with the η_i replaced by an orthonormal system of G_1.

The linear operator $S_d = V_1 \otimes V_2 \otimes \cdots \otimes V_d$ is given by

$$S_d(f) = \sum_{i_1,\dots,i_d} \lambda_{i_1,1}\cdots\lambda_{i_d,d}\, \langle f, \eta_{i_1}\cdots\eta_{i_d}\rangle_{F_d}\, g_{i_1,1}\cdots g_{i_d,d}.$$

Clearly, for $f = f_1 \cdots f_d$ with $f_i \in F_1$ we have

$$S_d(f) = V_1(f_1)\, V_2(f_2) \cdots V_d(f_d).$$

It is easy to check that the eigensystem of $W_d = (S_d^* S_d)^{1/2} : F_d \to F_d$,

$$\{\gamma_{i,d}\} = \{\lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_d,d}\}, \tag{4}$$

$$\{\eta_{i,d}\} = \{\eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}\}. \tag{5}$$

is given by

$$\begin{aligned} W_d\eta_{i,d} &= \gamma_{i,d}\eta_{i,d}, \\ \langle \eta_{i,d}, \eta_{j,d}\rangle_{F_d} &= \delta_{i,j}, \\ \gamma_{1,d} &\ge \gamma_{2,d} \ge \cdots \ge \gamma_{n,d} \to 0. \end{aligned}$$

Hence, the eigensystem $\{\gamma_{i,d}, \eta_{i,d}\}$ is given as the product of the eigensystems $\{\lambda_{i,j}, \eta_i\}$ of the one dimensional problems.

3 Strong Tractability in the Worst Case Setting

We approximate $S_d(f)$ for f from the unit ball of F_d by computing continuous linear functionals $L_j(f)$ with $L_j \in F_d^*$ for $j = 1, 2, \dots, n$. Let

$$y \;=\; [L_1(f), L_2(f) \dots, L_n(f)].$$

Let $\phi : \mathbb{R}^n \to G_d$. Then $\phi(y)$ serves as an approximation to $S_d(f)$.

We seek the minimal n for which it is possible to find an approximation with error at most ε. Let $e(n,d)$ denote the minimal (worst case) error which can be achieved by computing n continuous linear functionals,

$$e(n,d) \;=\; \inf_{L_1, L_2, \dots, L_n \in F_d^*} \; \inf_{\phi} \; \sup_{f \in F_d, \|f\| \le 1} \|S_d(f) - \phi\left(L_1(f), L_2(f), \dots, L_n(f)\right)\|.$$

It is well known[2], see [3, 6, 7], that the L_j and ϕ minimizing $e(n,d)$ are given by

$$U(f) \;=\; \sum_{j=1}^{n} \langle f, \eta_{j,d} \rangle_{F_d} \, S_d\left(\eta_{j,d}\right),$$

and that

$$e(n,d) \;=\; \gamma_{n+1,d}.$$

Hence, the (worst case) complexity is given by

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \;=\; \min\left\{n : \; \gamma_{n+1,d} \le \varepsilon\right\}. \tag{6}$$

Observe first that the largest eigenvalue $\gamma_{1,d} = 1$. If $\varepsilon \ge 1$ then $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) = 0$. To omit this trivial case, assume that

$$\varepsilon \;<\; 1.$$

Observe further that if $\lambda_2 = 0$ then $\gamma_{i,d} = 0$ for all $i \ge 2$. Then $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) = 1$ for all $\varepsilon < 1$. To omit also this case, assume that

$$\lambda_2 \;>\; 0.$$

Due to (4) we can rewrite (6) as

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \;=\; \mathrm{card}\left\{[i_1, i_2, \dots, i_d] : \; i_j = 1, 2, \dots, \text{ such that } \lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_d,d} \;>\; \varepsilon\right\}.$$

By varying the index $i_d = 1, 2, \dots$, and using the fact that $\lambda_{1,d} = 1$ and $\lambda_{i,d} = \beta_d \lambda_i$ we get

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \;=\; \mathrm{comp}^{\mathrm{wor}}(\varepsilon, d-1) \;+\; \sum_{i=2}^{\infty} \mathrm{comp}^{\mathrm{wor}}\left(\varepsilon/(\beta_d \lambda_i), d-1\right). \tag{7}$$

We are ready to check which weighted tensor products are strongly tractable, see (1).

[2]It is also known that the adaptive selection of functionals L_j does not help, see [1, 2, 7].

Theorem .1
Let p_λ and p_β be polynomial-exponents of $\{\lambda_i\}$ and $\{\beta_i\}$, see (2).

(i) *A weighted tensor product problem is strongly tractable in the worst case setting iff p_λ and p_β are positive, and then the strong exponent is*

$$p^* \,=\, \max\{1/p_\lambda, 1/p_\beta\}\,.$$

(ii) *If a weighted tensor product problem is strongly tractable and*

$$\lambda_i \,\le\, K\, i^{-1/p}, \quad \text{with } p > p^*$$

then

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \,\le\, C_d\, \varepsilon^{-p},$$

where

$$C_d \,=\, K^p \prod_{j=2}^{d} \left(1 + \beta_j \sum_{i=2}^{\infty} \lambda_i^p\right) \,\le\, C^* \,=\, K^p \exp\left(\sum_{j=2}^{\infty} \beta_j^p \sum_{j=2}^{\infty} \lambda_j^p\right) \,<\, +\infty.$$

Proof
We first prove (i). Assume that the problem is strongly tractable and (1) holds with K and p. Take $d = 1$. Then

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, 1) \,=\, \max\{n :\ \lambda_n > \varepsilon\} \,\le\, K\, \varepsilon^{-p}.$$

For $j \,=\, \mathrm{comp}^{\mathrm{wor}}(\varepsilon, 1) + 1$ we have $\lambda_j \,\le\, \varepsilon$. By varying ε we conclude that $\lambda_j \,=\, O\left(j^{-1/p}\right)$. This proves that the polynomial-exponent p_λ of $\{\lambda_i\}$ is positive and $p_\lambda \ge 1/p$. Since p in (1) can be arbitrarily close to the strong exponent p^*, we have $p_\lambda \ge 1/p^*$.

To prove that the polynomial-exponent p_β is positive, observe that $\lambda_2\beta_j$, $j = 2, 3 \ldots, d$ are among the eigenvalues $\gamma_{i,d}$. Indeed, it is enough to take $i_k = 1$ and $i_j = 2$ for $k \ne j$ in (4). Hence,

$$K\, \varepsilon^{-p} \,\ge\, \mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \ge \max\{j :\ j \,\le\, d \text{ such that } \lambda_2\beta_j \,>\, \varepsilon\}.$$

Since this holds for all d and λ_2 is positive, we conclude as before that p_β is positive and $p_\beta \ge 1/p^*$. This also proves that

$$p^* \,\ge\, \max\{1/p_\lambda, 1/p_\beta\}.$$

Assume now that p_λ and p_β are positive. Take $p > 1/p_\lambda$. Then there exists a positive $\delta < p_\lambda$ such that $p > 1/(p_\lambda - \delta)$ and $\lambda_i = O(i^{-(p_\lambda - \delta)})$. This yields that there exists a positive constant C such that

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, 1) \,\le\, C\, \varepsilon^{1/(p_\lambda - \delta)} \,\le\, C\, \varepsilon^{-p}.$$

Observe also that the series $\sum_{i=2}^{\infty} \lambda_i^p$ is convergent. Indeed,

$$\sum_{i=2}^{\infty} \lambda_i^p \;=\; O\left(\sum_{i=2}^{\infty} i^{-p(p_\lambda-\delta)}\right) \;<\; +\infty$$

since $p(p_\lambda - \delta) > 1$.

Assume inductively that $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, j) \;\le\; C_j \varepsilon^{-p}$ for some positive C_j and $j \;\le\; d-1$. Then (7) implies

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \;\le\; C_{d-1}\, \varepsilon^{-p} \;\left(1 \;+\; \beta_d^p \sum_{i=2}^{\infty} \lambda_i^p\right).$$

This yields

$$C_d \;=\; C_{d-1}\left(1 \;+\; \beta_d^p \sum_{i=2}^{\infty} \lambda_i^p\right) \;=\; C\, \prod_{j=2}^{d}\left(1 \;+\; \beta_j^p \sum_{i=2}^{\infty} \lambda_i^p\right).$$

Note that the sequence $\{C_d\}$ is nondecreasing and it is convergent to C_∞ iff the series $\sum_{j=2}^{\infty} \beta_j^p$ is convergent. As shown for the sequence $\{\lambda_j\}$, this holds if $p > 1/p_\beta$. Hence, if we take $p > \max\{1/p_\lambda, 1/p_\beta\}$ then we have

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \;\le\; C_\infty \varepsilon^{-p}.$$

This proves that the problem is strongly tractable. Since p can be arbitrarily close to $\max\{1/p_\lambda, 1/p_\beta\}$, we conclude that $p^* \le \max\{1/p_\lambda, 1/p_\beta\}$. This proves (i).

We are also almost done with (ii). Indeed, observe that $\lambda_i \le K i^{-1/p}$ implies that

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, 1) \le K^p \varepsilon^{-p}.$$

For $d \ge 2$, we estimate $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d)$ as above with $C = K^p$. Finally, it is enough to observe that $1 + x \;\le\; \exp(x)$ and therefore $C_\infty \;\le\; C^*$. This proves (ii) and completes the proof of Theorem .1. □

From the proof of Theorem .1 we can easily conclude the behavior of the ordered eigenvalues $\{\gamma_{j,d}\}$ of the operator W_d.

Corollary .2
If the weighted tensor product problem is strongly tractable with the strong exponent p^* then for every $p > p^*$ there exists a positive $K = K(p)$ such that

$$\gamma_{j,d} \;\le\; K\, j^{-1/p}, \quad \forall\, j, d \;=\; 1, 2, \ldots.$$

4 Strong Tractability in the Average Case Setting

In this section we consider a weighted tensor product problem from Section 2 in the average case setting. We equip the tensor product space F_d with a zero mean Gaussian measure μ_d. To define μ_d, it is enough to define its covariance operator $C_d : F_d \to F_d$. The operator C_d is self adjoint, nonnegative definite, and has finite trace. We select an eigensystem of C_d to preserve the importance of directions in the space F_d for the operator W_d, see (4). Hence, we assume that

$$C_d \eta_{i_1} \eta_{i_2} \cdots \eta_{i_d} \;=\; \lambda'_{i_1,1} \lambda'_{i_2,2} \cdots \lambda'_{i_d,d}\, \eta_{i_1} \eta_{i_2} \cdots \eta_{i_d},$$

where the eigenvalues $\lambda'_{i,j}$ resemble the eigenvalues of W_j, $j = 1, 2, \ldots$,

$$\lambda'_{1,j} \;=\; 1, \quad \lambda'_{i,j} \;=\; \beta'_j \lambda'_i, \quad i \;=\; 2, 3, \ldots,$$

and

$$1 \;=\; \lambda'_1 \;\geq\; \lambda'_2 \;\geq\; \cdots \;\geq\; 0, \qquad 1 \;=\; \beta'_1 \;\geq\; \beta'_2 \;\geq\; \cdots \;\geq\; 0.$$

Then the trace of the covariance operator is

$$\mathrm{trace}(C_d) \;=\; \prod_{j=1}^{d} \left(1 + \beta'_j \sum_{i=2}^{\infty} \lambda'_i \right).$$

To guarantee that the trace(C_d) is finite and bounded uniformly in d we assume that the series of λ'_i and the series of β'_i are finite. This is equivalent to assuming that the polynomial-exponents of $\{\lambda'_i\}$ and $\{\beta'_i\}$ are greater than one. Hence we assume

$$\min\{p_{\lambda'}, p_{\beta'}\} \;>\; 1. \tag{8}$$

In the average case setting, we want to compute an approximation with *average* error at most ε. Hence, the minimal (average case) error $e(n,d)$ which can be achieved by computing n continuous linear functionals is given by

$$e^2(n,d) \;=\; \inf_{L_1, L_2, \ldots, L_n \in F_d^*} \; \inf_{\phi} \int_{F_d} \| S_d(f) - \phi\,(L_1(f), L_2(f), \ldots, L_n(f)) \|^2 \, \mu_d(df).$$

It is known[3], see e.g., [6], that L_j and ϕ which minimize $e(n,d)$ are given by

$$U(f) \;=\; \sum_{j=1}^{n} \left\langle S_d(f), \eta^*_{j,d} \right\rangle_{G_d} \eta^*_{j,d},$$

[3] It is also known that the adaptive selection of functionals L_j can only help by at most one, see [8].

where $\eta^*_{j,d}$ are the eigenvectors of the covariance operator C_{ν_d} of the Gaussian measure $\nu_d = \mu S_d^{-1}$. The eigenvectors $\eta^*_{j,d}$ correspond to the n largest eigenvalues $\lambda^*_{1,d} \geq \lambda^*_{2,d} \geq \cdots \geq 0$.

We now check that

$$\begin{aligned} \left\{\lambda^*_{i,d}\right\} &= \left\{\lambda^2_{i_1,1}\lambda^{'}_{i_1,1}\lambda^2_{i_2,2}\lambda^{'}_{i_2,2}\cdots\lambda^2_{i_d,d}\lambda^{'}_{i_d,1}\right\}, \\ \left\{\eta^*_{i,d}\right\} &= \left\{S_d\left(\eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}\right)\right\}. \end{aligned}$$

Indeed, observe first that the measure ν_d is concentrated on the subspace $S_d(F_d)$, and that

$$\left\langle S_d(\eta_{\vec{i}}), S_d(\eta_{\vec{j}})\right\rangle_{G_d} = \left\langle W_d^2(\eta_{\vec{i}}), \eta_{\vec{j}}\right\rangle_{F_d} = \lambda^2_{\vec{i}}\,\delta_{\vec{i},\vec{j}},$$

where we use the notation $\eta_{\vec{i}} = \eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}$, and $\lambda_{\vec{i}} = \lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_d,d}$. Hence, $\{S_d(\eta_{\vec{i}})\}$ is an orthogonal basis of $S_d(F_d)$. Consider

$$a = \int_{F_d} \langle S_d(f), S_d(\eta_{\vec{i}})\rangle_{G_d} \left\langle S_d(f), S_d(\eta_{\vec{j}})\right\rangle_{G_d} \mu_d(df).$$

Change variables by setting $g = S_d(f)$. Then

$$a = \int_G \langle g, S_d(\eta_{\vec{i}})\rangle_{G_d} \left\langle g, S_d(\eta_{\vec{j}})\right\rangle_{G_d} \nu_d(dg) = \left\langle C_{\nu_d}(S_d(\eta_{\vec{i}})), S_d(\eta_{\vec{j}})\right\rangle_{G_d}.$$

On the other hand,

$$\begin{aligned} a &= \int_{F_d} \left\langle f, W_d^2(\eta_{\vec{i}})\right\rangle_{F_d} \left\langle f, W_d^2(\eta_{\vec{j}})\right\rangle_{F_d} \mu_d(df) \\ &= \lambda^2_{\vec{i}}\,\lambda^2_{\vec{j}} \int_{F_d} \langle f, \eta_{\vec{i}}\rangle_{F_d} \left\langle f, \eta_{\vec{j}}\right\rangle_{F_d} \mu_d(df) \\ &= \lambda^2_{\vec{i}}\,\lambda^2_{\vec{j}} \left\langle C_d(\eta_{\vec{i}}), \eta_{\vec{j}}\right\rangle_{F_d} \\ &= \lambda^2_{\vec{i}}\,\lambda^2_{\vec{j}}\,\lambda^{'}_{\vec{i}}\,\delta_{\vec{i},\vec{j}}. \end{aligned}$$

Hence,

$$\left\langle C_{\nu_d}(S_d(\eta_{\vec{i}})), S_d(\eta_{\vec{j}})\right\rangle_{G_d} = \lambda^2_{\vec{i}}\,\lambda^2_{\vec{j}}\,\lambda^{'}_{\vec{i}}\,\delta_{\vec{i},\vec{j}}.$$

This shows that $C_{\nu_d}(S_d(\eta_{\vec{i}}))$ is parallel to $S_d(\eta_{\vec{i}})$ and

$$C_{\nu_d}(S_d(\eta_{\vec{i}})) = \lambda^2_{\vec{i}}\,\lambda^{'}_{\vec{i}}\,S_d(\eta_{\vec{i}}),$$

as claimed.

It is also known that

$$e^2(n,d) = \sum_{i=n+1}^{\infty} \lambda^*_{n+1,d}.$$

Hence, the (average case) complexity is given by

$$\mathrm{comp}^{\mathrm{ave}}(\varepsilon, d) \;=\; \min\left\{ n : \sum_{i=n+1}^{\infty} \lambda^*_{n+1,d} \le \varepsilon^2 \right\}.$$

We are ready to prove that (8) implies strong tractability of the weighted tensor product problem in the average case setting independently of the behaviour of sequences $\{\lambda_i\}$ and $\{\beta_i\}$.

Theorem .3
If $\min\{p_{\lambda'}, p_{\beta'}\} \;>\; 1$ *then*

(i) *the weighted tensor product problem is strongly tractable in the average case setting and the strong exponent is*

$$p^* \;=\; \frac{2}{\min\left\{2p_\lambda + p_{\lambda'}, 2p_\beta + p_{\beta'}\right\} \,-\, 1}.$$

(ii) *If*

$$\lambda_i^2 \lambda_i' \;\le\; K\, i^{-1/p} \quad \text{for} \quad p \,\in\, (1/\min\{2p_\lambda + p_{\lambda'}, 2p_\beta + p_{\beta'}\}, 1)$$

then

$$\mathrm{comp}^{\mathrm{ave}}(\varepsilon, d) \;\le\; \left\lceil C^*\, \varepsilon^{-2p/(1-p)} \right\rceil$$

where

$$C^* \;=\; \left(2\, K^p \,\exp\left(\sum_{j=2}^{\infty} (\beta_j^2 \beta_j')^p \sum_{j=2}^{\infty} (\lambda_i^2 \lambda_i')^p \right) \right)^{1/(1-p)} \left(\frac{p}{1-p} \right)^{p/(1-p)}.$$

Proof
We first show (i). It is proven in [11], Theorem 5.1, that the problem is strongly tractable in the average case setting iff there exist $p \in (0,1)$, a positive A, and an integer k such that

$$\lambda^*_{j,d} \;\le\; A\, j^{-1/p}, \quad \forall\, d \;\text{ and for}\, j \;=\; k, k+1, \ldots, \tag{9}$$

and then the strong exponent $p^* = \inf\, 2p/(1-p)$ for p satisfying (9). In our case all $\lambda^*_{j,d} \le 1$ and we can set $k = 1$ in (9).

Hence, it is enough to deal with the sequence $\{\lambda^*_{j,d}\}$ which corresponds to the sequence $\{\gamma_{j,d}\}$ of Section 2 with the change of the sequences $\{\lambda_i\}$ and $\{\beta_i\}$ by the sequences $\{\lambda_i^2 \lambda_i'\}$ and $\{\beta_i^2 \beta_i'\}$.

Observe that the polynomial-exponent $p_{\lambda^2\lambda'}$ of the sequence $\{\lambda_i^2\lambda_i'\}$ is $2p_\lambda + p_{\lambda'}$. Similarly, $p_{\beta^2\beta'} = 2p_\beta + p_{\beta'}$. Since $\min\{p_{\lambda'}, p_{\beta'}\} > 1$ then

$$q \;=\; \max\left\{1/p_{\lambda^2\lambda'}, 1/p_{\beta^2\beta'}\right\} \;=\; \max\left\{1/(2p_\lambda + p_{\lambda'}), 1/(2p_\beta + p_{\beta'})\right\} \;<\; 1.$$

We can now formally consider the weighted tensor product with the eigenvalues $\lambda_i^2\lambda_i'$ and $\beta_i^2\beta_i'$ in the worst case setting, and conclude from Theorem .1 that the problem is strongly tractable with strong exponent q. Then we apply Corollary .2 to conclude that (9) holds for $q < p < 1$. Hence, our problem is strongly tractable in the average case setting and its strong exponent p^* is obtained when p approaches q, i.e.,

$$p^* \;=\; 2q/(1-q) \;=\; 2/(\min\{2p_\lambda + p_{\lambda'}, 2p_\beta + p_{\beta'}\} - 1),$$

as claimed in (i).

To prove (ii), we first estimate the worst case complexity $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d)$ of the weighted tensor product problem for the eigenvalues $\{\lambda_i^2\lambda_i'\}$ and $\{\beta_i^2\beta_i'\}$. Since $\lambda_i^2\lambda_i' \;\le\; K\, i^{-1/p}$ then (ii) of Theorem .1 yields

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) \;\le\; C_\infty\, \varepsilon^{-p},$$

where $C_\infty = K^p \exp\left(\sum_{j=2}^\infty(\beta_j^2\beta_j')^p \sum_{j=2}^\infty(\lambda_i^2\lambda_i')^p\right)$.

Since $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, d) = \min\{n : \lambda_{n+1,d}^* \le \varepsilon\}$, $\forall\, \varepsilon$, we conclude that

$$\lambda_{j,d}^* \le\; (2C_\infty)^{1/p}\, j^{-1/p}.$$

To estimate the average case complexity $\mathrm{comp}^{\mathrm{ave}}(\varepsilon, d)$, we need to estimate $\sum_{j=n+1}^\infty \lambda_{j,d}^*$. We have

$$\sum_{j=n+1}^\infty \lambda_{j,d}^* \;\le\; (2C_\infty)^{1/p} \int_n^\infty x^{-1/p} dx \;=\; \frac{(2C_\infty)^{1/p} p}{1-p}\, n^{-(1-p)/p}.$$

Hence,

$$\mathrm{comp}^{\mathrm{ave}}(\varepsilon, d) \;=\; \min\left\{ n : \sum_{j=n+1}^\infty \lambda_{j,d}^* \;\le\; \varepsilon^2 \right\} \;\le\; \left\lceil C^*\, \varepsilon^{-2p/(1-p)} \right\rceil$$

with $C^* = (2C_\infty)^{1/(1-p)}(p/(1-p))^{p/(1-p)}$, as claimed. $\square$

We briefly compare the strong exponents in the worst case,

$$p^{\mathrm{wor}} \;=\; \max\{1/p_\lambda, 1/p_\beta\},$$

and in the average case,

$$p^{\text{ave}} \;=\; 2/\left(\min\{2p_\lambda + p_{\lambda'}, 2p_\beta + p_{\beta'}\} \;-\; 1\right).$$

Clearly,

$$p^{\text{ave}} \;<\; p^{\text{wor}}.$$

However, if both $p_{\lambda'}$ and $p_{\beta'}$ are close to one then $p^{\text{ave}} \approx p^{\text{wor}}$.

Suppose now that $\beta_i = 1, \forall\, i$. Then we may take $V_j \;=\; V_1$ and obtain a tensor product problem. This tensor product problem is not strongly tractable in the worst case setting, however, it is strongly tractable in the average case with the strong exponent

$$p^{\text{ave}} \;=\; \frac{2}{\min\{2p_\lambda + p_{\lambda'}, p_{\beta'}\} \;-\; 1}.$$

This exponent is large if $p_{\beta'}$ is close to one.

5 Strong Tractability in the Randomized and Probabilistic Settings

We briefly discuss two other settings. In the randomized setting, we assume that continuous linear functionals L_j as well as the mapping ϕ can be randomly chosen. That is, an approximation

$$U_t(f) \;=\; \phi_t\left(L_{1,t}(f), L_{2,t}(f), \dots, L_{n,t}(f)\right)$$

depends on a random parameter t from a space T with distribution ρ. The (randomized) error of U is defined as the expected error with respect to t and the worst case with respect to f,

$$e^{\text{ran}}(U)^2 \;=\; \sup_{f\in F_d,\, \|f\|\le 1} \int_T \|S_d(f) - U_t(f)\|^2 \rho(dt).$$

Let $e^{\text{ran}}(n,d)$ denote the minimal (randomized) error of such U. Then the (randomized) complexity is

$$\text{comp}^{\text{ran}}(\varepsilon, d) \;=\; \min\left\{n:\; e^{\text{ran}}(n,d) \;\le\; \varepsilon\right\}.$$

It is proven in [4] that the randomized and worst case complexity are closely related,

$$\frac{1}{2}\,\text{comp}^{\text{wor}}\left(\sqrt{2}\,\varepsilon, d\right) \;\le\; \text{comp}^{\text{ran}}(\varepsilon, d) \;\le\; \text{comp}^{\text{wor}}(\varepsilon, d).$$

This estimate implies that strong tractability in the randomized setting is equivalent to strong tractability in the worst case setting, and the strong exponents are the same. Hence, Theorem .1 can be also used for the randomized setting.

We finally consider the probabilistic setting. In this setting we use deterministic approximations

$$U(f) \;=\; \phi\,(L_1(f), L_2(f), \dots, L_n(f))$$

and define the (probabilistic) error of such U as in the worst case setting but by neglecting a set of measure at most δ. That is,

$$e^{\text{pro}}(U) \;=\; \inf_{A:\, \mu_d(A)\le\delta}\; \sup_{f\in F_d - A} \|S_d(f) - U(f)\|.$$

Let $e^{\text{pro}}(n,d)$ denote the minimal (probabilistic) error of such U. Then the (probabilistic) complexity is

$$\text{comp}^{\text{pro}}(\varepsilon,d) \;=\; \min\,\{n:\; e^{\text{pro}}(n,d) \;\le\; \varepsilon\,\}\,.$$

Here, the parameter $\delta \in [0,1]$ and obviously small δ is of particular importance. To stress the dependence on δ we write $\text{comp}^{\text{pro}}(\varepsilon,d) = \text{comp}^{\text{pro}}(\varepsilon,\delta,d)$.

For Gaussian measures, the probabilistic complexity depends very weakly on δ through a power of $\ln\, 1/\delta$, see [6]. Therefore, we define strong tractability in the probabilistic setting iff there exists nonnegative K, p, q such that

$$\text{comp}^{\text{pro}}(\varepsilon,\delta,d) \;\le\; K\,\varepsilon^{-p}\,(\ln\, 1/\delta)^q, \quad \forall\,\varepsilon,\delta \in (0,1] \;\text{ and }\; d = 1,2,\dots.$$

It is known that the probabilistic and average case settings for Gaussian measures and approximations of linear operators are closely related, see [6]. In particular, we have

$$\text{comp}^{\text{pro}}(\varepsilon,\delta,d) \;\le\; \text{comp}^{\text{ave}}\left(\varepsilon/\sqrt{2\ln\, 5/\delta}, d\right)$$

and for any positive α and any small δ there exists a positive ε_0 such that

$$\text{comp}^{\text{pro}}(\varepsilon,\delta,d) \;\ge\; \text{comp}^{\text{ave}}\left(\varepsilon^{1-\alpha}/\sqrt{2\ln\, 5/\delta}, d\right), \qquad \forall\,\varepsilon \le \varepsilon_0.$$

This shows that strong tractability in the probabilistic setting is equivalent to strong tractability in the average case setting, and the strong exponents in ε^{-1} are the same, whereas the strong exponent in $\ln\, 1/\delta$ is at most equal to half of the strong exponent in ε^{-1}. Hence, Theorem .3 can be also used for the probabilistic setting.

6 Approximation of Smooth Periodic Functions

We illustrate the results of the previous sections for the approximation problem of smooth periodic functions.

Consider the Fourier expansion of a complex valued periodic function f defined on $[0,1]^d$,

$$f(x) \;=\; \sum_{h\in\mathbf{Z}^d} \hat{f}(h)\, e_h(x)$$

where $e_h(x) \;=\; \exp\left(2\,\pi\, i\,(h,x)\right)$, $i \;=\; \sqrt{-1}$, and $(h,x) \;=\; \sum_{j=1}^d h_j x_j$ for integers h_j. We assume that the Fourier coefficients

$$\hat{f}(h) \;=\; \int_{[0,1]^d} f(x)\, \exp\left(-2\,\pi\, i\,(h,x)\right)\, dx$$

satisfy the condition

$$\sum_{h\in\mathbf{Z}^d} |\hat{f}(h)|^2 \left(\overline{h}_1 \overline{h}_2 \cdots \overline{h}_d\right)^{2\alpha} \;<\; +\infty,$$

where $h \;=\; [h_1,h_2,\ldots,h_d]$ and $\overline{h}_j \;=\; \max\{1, |h_j|/\xi_j\}$. Here the parameter $\alpha > 0$, and the sequence $\{\xi_j\}$ is such that

$$\xi_{-j} \;=\; \xi_j, \qquad \text{and} \;\; 1 \ge \xi_1 \;\ge\; \xi_2 \;\ge\; \cdots \;>\; 0.$$

We define H_d as the Hilbert space of such functions with the inner product

$$\langle f,g\rangle_{H_d} \;=\; \sum_{h\in\mathbf{Z}^d} \hat{f}(h)\, \overline{\hat{g}(h)} \left(\overline{h}_1 \overline{h}_2 \cdots \overline{h}_d\right)^{2\alpha}.$$

The space H_d (with $\xi_j = 1$) is often studied for lattice methods for multivariate integration, see [5]. It is known that the parameter α indicates the smoothness of the functions from H_d. For example, if $\alpha > 1$ is an integer then f is $(\alpha-1)$ times differentiable with respect to all variables, and after such differentiations the function has bounded variation in the sense of Hardy and Krause, see [5].

We want to approximate functions from H_d in the $G_d \;=\; L_2([0,1])^d$ norm. That is, we approximate $A_d(f)$, where

$$A_d : H_d \;\to\; G_d, \qquad A_d(f) \;=\; f.$$

We first show that this approximation problem can be viewed as a weighted tensor product problem. We begin with the worst case setting. Let $T_d : H_d \to G_d$ be given by

$$T_d f \;=\; \sum_{h\in\mathbf{Z}^d} \hat{f}(h) \left(\overline{h}_1 \overline{h}_2 \cdots \overline{h}_d\right)^{\alpha}\, e_h.$$

Then $\|f\|_{H_d} \;=\; \|T_d f\|_{G_d}$ and for any linear $U : H_d \to G_d$ we have

$$\sup_{F\in H_d,\, \|f\|_{H_d}\le 1} \|A_d(f) - U(f)\|_{G_d} \;=\; \sup_{g\in G_d,\, \|g\|_{G_d}\le 1} \|T_d^{-1}(g) - U T_d^{-1}(g)\|_{G_d}.$$

Hence, approximation of the operator A_d over the unit ball of H_d is equivalent to approximation of the operator $T_d^{-1} : G_d \to H_d \subset G_d$ over the unit ball of G_d where

$$T_d^{-1}(f) \;=\; \sum_{h \in \mathbf{Z}^d} \hat{f}(h) \left(\overline{h}_1 \overline{h}_2 \cdots \overline{h}_d\right)^{-\alpha} e_h, \qquad \forall\, f \in G_d.$$

Define $F_1 \;=\; L_2([0,1])$ and

$$V_j(f) \;=\; \sum_{h=-\infty}^{+\infty} \hat{f}(h) \max\{1, |h|/\xi_j\}^{-\alpha} e_h, \qquad j \;=\; 1, 2, \dots .$$

Observe that $V_j^* \;=\; V_j$, and therefore $W_j \;=\; (V_j^* V_j)^{1/2} \;=\; V_j$ and

$$\begin{aligned} V_j\, e_0 &= e_0, \\ V_j\, e_h &= \xi_j^{\alpha}\, |h|^{-\alpha}\, e_h. \end{aligned}$$

Hence, the sequences $\{\lambda_j\}$ and $\{\beta_j\}$ from Section 2 are now given by

$$\lambda_{2j} \;=\; \lambda_{2j-1} \;=\; j^{-\alpha} \qquad \beta_j \;=\; \xi_j^{\alpha},$$

and the eigenvectors η_j are given by

$$\eta_{2j} \;=\; e_j, \qquad \eta_{2j-1} \;=\; e_{-j}.$$

Clearly,

$$T_d^{-1} \;=\; V_1 \otimes V_2 \otimes \cdots \otimes V_d.$$

Therefore, setting $F_d \;=\; G_d$ we have a weighted tensor product problem. Since $(T_d^{-1})^* \;=\; T_d^{-1}$ and $W_d \;=\; \left((T_d^{-1})^* T_g^{-1}\right)^{1/2} \;=\; T_d^{-1}$, the eigensystem of (4) is defined in terms of ξ_j and e_j as above.

From Theorem .1 we know that strong tractability holds iff $p_\lambda \;>\; 0$ and $p_\beta \;>\; 0$. In our case, $p_\lambda \;=\; \alpha$ is positive, and $p_\beta \;=\; \alpha\, p_\xi$. Hence, strong tractability in the worst case setting holds iff $p_\xi \;>\; 0$. Then the strong exponent is

$$p^{\mathrm{wor}} \;=\; \alpha^{-1} \max\{1/p_\xi, 1\}.$$

We briefly consider the average case setting. Let μ_{H_d} be a zero mean Gaussian measure on the Hilbert space H_d. Then

$$\int_{H_d} \|A_d(f) - U(f)\|^2\, \mu_{H_d}(df) \;=\; \int_{G_d} \|T^{-1}(g) - UT^{-1}(g)\|^2 \mu_d(df),$$

where $\mu_d \;=\; \mu_{H_d} T^{-1}$ is also a zero mean Gaussian whose covariance operator C_d is given in terms of the covariance operator C_{H_d} of the measure μ_{H_d} by

$$C_d \;=\; T_d\, C_{H_d}\, T_d^{-1}.$$

This shows that approximation of A_d in the average case setting with the Gaussian measure μ_{H_d} is equivalent to approximation of T_d^{-1} in the average case setting with the Gaussian measure μ_d.

Choosing the Gaussian measure μ_d as in Section 3, we get strong tractability independently of the sequence ξ_j, and the strong exponent is

$$p^{\mathrm{ave}} = \frac{2}{\min\left\{2\alpha + p_{\lambda'}, 2\alpha p_\xi + p_{\beta'}\right\} - 1}.$$

Acknowledgments

I thank G. W. Wasilkowski and A. G. Werschulz for valuable comments.

References

[1] N. S. Bakhvalov, On the optimality of linear methods for operator approximations in convex classes of functions, in Russian, Zh. Vychisl. Mat. Mat. Fiz., 11 (1971) 1014–1018. [English transl.: USSR Comput. Math. Math. Phys. 11 (1971) 244–249].

[2] S. Gal and C. A. Micchelli, Optimal sequential and non-sequential procedures for evaluating a functional, Appl. Anal. 10 (1980) 105–120.

[3] C. A. Micchelli and T. J. Rivlin, A survey of optimal recovery, in *Optimal Estimation in Approximation Theory*, C. A. Micchelli and T. J. Rivlin eds., Plenum Press, New York, (1977) 1–54.

[4] E. Novak, Optimal linear randomization methods for linear operators in Hilbert spaces, J. Complexity 8 (1992) 22–36.

[5] I. H. Sloan and S. Joe, *Lattice methods for multiple integration*, Oxford University Press, Oxford, 1994.

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

[7] J. F. Traub and H. Woźniakowski, *A general theory of optimal algorithms*, Academic Press, New York, 1980.

[8] G. W. Wasilkowski, Information of varying cardinality, J. Complexity 2 (1986) 204–228.

[9] G. W. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, J. Complexity 11 (1995) 1–56.

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

[11] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, J. of Computing and Information 4 (1994) 1–19.