On the tractability of linear tensor product problems in the worst case

Anargyros Papageorgiou and Iasonas Petras
Department of Computer Science
Columbia University
New York, NY 10027

May 14, 2009

Abstract

It has been an open problem to derive a necessary and sufficient condition for a linear tensor product problem to be weakly tractable in the worst case. The complexity of linear tensor product problems in the worst case depends on the eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ of a certain operator. It is known that if $\lambda_1 = 1$ and $\lambda_2 \in (0,1)$ then $\lambda_n = o((\ln n)^{-2})$, as $n \to \infty$, is a necessary condition for a problem to be weakly tractable. We show this is a sufficient condition as well.

1 Introduction

Traditionally, the complexity of multivariate problems has been studied with respect to the accuracy demand ε while considering the number of variables d to be arbitrary but fixed; for example, see [2] and the references therein. The resulting asymptotic estimates tend to ignore components of the cost of the algorithms and thereby of the complexity that are independent of ε but depend on d, even though they may be substantial and perhaps exponentially large in d. The study of the complexity of multivariate problems as a function of the number of variables and the accuracy requires a significant amount of new research.

About fifteen years ago, Henryk Woźniakowski introduced these ideas and initiated research in this area that has produced numerous results. Many results, some of them very recent, are presented in the book *Tractability of Multivariate Problems, Volume 1: Linear Information*, by Erich Novak and Henryk Woźniakowski, which has just been published [1]. However, many questions remain open, thirty of which are stated in this book as open problems. In this paper we solve Open Problem 26.

Linear multivariate problems deal with the approximation of a problem $S = \{S_d\}$, where each of the S_d, $d \geq 1$, is a continuous linear operator defined on a space of functions f of d variables. Moreover, the algorithms that approximate $S_d(f)$ can use n evaluations of arbitrary continuous linear functionals. The information complexity (for brevity, the complexity) is the minimal number of evaluations required to approximate S_d with accuracy ε. Accordingly, the complexity is denoted by $n(\varepsilon, d)$ to emphasize its dependence on ε and d. We remark that there are a variety of error criteria that one may consider for the accuracy of the algorithms but we limit ourselves to the worst case error. We will give all the necessary definitions and details in the next section.

A problem $S = \{S_d\}$ is *weakly tractable* iff

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln n(d,\varepsilon)}{\varepsilon^{-1}+d} = 0,$$

otherwise it is intractable. Thus, a problem is intractable if its complexity is an exponential function of either d or ε^{-1}. Observe that weakly tractable problems may have complexity that grows faster than a polynomial in ε^{-1} and d.

Linear multivariate tensor product problems are the linear multivariate problems obtained by taking the tensor product of d copies of a single univariate linear problem. Thus

$$S_d = \bigotimes_{j=1}^{d} S_1,$$

where S_1 is a given continuous linear operator. In this case, the complexity of approximating S_d with accuracy ε depends on the singular values of S_1 and, particularly, on their rate of decay [1, Ch. 5.2]. The squares of the singular values of S_1 are the eigenvalues, $\{\lambda_i\}_{i\in\mathbb{N}}$, of the operator $S_1^*S_1$, where the eigenvalues are indexed in non increasing order. Moreover, the relationship between the tractability of $S = \{S_d\}$ and the $\{\lambda_i\}_{i\in\mathbb{N}}$ is studied in detail in [1, Thm. 5.5]. In particular, we know that if a problem is weakly tractable with $\lambda_1 = 1$ and $\lambda_2 \in (0,1)$ then $\lambda_n = o((\ln n)^{-2})$, as $n \to \infty$. Proving the converse is Open Problem 26, which we solve in this paper. We remark that [1, Thm. 5.5] shows a stronger condition, namely, that if $\lambda_1 = 1$, $\lambda_2 < 1$ and $\lambda_n = o((\ln n)^{-2}(\ln\ln n)^{-2})$, as $n \to \infty$, then S is weakly tractable.

2 Linear Tensor Product Problems

A linear tensor product problem is defined in [1, Ch. 5.2] as a tensor product of a single univariate linear problem.

Let H_1 be an infinite dimensional separable Hilbert space of real univariate functions with its inner product denoted by $\langle\cdot,\cdot\rangle_{H_1}$, and let G_1 be an arbitrary Hilbert space. Assume that $S_1 : H_1 \to G_1$ is a compact linear operator. The operator

$$W_1 := S_1^* S_1 : H_1 \to H_1$$

is positive semi-definite, self-adjoint and compact. Let us denote its ordered eigenvalues by $\{\lambda_i\}$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \geq \ldots$. They are the squares of the singular values of S_1. We denote the eigenpairs of W_1 by $\{(\lambda_i, e_i)\}_{i\in\mathbb{N}}$.

For $d \geq 1$, define $H_d = \bigotimes_{j=1}^d H_1$ to be the tensor product of the space H_1. This is a space of real functions of d variables. Similarly, let $G_d = \bigotimes_{j=1}^d G_1$. The linear tensor product problem is defined by considering the operator

$$S_d := \bigotimes_{j=1}^{d} S_1 : H_d \to G_d.$$

Observe that S_d is compact and that $\|S_d\|_{H_d} = \prod_{j=1}^d \big[\lambda_1\big]^{1/2}$. The problem $S = \{S_d\}$ is called the *linear tensor product problem*.

The non-negative definite, self adjoint and compact operator

$$W_d := S_d^* S_d : H_d \to H_d$$

has eigenpairs $\{\lambda_{d,i}, e_{d,i}\}_{i\in\mathbb{N}^d}$ with $\lambda_{d,i} = \prod_{j=1}^d \lambda_{i_j}$, and $e_{d,i} = \bigotimes_{j=1}^d e_{i_j}$ for all $i = [i_1, i_2, \ldots, i_d] \in \mathbb{N}^d$. Let λ_{d,β_j} denote the j-th largest of all the $\lambda_{d,i}$ and let e_{d,β_j} denote the corresponding eigenvector. Clearly, $\lambda_{d,\beta_1} = \lambda_{d,1,\ldots,1} = \lambda_1^d$.

Suppose we can use arbitrary linear continuous functionals as information operations. Then it is known, see e.g. [3], that the algorithm

$$A_{n,d}(f) = \sum_{j=1}^{n} \langle f, e_{d,\beta_j} \rangle_{H_d} S_d e_{d,\beta_j}$$

minimizes the worst case error among all possible algorithms using at most n information operations. The worst case error is defined as

$$e(A_{n,d}) = \sup_{f \in H_d,\, \|f\|_{H_d} \leq 1} \|S_d f - A_{n,d}(f)\|_{G_d}.$$

It is also known that $e(A_{n,d}) = \sqrt{\lambda_{d,\beta_{n+1}}}$.

For accuracy ε, the worst case information complexity of the problem S_d for the absolute error criterion is defined as the minimal number of information operations needed to guarantee that the worst case error is at most ε, and is given by

$$n(d,\varepsilon) = |\{ i \in \mathbb{N}^d \ : \ \lambda_{d,i} > \varepsilon^2 \}|,$$

where $|\{\cdot\}|$ denotes the cardinality of the set.

As we have already mentioned, the problem S_d is *weakly tractable* iff

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln\, n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0.$$

The reader is referred to [1] for more details.

3 Weak Tractability

Recall that the relationship between the complexity $n(\varepsilon, d)$ of linear tensor product problems and the singular values of S_1 is extensively studied in [1, Thm. 5.5]. More precisely, the complexity depends on the eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ of the operator W_1. The problem $S = \{S_d\}$ is intractable when $\lambda_1 > 1$ and $\lambda_2 > 0$. When $\lambda_1 = \lambda_2 = 1$ the problem remains intractable.

When $\lambda_1 = 1$ and $\lambda_2 \in (0,1)$ the problem is weakly tractable as long as the remaining eigenvalues decay sufficiently fast. Theorem 5.5 shows that $\lambda_n = o((\ln n)^{-2}(\ln\ln n)^{-2})$, as $n \to \infty$ is a sufficient condition. It also shows that if a problem is weakly tractable then $\lambda_n = o((\ln n)^{-2})$, as $n \to \infty$.

The question in Open Problem 26 in [1] is whether the latter is a necessary and sufficient condition for a problem to be weakly tractable. We give an affirmative answer below.

Theorem 1. *: Consider the linear tensor product problem in the worst case setting $S = \{S_d\}$ with $\lambda_1 = 1$ and $\lambda_2 \in (0, 1)$ with the absolute error criterion. Then S is weakly tractable iff*

$$\lambda_n = o((\ln n)^{-2}) \ \text{ as } \ n \to \infty.$$

Proof. We know that $\lambda_n = o((\ln n)^{-2})$ is a necessary condition for weak tractability of S [1, Thm. 5.5]. We show that it is also a sufficient condition.

When $\lambda_n = o(\ln^{-2} n)$ one may proceed as in [1] to obtain $\ln n(\varepsilon, 1) = o(\varepsilon^{-1})$. Indeed, $n(\varepsilon, 1) = \min\{n : \lambda_{n+1} < \varepsilon^2\} \leq \min\{n : \ln n = o(\varepsilon^{-1})\}$.

When $\lambda_2 \leq \varepsilon^2$ we know that $n(\varepsilon, 1) \leq 1$ and so we consider the case $\lambda_2 > \varepsilon^2$.

For $d \geq 2$, we are interested in eigenvalue products satisfying

$$\lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d} > \varepsilon^2. \tag{1}$$

Let k be the number of indices $j_i \geq 2$, i.e., $\lambda_{j_i} < 1$. The inequality above implies

$$\lambda_2^k > \varepsilon^2, \tag{2}$$

and we know that $k \leq a_d(\varepsilon)$, where

$$a_d(\varepsilon) = \min\left(d, \left\lceil 2\frac{\ln \varepsilon^{-1}}{\ln \lambda_2^{-1}} \right\rceil - 1\right),$$

see [1, Thm. 5.5] for the details.

There are $\binom{d}{a_d(\varepsilon)}$ ways to select the $(d - a_d(\varepsilon))$ indices j_r that must be equal to 1, i.e., $\lambda_{j_r} = 1$, due to (1,2).

Let $j_{\max}$ be the largest index of the eigenvalues in (1), then $\lambda_{j_{\max}} \geq \lambda_{j_1} \cdots \lambda_{j_d} > \varepsilon^2$, which implies $j_{\max} \leq n(\varepsilon, 1)$. Note that there are no more than $a(d) \leq d$ choices for the location of the largest index.

Consider now the second largest index $j'_{\max}$ of the eigenvalues in (1) , then $\lambda^2_{j'_{\max}} \geq \lambda_{j'_{\max}} \lambda_{j_{\max}} \geq \lambda_{j_1} \cdots \lambda_{j_d} > \varepsilon^2$, which implies that $\lambda_{j'_{\max}} > \varepsilon$ and so $j'_{\max} \leq n(\varepsilon^{1/2}, 1)$.

(Similarly, we see that the i-th largest index is at most $n(\varepsilon^{1/i}, 1)$.)

Thus, we estimate $n(\varepsilon, d)$ by

$$n(\varepsilon, d) \leq \binom{d}{a_d(\varepsilon)} [n(\varepsilon^{1/2}, 1)]^{a_d(\varepsilon)-1} n(\varepsilon, 1) d.$$

Taking the logarithm we obtain

$$\begin{aligned}
\ln n(\varepsilon, d) &\leq \ln \left[\binom{d}{a_d(\varepsilon)} [n(\varepsilon^{1/2}, 1)]^{a_d(\varepsilon)-1} n(\varepsilon, 1) d \right] \\
&= \ln \binom{d}{a_d(\varepsilon)} + (a_d(\varepsilon) - 1) \ln n(\varepsilon^{1/2}, 1) + \ln n(\varepsilon, 1) + \ln d \\
&\leq a_d(\varepsilon) \ln d - \ln(a_d(\varepsilon)!) + a_d(\varepsilon) \ln n(\varepsilon^{1/2}, 1) + \ln n(\varepsilon, 1) + \ln d \\
&\leq a_d(\varepsilon) \ln d + a_d(\varepsilon) \ln n(\varepsilon^{1/2}, 1) + \ln n(\varepsilon, 1) + \ln d.
\end{aligned}$$

Dividing by $(\varepsilon^{-1} + d)$ and taking the limit as $\varepsilon^{-1} + d \to \infty$ yields

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} \leq \lim_{\varepsilon^{-1}+d\to\infty} \left[\frac{a_d(\varepsilon) \ln d}{\varepsilon^{-1} + d} + \frac{a_d(\varepsilon) \ln[n(\varepsilon^{1/2}, 1)]}{\varepsilon^{-1} + d} + \frac{\ln n(\varepsilon, 1)}{\varepsilon^{-1} + d} + \frac{\ln d}{\varepsilon^{-1} + d} \right].$$

Using $\ln n(\varepsilon, 1) = o(\varepsilon^{-1})$ and $a_d(\varepsilon) = \Theta(\min(d, \ln \varepsilon^{-1}))$, we consider the limit of each of the four terms in the right hand side above.

The limit of the first term is zero. Indeed, as in [1], if $x = \max(d, \varepsilon^{-1})$, then $\min(d, \ln \varepsilon^{-1}) \leq \ln x$, and

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\min(d, \ln \varepsilon^{-1}) \ln d}{(\varepsilon^{-1} + d)} \leq \lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln^2 x}{x} = 0.$$

The limit of the second term is zero since

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\min(d, \ln \varepsilon^{-1}) \cdot o(\varepsilon^{-1/2})}{(\varepsilon^{-1} + d)} = 0.$$

Observe that if we had $o(\varepsilon^{-1})$ instead of $o(\varepsilon^{-1/2})$ in the numerator, then for $d = \Theta(\varepsilon^{-1})$ the limit would not be zero, which was the complicating factor in [1].

For the third term, it is easy to see that

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln n(\varepsilon, 1)}{\varepsilon^{-1} + d} = \lim_{\varepsilon^{-1}+d\to\infty} \frac{o(\varepsilon^{-1})}{\varepsilon^{-1} + d} = 0.$$

Finally, the limit of the fourth term is trivially zero.

Hence,

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln n(\varepsilon, d)}{(\varepsilon^{-1}+d)} = 0,$$

and the problem is weakly tractable. $\square$

Acknowledgments

We thank Joseph F. Traub and Henryk Woźniakowski for their helpful comments and suggestions.

References

[1] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*, Volume 1, Linear Information, European Math. Soc., Zürich, 2008.

[2] J. F. Traub and A. G. Werschulz, *Complexity and Information*, Cambridge University Press, Cambridge, 1999.

[3] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988.