JOURNAL OF COMPLEXITY **16**, 459–473 (2000)
doi:10.1006/jcom.2000.0544, available online at http://www.idealibrary.com on IDEAL®

On the Communication Complexity of Lipschitzian Optimization for the Coordinated Model of Computation

Mehran Mesbahi[1]

Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, Minnesota

and

George P. Papavassilopoulos[2]

Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, California

Received January 20, 1999

We consider the problem of approximating the maximum of the sum of m Lipschitz continuous functions. The values of each function are assumed to reside at a different memory element. A single processing element is designated to approximate the value of the maximum of the sum of these functions by adopting a certain protocol. Under certain assumptions on the class of permissible protocols, we obtain the minimum number of real-valued messages that has to be transferred between the processing element and the memory elements in order to find the desired approximation of this maximum. In particular, we exploit the optimality of the nonadaptive protocols for the Lipschitzian optimization problem, studied in the context of information-based complexity, to prove our main result. © 2000 Academic Press

Key Words: Lipschitzian optimization; communication complexity; coordinated model of computation.

1. INTRODUCTION

Five people are each given a number between 1 and 100. A questioner comes along and wants to figure out the sum of these five numbers by

[1] To whom correspondence should be addressed. E-mail: mesbahi@aem.umn.edu.
[2] E-mail: yorgos@bode.usc.edu.

0885-064X/00 $35.00

asking each person questions of the type: "Is your number greater than or equal to x?," for some integer x. In return, the questioner expects to receive a "yes" or a "no" response from the participant. The questioner has been told by an informer that the sum of the numbers held by the participants does not exceed 100. Suppose that the questioner asks the first person whether the number held by him or her is greater than or equal to 75 and receives in return a "yes" answer. Knowing that the total sum is not greater than or equal to 100, it would not be wise for the questioner to ask the second person "Is your number greater than or equal to 75?" since the reply is *definitely* a "no." In fact, the questioner can use the previous responses to formulate the question for the next person in some intelligent manner. In a sense, the questioner can *adapt* the next question by incorporating the answers for the previous questions in its formulation.

It is also conceivable that the questioner first figures out the exact number that each person has, without considering other people's numbers, and then sums them up. In this case, the total number of questions that the questioner has to ask would just be five times the number of questions needed to figure out one person's number. In fact one might suspect that the questioner cannot really do better than this, in terms of the minimizing the total number of questions asked, at least in the *worst case*.

In this paper we consider a similar problem. There are m function storage devices, or memory elements, storing the values of the functions $f_1, ..., f_m$, and each function is known to be in the class of Lipschitz continuous functions with modulus k; the class of such functions will be denoted by F_k. There is a processing element which is designated to approximate the value of $z := \max_x \sum_{j=1}^{m} f_j(x)$ by asking each memory element about the value of the function residing in that memory element, at a given point. We are interested to know, under the above restrictions, the minimum number of questions that the processing element has to ask the memory elements, in order to be able to approximate the value of z within an accuracy $0 < \varepsilon < 1$, for all possible $f_j \in F_k (j = 1, ..., m)$. Although in this case, the processing element cannot find z by figuring out the maximum of each function f_j $(j = 1, ..., m)$ separately, our result indicates that in terms of the total number of questions needed to be asked from the memory elements, the processing element cannot do any better than this, at least in the *worst case*. The minimum number of questions needed to approximate the maximum of the sum of functions in F_k as described above, shall be referred to as the communication complexity of the k-Lipschitzian optimization for the coordinated model of computation.

The problem of determining the communication complexity is important in several settings. First, is the area of parallel and distributed computation (Bertsekas and Tsitsiklis, 1989; Hwang, 1994). In this setting, one can consider the processors as having partial information regarding the computational task at a given time, and hence they communicate among themselves in order to solve the problem in a distributed manner. It is believed that the amount of communication needed to complete the computation in the distributed manner is one of the main factors that determine the efficiency of the parallelism employed (Gentleman, 1978; Saad 1986). The communication requirements become very important in the context of very large integrated circuits (VLSI) (Aho *et al.*, 1983; Ullman, 1984). In particular, it is known that the number of bits that is needed to be exchanged between the different parts of the chip is related to the product of the area of the chip and the computation time (Ullman, 1984).

The issue of communication complexity is also of relevance in the setting of distributed data acquisition and control. In this case, one can consider the processing element as the controller which has access to the state of the environment through two or more sensors. The sensors, due to their limited computational power can only send functionals of the state of the environment upon receiving a correspondence from the controller. If the communication among the controller and the sensors is costly (for example due to the congestion of the network), the issue of communication complexity becomes important.

In this paper, we will show that for the Lipschitzian optimization problem, the methodology developed in the context of information-based complexity can be extended to the coordinated model of computation. This will be done mainly by utilizing the results pertaining to the optimality of the nonadaptive protocols for the case where there is a single pair of processing and memory elements.

The organization of the paper is as follows. We first provide a very brief survey of the works that have been done in the area of communication complexity. In Section 2, we provide the minimum amount of notation and preliminaries which enables us to state, more formally, the problem and the main result discussed in the paper. Section 3 is devoted to the proof of the main result.

1.1. *Related Works*

The study of communication complexity was initiated by Abelson (1980) where functions of the form $f\colon \Re^m \times \Re^n \to \Re$, $f \in C^2$ (the class of twice continuously differentiable functions) were studied. In this setting, $x \in \Re^m$ and $y \in \Re^n$ are known by different processors, and each processor can transmit functions of their data which are also assumed to be in C^2. The communication complexity is defined to be the minimum number of messages that has to be exchanged between the processors in order to exactly evaluate $f(x, y)$. It should be noted that functions considered by Abelson (1980) have a very special structure; namely, it is assumed that *there exists* a communication protocol which can be employed by the processors in order to obtain the exact value of $f(x, y)$. The work that has been done in the spirit of Abelson (1980) includes that of Luo and Tsitsiklis (1991).

Another stream of work on the communication complexity was initiated by the work of Yao (1979). This line of work is concerned with obtaining the communication complexity of evaluating a Boolean function $f(x, y)$, where $f: X \times Y \rightarrow \{0, 1\}$, and X and Y are finite sets. It is assumed that $x \in X$ and $y \in Y$ are known by two different processors. The communication complexity is then defined to be the minimum number of bits that has to be exchanged between the processors in order to exactly evaluate $f(x, y)$, for all possible values of $x \in X$ and $y \in Y$. In this setting, X and Y have a simple structure and the communication complexity, in essence, is an indication of the behavior of f on the lattice $X \times Y$. We refer the reader to the survey of Orlitsky and El Gamal (1988) for a summary of this approach and many possible extensions.

More closely related to the approach of the present paper is the work done in the area of information-based complexity (Traub *et al.*, 1988; Nemirovsky and Yudin, 1983), and in particular the work of Sukharev (1992). This line of work is concerned with the efficiency of algorithms for problems defined on the infinite-dimensional spaces, such as the function integration problem, approximation, and optimization. In this context, the processing element can obtain the values of the function (a member of an infinite-dimensional space) through an oracle. In general, for these problems only approximate solutions can be obtained. Therefore there is a presence of the parameter ε in all the complexity results. It turns out that in many situations, the cost of the oracle calls dominates the cost of the entire algorithm. One is thus led to consider the communication complexity, i.e., the minimum number of oracle calls needed by the algorithm in order to be able to approximate the solution (the minimizer, the value of the integral, etc.) within an error ε. The work of Sukharev (1992) is concerned with the same issue, but his approach relies more heavily on the minimax models and various notions of adaptive algorithms. In fact in the introduction of Sukharev (1992) it is stated that the main feature of the work is that "the process of computation has been regarded as a controlled process and the algorithm as a control strategy." Consequently, the optimal algorithms are obtained by employing methods of operations research, game theory, and system analysis.

2. PRELIMINARIES AND THE MAIN RESULT

In this section, we provide certain notions which enable us to state the problem considered in the paper more formally. We then state the main result.

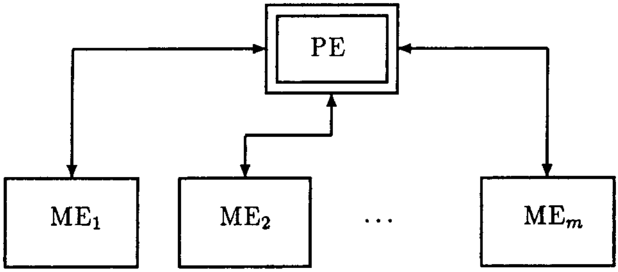

FIG. 1. The coordinated model of computation.

We consider a model of computation, referred to as the coordinated model of computation, where a single processing element (PE) is connected to m memory elements (MEs) via m dedicated channels. This model is shown in Fig. 1.

Each memory element ME_j has access to the value of the Lipschitz continuous function f_j ($j=1, ..., m$), with Lipschitz modulus k, at an arbitrary point of its domain. Let us denote by F_k the class of Lipschitz continuous functions with constant k, defined on the p-dimensional unit cube $[0, 1]^p$. The value of f_j at the point x will be denoted by $x(f_j)$, $j=1, ..., m$. The PE is allowed to specify x to each ME in an arbitrary manner and receive in return the value $x(f_j)$ with infinite precision. The operation of specifying x by the PE and receiving the value $x(f_j)$ from the ME_j is counted as one communication operation. The objective of the PE is to approximate

$$z := \max_{x \in [0,1]^p} \sum_{j=1}^{m} f_j(x) \tag{2.1}$$

within an accuracy $0 < \varepsilon < 1$. Let us denote the total information gathered by the PE after n such *information gathering* operations by I^n; i.e., I^n contains the values of different f_j's at various points of the domain $[0, 1]^p$. The PE then applies a functional $\tilde{\beta}$ to I^n and comes up with an estimate of z (2.1). We shall refer to $\tilde{\beta}$ as the *terminal operation*. The process of gathering the information I^n, and applying the terminal operation $\tilde{\beta}$, will be referred to as the *communication protocol*.

Under the aforementioned restrictions on the communication protocol, let us define the *communication complexity* of the k-Lipschitzian optimization for the coordinated model of computation (with m memory elements) as

$$\Delta_m(\varepsilon, k) := \{\min n: |\tilde{\beta}(I^n) - z| \leqslant \varepsilon, \forall f \in F_k\}. \tag{2.2}$$

The communication complexity $\Delta_m(\varepsilon, k)$ is the minimum number of question–answer sessions that the PE needs to conduct in order to come up with an $0<\varepsilon<1$ approximation of $\max_{[0,1]^p} \sum_{j=1}^m f_j$. The PE comes up with this approximation by applying $\tilde{\beta}$ to the information gathered during the n operations of the form $x(f_j)$. Moreover, the approximation should be valid for all possible $f_j \in F_k$ $(j=1, ..., m)$.

The main result of the paper can now be stated as follows.

THEOREM 1. *For the k-Lipschitzian optimization and for* $0<\varepsilon<1$,

$$\Delta_m(\varepsilon, k) = m\, \Delta_1(\varepsilon, mk). \tag{2.3}$$

In the rest of the paper, we shall present the proof of Theorem 1. First however, we need some more preliminaries.

2.1. *More Preliminaries*

Consider the PE–ME's configuration shown in Fig. 1. We assume that f_j: $[0,1]^p \to \Re$, $f_j \in F_k$, and that ME_j has access to the values of f_j at any point $x \in [0,1]^p$, which will be denoted by $x(f_j)$ $(j=1, ..., m)$. As it is customary in (elementary) functional analysis, we think of x both as a point in $[0,1]^p$ and as a functional on F_k, where

$$F_k := \{f\colon [0,1]^p \to \Re,\ |f(x)-f(y)| \leqslant k\, \|x-y\|,\ \forall x, y \in [0,1]^p\}, \tag{2.4}$$

and $\|\cdot\|$ denotes the infinity (supremum) norm. Each of the m channels shown in Fig. 1, between the PE and the MEs, can carry a real number, $x(f_j)$, in response to the x submitted by the PE to the ME_j. The point submitted to the ME_j at time i will be denoted by $x_i(f_j)$.

Let $I^n = (x_1, ..., x_n, x_1(f_{k_1}), ..., x_n(f_{k_n}))$, where $\{k_1, ..., k_n\} \subseteq \{1, ..., m\}^n$. The PE can stop the information gathering process at time n and apply the terminal operation $\tilde{\beta}$ to I^n, in order to approximate z (2.1). The basic question considered in the present work is the minimal value of n, as a function of $0<\varepsilon<1$, such that for all possible choices of $f_j \in F_k$ $(j=1, ..., m)$, $|\tilde{\beta}(I^n) - z| \leqslant \varepsilon$.

In order for the PE to come up with the point x_i to be submitted to some ME at time i, it employs the previously gathered information by using some *strategy*. Let us denote by $\tilde{x}_i$ the strategy employed by the PE at time i to come up with the new point x_i that is to be submitted to some ME.

The $(n+1)$ tuple $\tilde{\alpha}^n := (\tilde{x}_1, ..., \tilde{x}_n, \tilde{\beta})$ will be called an nth degree (deterministic) protocol for approximating z (2.1). The protocol $\tilde{\alpha}^n$ can in fact be described by the following sequence of mappings and functionals,

$$\tilde{x}_1 \equiv x_1: F_k \to \Re; \qquad x_1 \in [0,1]^p,$$

$$\tilde{x}_2: [0,1]^p \times \Re \to [0,1]^p \quad \text{(which yields } x_2\text{)}; \qquad x_2: F_k \to \Re,$$

$$\vdots$$

$$\tilde{x}_n: \overbrace{[0,1]^p \times \cdots \times [0,1]^p}^{n-1} \times \overbrace{\Re \times \cdots \times \Re}^{n-1} \to [0,1]^p \quad \text{(which yields } x_n\text{)};$$

$$x_n: F_k \to \Re,$$

$$\tilde{\beta}: \overbrace{[0,1]^p \times \cdots \times [0,1]^p}^{n} \times \overbrace{\Re \times \cdots \times \Re}^{n} \to \Re.$$

Note that in general, the PE uses all the previous points chosen, and all the corresponding function values at those points, to come up with the new point.

If the new point to be submitted to the ME_j is *independent* of the previous questions, for all $j = 1, ..., m$ and for all time instances $i \geqslant 1$, then we call the protocol (strongly) nonadaptive. In this case, $\tilde{x}_i \equiv x_i: F_k \to \Re$, for all $i \geqslant 1$.

For the nonadaptive protocols, since the point x_i submitted to the ME_j solely depends on F_k, it follows that the same point should also be chosen for all j $(j = 1, ..., m)$ and that the number of points specified for all MEs should be equal. Without loss of generality, for the nonadaptive case, we shall assume that the points are submitted to the MEs by the PE in the round-robin manner; i.e., the MEs are indexed from 1 through m, and the points are submitted to the MEs starting from ME_1 through ME_m, and then back to ME_1, and so on. For the nonadaptive protocols, let l denote the number of points specified to each one of the MEs; $n = ml$. The judicious choice of the terminal operation in the *nonadaptive* case would be

$$\tilde{\beta}(I^n) = \max_{k = 0, ..., l-1} \sum_{j=1}^{m} x_{km+j}(f_j). \tag{2.5}$$

For our purpose, the role played by the nonadaptive protocols is of central importance; the results pertaining to the nonadaptive protocols are easily extendible from $m = 1$ to the case where $m > 1$, as will be shown in the next section.

Let $\tilde{A}^n$ and A^n denote the class of nth degree adaptive and nonadaptive protocols, respectively. Then for a specific set of $f_1, ..., f_m \in F_k$, residing at ME_j $(j = 1, ..., m)$, the error associated with using the protocol $\tilde{\alpha}^n$ to approximate z (2.1) is defined to be

$$e(\tilde{\alpha}^n, k, f_1, ..., f_m) := |z - \tilde{\beta}(I^n)|. \tag{2.6}$$

The least guaranteed error for the class of nth degree adaptive protocols for the k-Lipschitzian optimization will then be

$$E_m^{\text{adaptive}}(n, k) := \inf_{\tilde{\alpha}^n \in \tilde{A}^n} \sup_{f_1, ..., f_m \in F_k} e(\tilde{\alpha}^n, k, f_1, ..., f_m) \tag{2.7}$$

and for the nonadaptive case it will be

$$E_m^{\text{non-adaptive}}(n, k) := \inf_{\tilde{\alpha}^n \in A^n} \sup_{f_1, ..., f_m \in F_k} e(\tilde{\alpha}^n, k, f_1, ..., f_m). \tag{2.8}$$

Although $E_m^{\text{adaptive}}(n, k) \leqslant E_m^{\text{nonadaptive}}(n, k)$ by definition (since $A^n \subseteq \tilde{A}^n$), we shall nevertheless use the following notation:

$$E_m(n, k) := \min\{E_m^{\text{nonadaptive}}(n, k), E_m^{\text{adaptive}}(n, k)\}. \tag{2.9}$$

Clearly the only interesting scenario would be when

$$E_m(n, k) = E_m^{\text{nonadaptive}}(n, k),$$

which is the case if and only if $E_m^{\text{nonadaptive}}(n, k) = E_m^{\text{adaptive}}(n, k)$.

We now define similar definitions for the communication complexity of the k-Lipschitzian optimization. In particular, let us define the following:

$$\Delta_m^{\text{adaptive}}(\varepsilon, k) := \min\{n\colon E_m^{\text{adaptive}}(n, k) \leqslant \varepsilon\}, \tag{2.10}$$

$$\Delta_m^{\text{nonadaptive}}(\varepsilon, k) := \min\{n\colon E_m^{\text{nonadaptive}}(n, k) \leqslant \varepsilon\}, \tag{2.11}$$

and,

$$\Delta_m(\varepsilon, k) := \min\{n\colon E_m(n, k) \leqslant \varepsilon\}. \tag{2.12}$$

Our approach for proving Theorem 1 is along the proofs of the following two propositions.

PROPOSITION 2.

$$\Delta_m^{\text{nonadaptive}}(\varepsilon, k) = m\Delta_1(\varepsilon, mk). \tag{2.13}$$

PROPOSITION 3.

$$\Delta_m^{\text{nonadaptive}}(\varepsilon, k) = \Delta_m(\varepsilon, k). \tag{2.14}$$

Theorem 1 will then follow by combining the statements of Propositions 2 and 3.

3. THE PROOF OF THE MAIN RESULT

In this section, we present the proofs of Propositions 2 and 3. The statement of Theorem 1 then follows immediately from the results of these two propositions.

Proposition 2 deals with the problem of extending the result pertaining to the communication complexity for $m=1$ to the case of $m>1$. When $m=1$, the coordinated model of computation shown in Fig. 1 reduces to the oracle-type machine considered in the context of information-based complexity.

The following result of Sukharev (1992) plays a central role in our analysis. We remind the reader that we are dealing only with the Lipschitzian optimization problem.

THEOREM 4 (Sukharev (1992, Theorem 1.2, p. 125)). *For all* $k>0$,

$$E_1^{\text{nonadaptive}}(n,k)=E_1(n,k)=\frac{k}{2\lfloor n^{1/p}\rfloor}.$$

An obvious implication of Theorem 4 is that for all $m>1$, and for fixed $k>0$,

$$E_1^{\text{nonadaptive}}(n,mk)=E_1(n,mk)=mE_1(n,k).$$

As a corollary to Theorem 4 we also obtain

COROLLARY 5. *For all* $k>0$,

$$\Delta_1^{\text{nonadaptive}}(\varepsilon,k)=\Delta_1(\varepsilon,k)=\left\lceil\left(\frac{k}{2\varepsilon}\right)^p\right\rceil.$$

We now present the proof of Proposition 2.

Proof (Proposition 2). As it was pointed out in Introduction, for the nonadaptive protocols we shall fix the operation $\tilde{\beta}$, as defined in (2.5). In this case, the points submitted to each ME_j ($j=1, ..., m$) by the PE depend solely on the functional class F_k.

If we use the optimal nonadaptive protocol for the single ME case on all the m MEs and use the terminal operation (2.5), we obtain a nonadaptive protocol for the m ME case and thus,

$$\begin{aligned} E_m^{\text{nonadaptive}}(n, k) &= E_m^{\text{nonadaptive}}(ml, k) \\ &\leqslant E_1^{\text{nonadaptive}}(l, mk) \\ &= mE_1^{\text{nonadaptive}}(l, k), \end{aligned}$$

since this protocol is now exactly an lth degree protocol for the single ME case as applied to a function in F_{mk}. Consequently,

$$\Delta_m^{\text{nonadaptive}}(\varepsilon, k) \leqslant m\Delta_1^{\text{nonadaptive}}\left(\frac{\varepsilon}{m}, k\right) = m\,\Delta_1^{\text{nonadaptive}}(\varepsilon, mk).$$

Consider now a situation where the functions residing in the m MEs are all identical and, moreover, where the PE *is aware* of this fact when submitting the questions to the MEs. The best achievable error bound in this case would be $E_1^{\text{nonadaptive}}(l, mk)$, indicating that this scenario is exactly the same as the one where the PE is communicating with one ME, with the knowledge that the function residing in that ME is a member of $F_{m,k}$. Thereby,

$$\begin{aligned} \inf_{\tilde{\alpha}^n \in A^n} \sup_{f_1 = f_2 = \cdots = f_m \in F_k} e(\tilde{\alpha}^n, k, f_1, ..., f_m) &= \inf_{\tilde{\alpha}^l \in A^l} \sup_{f \in F_{mk}} e(\tilde{\alpha}^l, mk, f) \\ &= E_1^{\text{nonadaptive}}(l, mk) \\ &\leqslant E_m^{\text{nonadaptive}}(n, k), \end{aligned}$$

which translates to

$$mE_1^{\text{nonadaptive}}(l, k) \leqslant E_m^{\text{nonadaptive}}(n, k).$$

Thus,

$$m\,\Delta_1^{\text{nonadaptive}}\left(\frac{\varepsilon}{m}, k\right) = m\,\Delta_1^{\text{nonadaptive}}(\varepsilon, mk) \leqslant \Delta_m^{\text{nonadaptive}}(\varepsilon, k). \quad \blacksquare$$

Using Corollary 5, we also conclude that

$$\Delta_m^{\text{nonadaptive}}(\varepsilon, k) = m\,\Delta_1(\varepsilon, mk).$$

We now present the proof of Proposition 3. Proposition 3 states that for the k-Lipschitzian optimization on the coordinated model of computation, the PE cannot do any better than using an optimal nonadaptive protocol (in terms of reducing the amount of communication needed in the worst case).

Proof (Proposition 3). It suffices to show that for all $n \geqslant 1$,

$$E_m^{\text{nonadaptive}}(n, k) = E_m^{\text{adaptive}}(n, k).$$

The proof is essentially a straightforward generalization of the proof of Sukharev for the single memory case. We present the proof in three steps. First, some notations are introduced.

Consider a nonadaptive protocol with the fixed terminal operation defined as in (2.5). Then the protocol is merely specified by the points x_i submitted to each PE. In particular, define $x^n := (x_1, ..., x_n)$ and write

$$E_m^{\text{nonadaptive}}(n, k) = \inf_{x^n} \sup_{f_1, ..., f_m \in F_k} e(x^n, k, f_1, ..., f_m),$$

where $e(x^n, k, f_1, ..., f_m)$ is the error of the approximation when the terminal operation is fixed (similarly we shall use the notation $e(\tilde{x}^n, k, f_1, ..., f_m)$ for the adaptive case).

Define the set

$$\mathscr{F}(x^n, f_1, ..., f_n) := \{(f'_1, ..., f'_m) \mid x_i(f'_j) = x_i(f_j);\ i = j \bmod m\}$$

and

$$e_{\mathscr{F}}(x^n, f_1, ..., f_m) = \sum_{(f'_1, ..., f'_m) \in \mathscr{F}(x^n, f_1, ..., f_n)} e(x^n, k, f'_1, ..., f'_m).$$

The three steps of the proof are as follows:

1. There exist functions $\bar{f}_j$ $(j \leqslant j \leqslant m)$ such that $e_{\mathscr{F}}(x^n, \bar{f}_1, ..., \bar{f}_m) \geqslant e_{\mathscr{F}}(x^n, f_1, ..., f_m)$, for all $f_j \in F_k$.

2. Provided that (1) holds, $e_{\mathscr{F}}(x^n, f_1, ..., f_m)$ has a generalized saddle point, that is,

$$\inf_{x^n} \sup_{f_1, ..., f_m} e_{\mathscr{F}}(x^n, f_1, ..., f_m) = \sup_{f_1, ..., f_m} \inf_{x^n} e_{\mathscr{F}}(x^n, f_1, ..., f_m).$$

3. Provided that (2) holds, the statement of the Proposition is then proved.

We provide the proof for each part.

1. Fix x^n and let $l = n/m$. For $f'_j \in F_k$ $(1 \leqslant j \leqslant m)$ define,

$$\tilde{f}_j := \frac{1}{m}\left(\sum_j f'_j - \max_{k=0, ..., l-1} \sum_{j=1}^{m} x_{km+j}(f_j)\right)_+,$$

where $f_+ = \max(f, 0)$. Note that for each $j, \tilde{f}_j \in F_k$, since if $g \in F_{mk}$, then $\max(g, 0) \in F_{mk}$, and $\frac{1}{m} g \in F_k$. Now,

$$x_i(\tilde{f}_j) = 0, \qquad 1 \leqslant i \leqslant n, \qquad i = j \bmod m.$$

Thereby,

$$e(x^n, k, f'_1, ..., f'_m) = e(x^n, k, \tilde{f}_1, ..., \tilde{f}_m) \leqslant e_{\mathscr{F}}(x^n, 0, ..., 0),$$

and consequently,

$$e_{\mathscr{F}}(x^n, f'_1, ..., f'_m) \leqslant e_{\mathscr{F}}(x^n, 0, ..., 0)$$

for all $f'_j \in F_k$ $(1 \leqslant j \leqslant m)$.

2. In general one has,

$$\sup_{f_1, ..., f_m} \inf_{x^n} e_{\mathscr{F}}(x^n, f_1, ..., f_m) \leqslant \inf_{x^n} \sup_{f_1, ..., f_m} e_{\mathscr{F}}(x^n, f_1, ..., f_m).$$

To show that the inequality also holds in the reverse direction, we observe that,

$$\begin{aligned}\sup_{f_1, ..., f_m} \inf_{x^n} e_{\mathscr{F}}(x^n, f_1, ..., f_m) &\geqslant \inf_{x^n} e_{\mathscr{F}}(x^n, \tilde{f}_1, ..., \tilde{f}_m)\\ &\geqslant \inf_{x^n} \sup_{f_1, ..., f_m} e_{\mathscr{F}}(x^n, f_1, ..., f_m).\end{aligned}$$

3. We now show that in view of (1) and (2) above,

$$E_m^{\text{nonadaptive}}(n, k) \leqslant E_m^{\text{adaptive}}(n, k).$$

By the definition of $e_{\mathscr{F}}(x^n, f_1, ..., f_m)$, one has

$$E_m^{\text{nonadaptive}}(n, k) = \inf_{x^n} \sup_{f_1, ..., f_m} e_{\mathscr{F}}(x^n, f_1, ..., f_m).$$

For any $\delta > 0$, there exists f_j^δ $(1 \leqslant j \leqslant m)$ such that

$$\begin{aligned}\inf_{x^n} e_{\mathscr{F}}(x^n, f_1^\delta, ..., f_m^\delta) &= \inf_{x^n} \sup_{f_1, ..., f_m} e_{\mathscr{F}}(x^n, f_1, ..., f_m) - \delta\\ &\geqslant \inf_{x^n} \sup_{f_1, ..., f_m} e(x^n, k, f_1, ..., f_m) - \delta,\end{aligned}$$

which implies that for all x^n,

$$e_{\mathscr{F}}(x^n, f_1^\delta, ..., f_m^\delta) \geqslant \inf_{x^n} \sup_{f_1, ..., f_m} e(x^n, k, f_1, ..., f_m) - \delta$$

For any fixed adaptive protocol $\tilde{\alpha}^n = (\tilde{x}^n, \tilde{\beta})$, let x_δ^n be the realization of the strategy $\tilde{x}^n$ for the particular functions f_j^δ $(1 \leqslant j \leqslant n)$. Then,

$$\sup_{f_1, ..., f_m} e'(\tilde{x}^n, k, f_1, ..., f_m) \geqslant e_{\mathscr{F}}(x_\delta^n, f_1^\delta, ..., f_m^\delta)$$
$$\geqslant \inf_{x^n} \sup_{f_1, ..., f_m} e_{\mathscr{F}}(x^n, f_1, ..., f_m) - \delta.$$

Since $\tilde{\alpha}^n = (\tilde{x}^n, \tilde{\beta})$ and $\delta > 0$ were arbitrary,

$$\inf_{\tilde{\alpha}^n} \sup_{f_1, ..., f_m} e(\tilde{\alpha}^n, k, f_1, ..., f_m) \geqslant \inf_{x^n} \sup_{f_1, ..., f_m} e(x^n, k, f_1, ..., f_m).$$

Hence,

$$E_m^{\text{nonadaptive}}(n, k) \leqslant E_m^{\text{adaptive}}(n, k). \quad \blacksquare$$

Having proved Propositions 2 and 3, we now compare the equations (2.13) and (2.14) and obtain,

$$\Delta_m(\varepsilon, k) = m\, \Delta_1(\varepsilon, mk),$$

which is the statement of Theorem 1.

As a corollary of Theorem 1, we also obtain an expression for the communication complexity of the k-Lipschitzian optimization for the coordinated model of computation.

COROLLARY 6. *For* $0 < \varepsilon < 1$,

$$\Delta_m(\varepsilon, k) = m \left\lceil \left(\frac{mk}{2\varepsilon} \right)^p \right\rceil. \tag{3.15}$$

Proof. Since $E_1(n, k) = k/2 \lfloor n^{1/p} \rfloor$ by Theorem 4, $\Delta_1(\varepsilon, k) = \lceil (k/2\varepsilon)^p \rceil$. The statement of the corollary now follows using Theorem 1. $\blacksquare$

4. CONCLUSION

We have addressed the problem of determining the communication complexity of Lipschitzian optimization for the coordinated model of computation. The main concept that has been exploited in this direction is the optimality of a *nonadaptive* protocol among the class of all permissible protocols. This result can be viewed as a generalization of the result of Sukharev for the oracle-type machines considered traditionally in the context of information-based complexity.

There are several directions along which this work can be continued. For example, it would be of interest to consider the communication complexity for more general distributed configurations, such as the case where more than one processing element is present.

ACKNOWLEDGMENTS

The research of G. P. Papavassilopoulos has been supported in part by the National Science Foundation Grant CCR-9222734. The authors thank the referee for the correction to the original proof of Proposition 2.

REFERENCES

1. H. Abelson, Lower bounds on information transfer in distributed computation, *J. Assoc. Comput. Mach.* **27** (1980), 384–392.
2. A. V. Aho, J. D. Ullman, and M. Yannakakis, On notion of information transfer in VLSI circuits, *in* “Proceedings of 15th STOC,” pp. 133–139, 1983.
3. D. P. Bertsekas and J. N. Tsitsiklis, “Parallel and Distributed Computation,” Prentice–Hall, New York, 1989.
4. W. M. Gentleman, Some complexity results for matrix computations on parallel machines, *J. Assoc. Comput. Mach.* **25** (1978), 112–115.
5. K. Hwang, “Advanced Computer Architecture,” McGraw–Hill, New York, 1994.
6. Z. Luo and J. N. Tsitsiklis, On the communication complexity of solving a polynomial equation, *SIAM J. Comput.* **20** (1991), 936–950.
7. A. S. Nemirovsky and D. B. Yudin, “Problem Complexity and Method Efficiency in Optimization,” Wiley, New York, 1983.
8. A. Orlitsky and A. El Gamal, Communication complexity, *in* “Complexity in Information Theory” (Y. S. Abu-Mostafa, Ed.), pp. 20–61, Springer-Verlag, Berlin/New York, 1988.
9. Y. Saad, Communication complexity of Gaussian elimination algorithm on multiprocessors, *Linear Algebra Appl.* **77** (1986), 315–340.
10. A. G. Sukharev, “Minimax Models in the Theory of Numerical Methods,” Kluwer Press, Dordrecht/Norwell, MA, 1992.

11. J. F. Traub, G. W. Wasilkowski, and H. Wazniakowski, "Information-Based Complexity," Academic Press, San Diego, 1988.

12. J. N. Tsitsiklis and Z. Q. Luo, Communication complexity of convex optimization, *J. Complexity* **3** (1987), 231–243.

13. J. Ullman, "Computational Aspects of VLSI," Computer Science Press, Rockville, MD, 1984.

14. A. C. Yao, Some complexity questions related to distributed computing, *in* "Proceedings of 11th STOC," pp. 209–213, 1979.