SIAM J. COMPUT.
Vol. 20, No. 5, pp. 936-950, October 1991

008

ON THE COMMUNICATION COMPLEXITY OF SOLVING A POLYNOMIAL EQUATION*

ZHI-QUAN LUO†‡ AND JOHN N. TSITSIKLIS†

Abstract. This paper considers the problem of evaluating a function $f(x, y)$ $(x \in \Re^m, y \in \Re^n)$ using two processors P_1 and P_2, assuming that processor P_1 (respectively, P_2) has access to input x (respectively, y) and the functional form of f. A new general lower bound is established on the communication complexity (i.e., the minimum number of real-valued messages that have to be exchanged). The result is then applied to the case where $f(x, y)$ is defined as a root z of a polynomial equation $\sum_{i=0}^{n-1}(x_i + y_i)z^i = 0$ and a lower bound of n is obtained. This is in contrast to the $\Omega(1)$ lower bound obtained by applying earlier results of Abelson.

Key words. communication complexity, polynomial equation, lower bound

AMS(MOS) subject classifications. 05, 68

1. Introduction. In a computer network where a set of processors wishes to perform some computational task, communication can sometimes become a bottleneck, especially when communication resources are scarce. This is particularly so in the area of parallel and VLSI computation (see, e.g., [BT89], [U84]), where the communication issues have been studied extensively. In such contexts, it is desirable to design algorithms that require as little information exchange as possible. Problems of minimizing the amount of exchanged information also arise in the context of decentralized signal processing, where each local processor collects some partial data to be processed collectively. In this paper, we study the "communication complexity" (i.e., the minimum possible amount of information exchange) of some particular computational tasks.

Generally speaking, communication complexity depends both on the topology of a computer network and on the nature of the computational task under consideration. In this paper, we ignore the topological issues by assuming that there are only two processors, say P_1 and P_2. We use the following model of communications introduced by Abelson [A80]. Let there be given a continuously differentiable function $f: D_x \times D_y \mapsto \Re$, where D_x and D_y are some open subsets of $\Re^m$ and $\Re^n$, respectively. It is assumed that processor P_1 (respectively, P_2) has access to a vector $x \in D_x$ (respectively, $y \in D_y$) and the formula defining f. The processors P_1, P_2 proceed to evaluate $f(x, y)$ by exchanging messages, using a *two-way communication protocol*, in which messages can be sent in both directions. Let us use π to denote a two-way communication protocol and $r(\pi)$ to denote the number of messages exchanged in π. In addition, let $T_{1\to 2}$ (respectively, $T_{2\to 1}$) denote the set of indices i for which the ith message is sent from P_1 to P_2 (respectively, from P_2 to P_1). The protocol π consists of $r(\pi)$ functions $m_1, \cdots, m_{r(\pi)}: D_x \times D_y \mapsto \Re$, with $m_i(x, y)$ being interpreted as the value of the ith message. These message functions must depend on the inputs x and y in a very special way. Precisely, for each i, there must exist some real-valued function $\hat{m}_i$ such that

$$(1.1)\quad m_i(x, y) = \hat{m}_i(x, m_1(x, y), \cdots, m_{i-1}(x, y)) \quad \forall (x, y) \in D_x \times D_y \quad \text{if } i \in T_{1\to 2},$$

* Received by the editors July 17, 1989; accepted for publication (in revised form) December 5, 1990. This research was supported by Office of Naval Research contract N00014-84-K-0519 (NR649-003) and Army Research Office contract DAAL03-86-K-0171.

† Operations Research Center and Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

‡ Present address, Department of Electrical and Computer Engineering, Room 225/CRL, McMaster University, Hamilton, Ontario, L8S 4L7, Canada.

or

$$(1.2)\quad m_i(x, y) = \hat{m}_i(y, m_1(x, y), \cdots, m_{i-1}(x, y)) \quad \forall (x, y) \in D_x \times D_y \quad \text{if } i \in T_{2\to 1}.$$

Furthermore, we require that either:

(a) There exists a function h such that

$$(1.3)\qquad f(x, y) = h(x, m_1(x, y), \cdots, m_{r(\pi)}(x, y)) \quad \forall (x, y) \in D_x \times D_y,$$

(this corresponds to the case where processor P_1 performs the final computation) or

(b) There exists a function h such that

$$(1.4)\qquad f(x, y) = h(y, m_1(x, y), \cdots, m_{r(\pi)}(x, y)) \quad \forall (x, y) \in D_x \times D_y,$$

which corresponds to the case where processor P_2 computes the final result.

Typically, some smoothness constraints are imposed on the functions m_i, $\hat{m}_i$, and h. For example, [A80] considers the class of two-way communication protocols (denoted by $\Pi_2(f; D_x \times D_y)$) in which the functions m_i, $\hat{m}_i$, and h are twice continuously differentiable. In this paper, we consider a more general class of protocols in which the message functions m_i, $\hat{m}_i$ are once continuously differentiable and the final evaluation function h is continuous. We denote this class of two-way protocols for computing f by $\Pi_1(f; D_x \times D_y)$. We define the two-way communication complexity of computing f with protocols in $\Pi_2(f; D_x \times D_y)$ as

$$C_2(f; D_x \times D_y) = \inf_{\pi \in \Pi_2(f; D_x \times D_y)} r(\pi).$$

We define the quantity $C_1(f; D_x \times D_y)$ similarly. Notice that $\Pi_2(f; D_x \times D_y) \subset \Pi_1(f; D_x \times D_y)$. Thus, $C_2(f; D_x \times D_y) \geq C_1(f; D_x \times D_y)$. As discussed in [L89], $\Pi_1(f; D_x \times D_y)$ is, in some sense, the most general class of protocols for which the notion of communication complexity is well defined for problems involving continuous variables.

A general lower on $C_2(f; D_x \times D_y)$ was established in the fundamental work of Abelson [A80]. In particular, let $f: D_x \times D_y \mapsto \mathfrak{R}$ be a twice continuously differentiable function and let $H_{xy}(f)$ denote the matrix (of size $m \times n$) whose (i, j)th entry is given by $\partial^2 f / \partial x_i \partial y_j$. The following result was proved in [A80].

THEOREM 1.1. *For any $p \in D_x \times D_y$, we have*

$$C_2(f; D_x \times D_y) \geq \operatorname{rank}(H_{xy}(f))(p).$$

Note that Theorem 1.1 only takes into account the second-order derivatives of f and ignores the derivatives of other orders. Thus, this bound should not be expected to be tight, as was shown in [LT89].

In this paper, we derive a new general lower bound. Our result (Theorem 2.1) makes use of the first-order derivatives of f and is fairly intuitive, but surprisingly difficult to prove. Our work was motivated from the problem of distributed computation of a root of a polynomial equation of degree $n-1$. We apply our result to this problem and obtain a lower bound of n, in contrast to the $\Omega(1)$ lower bound obtained from Abelson's result. In [L89], a similar $\Omega(n)$ lower bound is established for the same problem, but under a more restricted class of communication protocols in which the functions m_i, $\hat{m}_i$ $(i = 1, \cdots, r(\pi))$ are assumed to be polynomials. The proof in [L89] makes use of a result from dimension theory and is algebraic in nature, in contrast to the analytic approach in the proof given here.

In related work ([LT89]), Abelson's result has been extended by considering a more restricted class of communication protocols; in particular, some improved lower

bounds on one-way and two-way communication complexity have been obtained by exploiting the algebraic structure present in certain problems. Communication complexity has also been studied under discrete communication models (see, e.g., [MS82], [PS82], [PT82], [Y79]). In these models, the messages are no longer real numbers, but binary strings. A substantial amount of research has been devoted to the study of the communication complexity of selected combinatorial problems ([AU83], [PE86], [U84]). A different model is introduced in [TL87] for the problem of approximately minimizing the sum of two convex functions under the assumption that each convex function is known to a different processor.

The rest of this paper is organized as follows. In § 2, we prove our main result (Theorem 2.1). In § 3, we apply the result of § 2 to establish a lower bound of n for the problem of computing a root of a polynomial equation of degree $n-1$. In § 4, we compare our result with Abelson's. Finally, the Appendix contains certain results from multidimensional calculus that are needed in § 2.

2. Main result. Let $f: D_x \times D_y \mapsto \Re$ be a continuously differentiable function, where D_x and D_y are some open subsets of $\Re^m$ and $\Re^n$, respectively. We use the notation $\nabla_x f(x, y)$ (respectively, $\nabla_y f(x, y)$) to denote the m-dimensional (respectively, n-dimensional) vector whose components are the partial derivatives of f with respect to the components of x (respectively, y). Also, for any set $S \subset D_x$, we use $[\nabla_y f(x, y); x \in S]$ to denote the subspace of $\Re^n$ spanned by the vectors $\nabla_y f(x, y)$, $x \in S$. Finally, for any set $S \subset D_y$, $[\nabla_x f(x, y); y \in S]$ is similarly defined.

ASSUMPTION 2.1. *For any $y \in D_y$, we let*

$$\mathcal{S}^{(2)}(y) = \{S \subset D_x \mid f(S, y) \text{ contains an open interval}\}.^1$$

(For any $x \in D_x$, $\mathcal{S}^{(1)}(x)$ is similarly defined.)

(a) *For any $y \in D_y$ and any nonempty open set $S \subset D_x$, we have $S \in \mathcal{S}^{(2)}$.*

(b) *For any $x \in D_x$ and any nonempty open set $S \subset D_y$, we have $S \in \mathcal{S}^{(1)}$.*

(c) *For some nonnegative integer n_f, we have*

$$\dim [\nabla_y f(x, y); x \in S] \geqq n_f \quad \forall y \in D_y \quad \forall S \in \mathcal{S}^{(2)}(y). \tag{2.1}$$

(d) *For some nonnegative integer m_f, we have*

$$\dim [\nabla_x f(x, y); y \in S] \geqq m_f \quad \forall x \in D_x \quad \forall S \in \mathcal{S}^{(1)}(x). \tag{2.2}$$

Our main result is the following.

THEOREM 2.1. *Under Assumption 2.1, the following is true*

$$C_1(f; D_x \times D_y) \geqq \min \{n_f, m_f\}. \tag{2.3}$$

The proof of Theorem 2.1 is a long and tedious argument based primarily on elementary differential geometry. Before proving Theorem 2.1, we first give a sketch of the basic proof ideas.

Consider an optimal protocol described by (1.1)–(1.2). By symmetry, we can assume that the final evaluation of $f(x, y)$ is performed by processor P_1, in which case the last message must have been transmitted by processor P_2.

We assume, in order to derive a contradiction, that the number r of messages in the protocol satisfies $r < n_f$. Let us fix a "crossing message sequence" $c = (c_1, \cdots, c_r)$,

[1] The notation $f(S, y)$ stands for the set $\{f(x, y) \mid x \in S\}$. Similar notation will be used later without further comment.

that is, the values $c_i = m_i(x, y)$ of the messages under some execution of the protocol. Fixing c imposes the following constraints on x and y:

$$c_i = \hat{m}_i(x, c_1, \cdots, c_{i-1}), \qquad i \in T_{1\to 2}, \tag{2.4}$$

$$c_i = \hat{m}_i(y, c_1, \cdots, c_{i-1}), \qquad i \in T_{2\to 1}. \tag{2.5}$$

Note that these constraints decouple and can be expressed in the form $x \in S_x(c)$ and $y \in S_y(c)$. With some technical work (making sure that certain Jacobians are nonsingular), we can show that $S_x(c)$, $S_y(c)$ are "smooth" (continuously differentiable) surfaces, depending smoothly on c.

The equation

$$f(x, y) = h(x, m_1(x, y), \cdots, m_r(x, y)) \tag{2.6}$$

shows that $f(x, y)$ depends on y through at most r functions. Taking derivatives and using the chain rule, we can show that for any y^*, and any crossing sequence c, the collection of vectors $\{\nabla_y f(x, y^*) \mid x \in S_x(c)\}$ spans a subspace of dimension at most r.

Note that if $y^* \in S_y(c)$, then $f(x, y^*) = h(x, c)$ for all $x \in S_x(c)$. We consider two cases.

Case 1. If there exists some open set of c's in which $h(x, c) = h(c)$ (i.e., independent of x), for all $x \in S_x(c)$, then there exists an open ball in which the equation $f(x, y) = h(m_1(x, y), \cdots, m_r(x, y))$ holds. But this would imply that $f(x, y)$ could have been evaluated by processor P_2 before transmitting the message $m_r(x, \hat{y})$, and we would have a protocol with $r-1$ messages, a contradiction.

Case 2. If Case 1 does not hold, a technical argument shows that there exists some particular c for which $h(x, c)$ is not independent of x. By continuity, $\{h(x, c) \mid x \in S_x(c)\}$ contains an open interval. Hence, $S_x(c)$ belongs to $\mathcal{S}^{(2)}(y^*)$. Therefore, using Assumption 2.1 (d) and the fact that the subspace spanned by the vectors $\{\nabla_y f(x, y^*) \mid x \in S_x(c)\}$ has dimension at most r, we have $n_f \leqq r$, which contradicts our earlier assumption.

To turn the above intuitive argument into a rigorous proof, we have to make sure that all the functions involved are properly defined and have the desired differentiability properties. The rest of this section is devoted to a formal proof of Theorem 2.1.

Let $r = C_1(f; D_x \times D_y)$. We first prove that it is sufficient to show the lower bound (2.3) under the additional assumption

$$r = \min_{\bar{D}_x, \bar{D}_y} C_1(f; \bar{D}_x \times \bar{D}_y), \tag{2.7}$$

where the minimum is taken over all nonempty open subsets $\bar{D}_x, \bar{D}_y$ of D_x, D_y, respectively. Suppose that we have already shown that Theorem 2.1 is true under the assumption (2.7). Let us now show that (2.3) is valid when (2.7) does not hold. In this case, there exists some $r' < r$ and some open subsets $\hat{D}_x \times \hat{D}_y$ of $D_x \times D_y$ such that

$$r' = C_1(f; \hat{D}_x \times \hat{D}_y) = \min_{\bar{D}_x, \bar{D}_y} C_1(f; \hat{D}_x \times \hat{D}_y),$$

where the minimum is taken over all nonempty open subsets $\bar{D}_x, \bar{D}_y$ of $\hat{D}_x, \hat{D}_y$. Thus (2.7) holds with r, D_x, and D_y replaced by r', $\hat{D}_x$, and $\hat{D}_y$, respectively. Since any nonempty open subset of $\hat{D}_x$ (respectively, $\hat{D}_y$) is also a nonempty subset of D_x (respectively, D_y), we see that Assumption 2.1 remains valid (with the same constants n_f, m_f) when D_x, D_y are replaced by $\hat{D}_x$, $\hat{D}_y$. Therefore, Theorem 2.1 applies and shows that $r > r' \geqq \min\{n_f, m_f\}$, which shows that Theorem 2.1 holds regardless of assumption (2.7).

In the rest of the proof, we will assume that (2.7) holds. Let us consider a protocol that uses exactly r messages, described by (cf. § 1)

$$m_i(x, y) = \hat{m}_i(x, m_1(x, y), \cdots, m_{i-1}(x, y)) \quad \forall (x, y) \in D_x \times D_y \quad \text{if } i \in T_{1\to 2}, \tag{2.8}$$

$$m_i(x, y) = \hat{m}_i(y, m_1(x, y), \cdots, m_{i-1}(x, y)) \quad \forall (x, y) \in D_x \times D_y \quad \text{if } i \in T_{2\to 1}, \tag{2.9}$$

where each m_i and $\hat{m}_i$ is a continuously differentiable function. By symmetry, we can assume that the final evaluation of f is performed by processor P_1. Thus there exists some continuous function h such that

$$f(x, y) = h(x, m_1(x, y), \cdots, m_r(x, y)) \quad \forall (x, y) \in D_x \times D_y. \tag{2.10}$$

Before presenting the main line of argument, we derive three lemmas. Let $u = (x, y)$ and let $D = D_x \times D_y$. Write $m(u) = (m_1(u), \cdots, m_r(u))$ and let $\nabla m(u)$ be the $(m+n) \times r$ matrix whose ith column is the gradient vector $\nabla m_i(u)$, $i = 1, \cdots, r$. Define

$$k = \max_{u \in D} \operatorname{rank} [\nabla m(u)]. \tag{2.11}$$

LEMMA 2.1. $k = r$.

Proof. We show this by contradiction. Suppose that $r > k$. Consider the continuously differentiable mapping $m: D \mapsto \Re^r$, where $D = D_x \times D_y$ is an open set and $m(u) = (m_1(u), \cdots, m_r(u))$. We claim that $\nabla m_1(x, y)$ is not identically zero on the set D. Indeed, if this was the case, then $m_1(x, y)$ would be equal to a constant on the set D, and the first message in the protocol would be redundant. Thus, there would exist a protocol that uses $r-1$ messages, contradicting definition of r. We can therefore apply Theorem A.2 in the Appendix (with the correspondence $m \leftrightarrow F$, $D \leftrightarrow Q$, $r \leftrightarrow s$) to conclude that there exists some positive integer i and some continuously differentiable function g such that

$$m_{i+1}(u) = g(m_1(u), \cdots, m_i(u)) \quad \forall u \in \bar{D}, \tag{2.12}$$

where $\bar{D}$ is some nonempty open subset of D. By taking a subset of $\bar{D}$ if necessary, we can assume that $\bar{D}$ is of the product form $\bar{D}_x \times \bar{D}_y$, where $\bar{D}_x$ and $\bar{D}_y$ are some open subsets of D_x and D_y, respectively. Then, (2.12) would imply that the $(i+1)$st message $m_{i+1}(x, y)$ is redundant for computing f over $\bar{D}_x \times \bar{D}_y$, which contradicts the definition of r (cf. (2.7)). □

Loosely speaking, Lemma 2.1 tells us that each message in an optimal protocol has to contain some "new information" and therefore the corresponding gradient vectors have to be linearly independent. Before we go on to the next lemma, we introduce some more notations. Let $\bar{D}_x \subset D_x$, $\bar{D}_y \subset D_y$ be nonempty open sets such that $\nabla m(u)$ has full rank for every $u \in \bar{D}_x \times \bar{D}_y$. (Such sets can be taken nonempty due to Lemma 2.1, and open due to the continuity of $\nabla m(u)$.) We use $\bar{D}$ as a short notation for $\bar{D}_x \times \bar{D}_y$. Furthermore, for any vector $c = (c_1, \cdots, c_r) \in \Re^r$ and for $i \leqq r$, we let $c^i = (c_1, c_2, \cdots, c_i)$. Let also r_1 (respectively, r_2) be the number of messages sent by processor P_1 (respectively, P_2). In addition, we use the notation $[\nabla_x m_i(x, y);\ i \in T_{1\to 2}]$ to denote the $m \times r_1$ matrix whose column vectors are $\nabla_x m_i(x, y) = (\partial m_i(x, y)/\partial x_1, \cdots, \partial m_i(x, y)/\partial x_m)$, $i \in T_{1\to 2}$. The $n \times r_2$ matrix $[\nabla_y m_i(x, y);\ i \in T_{2\to 1}]$ is defined similarly. As a refinement of Lemma 2.1, we have the following lemma.

LEMMA 2.2. *For any $(x, y) \in \bar{D}$, we have*

$$\operatorname{rank} [\nabla_x \hat{m}_i(x, c^{i-1});\ i \in T_{1\to 2}] = r_1,$$

and

$$\operatorname{rank} [\nabla_y \hat{m}_i(y, c^{i-1});\ i \in T_{2\to 1}] = r_2,$$

where $c = m(x, y)$.

Proof. By Lemma 2.1, we see that the matrix $\nabla m(x, y)$ has full rank (and its rank is equal to r) over the set $\bar{D}$. Note that by possibly reindexing the columns of the matrix $\nabla m(x, y)$, we can write $\nabla m(x, y)$ in the form

$$\nabla m(x, y) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $A_{11} = [\nabla_x m_i(x, y);\ i \in T_{1\to 2}]$ and $A_{22} = [\nabla_y m_i(x, y);\ i \in T_{2\to 1}]$. From (2.8)-(2.9), it is easily seen that for each $i \in T_{2\to 1}$, there exists a continuously differentiable function M_i such that

$$m_i(x, y) = M_i(y, \{m_l(x, y): l < i, l \in T_{1\to 2}\}), \qquad i \in T_{2\to 1}. \tag{2.13}$$

(In other words, a message sent by processor P_2 can be expressed as a function of y and the messages already received.) By differentiating (2.13), we obtain

$$\nabla_x m_i(x, y) = \sum_{l \in T_{1\to 2}, l<i} d_l(x, y) \nabla_x m_l(x, y), \qquad i \in T_{2\to 1}, \tag{2.14}$$

where each $d_l(x, y)$ is a suitable scalar. Thus,

$$\nabla_x m_i(x, y) \in \text{span}\,\{\nabla_x m_l(x, y);\ l \in T_{1\to 2}\} \quad \forall (x, y) \in \bar{D} \quad \forall i \in T_{2\to 1}.$$

This means that the columns of A_{12} belong to the span of the columns of A_{11} and therefore

$$\text{rank}\,[A_{11} \quad A_{12}] = \text{rank}\,(A_{11}) \leqq r_1.$$

Similarly, one can show that

$$\text{rank}\,[A_{21} \quad A_{22}] = \text{rank}\,(A_{22}) \leqq r_2.$$

On the other hand,

$$\begin{aligned} r &= r_1 + r_2 \\ &\geqq \text{rank}\,(A_{11}) + \text{rank}\,(A_{22}) \\ &= \text{rank}\,[A_{11} \quad A_{12}] + \text{rank}\,[A_{21} \quad A_{22}] \\ &\geqq \text{rank} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \\ &= \text{rank}\,[\nabla m(x, y)] \\ &= r \quad \forall (x, y) \in \bar{D}. \end{aligned}$$

This implies that

$$\text{rank}\,(A_{11}) = \text{rank}\,[\nabla_x m_i(x, y);\ i \in T_{1\to 2}] = r_1$$

and

$$\text{rank}\,(A_{22}) = \text{rank}\,[\nabla_y m_i(x, y);\ i \in T_{2\to 1}] = r_2.$$

To show that $\text{rank}\,[\nabla_x \hat{m}_i(x, c^{i-1});\ i \in T_{1\to 2}] = r_1$, we differentiate (2.8) to obtain

$$\nabla_x m_i(x, y) = \nabla_x \hat{m}_i(x, c^{i-1}) + \sum_{l=1}^{i-1} \frac{\partial \hat{m}_i}{\partial m_l}(x, c^{i-1}) \nabla_x m_l(x, y) \quad \text{if } i \in T_{1\to 2}, \tag{2.15}$$

where $c = m(x, y)$ and $(x, y) \in \bar{D}$. Using (2.14), we see that

$$\sum_{l=1}^{i-1} (\partial \hat{m}_i / \partial m_l)(x, c^{i-1}) \nabla_x m_l(x, y)$$

can be written as a linear combination of the vectors $\{\nabla_x m_l(x, y);\ l < i-1, l \in T_{1\to 2}\}$. Therefore, (2.15) shows that

$$[\nabla_x \hat{m}_i(x, c^{i-1});\ i \in T_{1\to 2}] = [\nabla_x m_i(x, y);\ i \in T_{1\to 2}]\ C = A_{11} C,$$

where C is some upper triangular matrix whose diagonal entries are equal to 1. Hence rank $[\nabla_x \hat{m}_i(x, c^{i-1}); i \in T_{1\to 2}] = \text{rank}\,(A_{11}) = r_1$. The equality

$$\text{rank}\,[\nabla_y \hat{m}_i(y, c^{i-1}); i \in T_{2\to 1}] = r_2$$

can be shown by a similar argument. □

Let us fix some more notations. For any vector $c = (c_1, \cdots, c_r) \in \Re^r$ and $1 \leqq i \leqq n$, we let

$$\begin{aligned} S(c) &= \{(x, y) \in D_x \times D_y \,|\, m_i(x, y) = c_i,\ i = 1, \cdots, r\}, \\ S_x(c) &= \{x \in D_x \,|\, \hat{m}_i(x, c^{i-1}) = c_i,\ \forall i \in T_{1\to 2}\}, \\ S_y(c) &= \{y \in D_y \,|\, \hat{m}_i(y, c^{i-1}) = c_i,\ \forall i \in T_{2\to 1}\}, \\ R^r &= \{(m_1(x, y), \cdots, m_r(x, y)) \,|\, (x, y) \in D_x \times D_y\}. \end{aligned} \tag{2.16}$$

LEMMA 2.3. *For any $c \in R^r$, we have*

$$S(c) = S_x(c) \times S_y(c). \tag{2.17}$$

Proof. We have, using definition (2.16) and (2.8)-(2.9),

$$\begin{aligned} S(c) &= \{(x, y) \in D_x \times D_y \,|\, \hat{m}_i(x, c^{i-1}) = c_i, \forall i \in T_{1\to 2},\ \hat{m}_i(y, c^{i-1}) = c_i, \forall i \in T_{2\to 1}\} \\ &= S_x(c) \times S_y(c). \end{aligned}$$

□

We now fix some $(x^*, y^*) \in \bar{D}$ and let $c^* = m(x^*, y^*)$. Let us define

$$F_i(x, c) = \hat{m}_i(x, c^{i-1}) - c_i \quad \forall c \in R^r, \qquad x \in D_x, \qquad i \in T_{1\to 2}.$$

Thus $F_i(x^*, c^*) = 0$ for all $i \in T_{1\to 2}$. Moreover, it follows from Lemma 2.2 that the matrix $[\nabla_x F(x^*, c^*)]$ has full rank. It is now clear that we are in a position to apply Theorem A.3 in the Appendix (with the correspondence that $u \leftrightarrow x$, and $v \leftrightarrow c$) to conclude that there exist an open subset U_1 of $\Re^r$ containing c^*, and an open subset $\hat{D}_x$ of D_x containing x^*, such that $S_x(c) \cap \hat{D}_x$ is nonempty and connected for all $c \in U_1$. Following a symmetrical argument, we see that there exist open subsets $U_2 \subset \Re^r$ and $\hat{D}_y \subset D_y$ such that $c^* \in U_2$, $y^* \in \hat{D}_y$, and $S_y(c) \cap \hat{D}_y$ is nonempty and connected for all $c \in U_2$. Let $U = U_1 \cap U_2$. Clearly, U is nonempty, since $c^* \in U$. In light of Lemma 2.3, we see that for all $c \in U$,

$$\begin{aligned} \hat{S}(c) &\triangleq S(c) \cap (\hat{D}_x \times \hat{D}_y) \\ &= (S_x(c) \cap \hat{D}_x) \times (S_y(c) \cap \hat{D}_y), \end{aligned}$$

and the set $\hat{S}(c)$ is nonempty and connected. Let us use $\hat{S}_x(c)$ and $\hat{S}_y(c)$ to denote the sets $S_x(c) \cap \hat{D}_x$ and $S_y(c) \cap \hat{D}_y$, respectively.

We now proceed to prove Theorem 2.1. Since we have assumed that the final result is evaluated by processor P_1, it follows that the last message $m_r(x, y)$ must have been sent by processor P_2. (Otherwise, processor P_1 would be able to evaluate $f(x, y)$ on the basis of $m_1(x, y), \cdots, m_{r-1}(x, y)$, and we would have a protocol with $r-1$ messages, thus contradicting (2.7).) Suppose that there exists some function $w: U \mapsto \Re$ such that

$$h(x, c) = w(c) \quad \forall c \in U \quad \forall x \in \hat{S}_x(c), \tag{2.18}$$

where h is the function given by (2.10). We claim that w is a continuous function of c in U. In fact, let c be an arbitrary vector in U and let $\{c_i \in U; i = 1, 2, \cdots\}$ be a sequence of vectors converging to c. By Theorem A.3 in the Appendix, we can pick a

convergent sequence of vectors $\{x_i \in \hat{S}_x(c_i);\ i = 1, 2, \cdots\}$ such that $\lim_{i\to\infty} x_i = x$ for some $x \in \hat{D}_x$. By using (2.18) and the continuity of h, we see that

$$\lim_{i\to\infty} w(c_i) = \lim_{i\to\infty} h(x_i, c_i) = h(x, c),$$

which implies that w is continuous on U. Since for any $(x, y) \in m^{-1}(U) \cap (\hat{D}_x \times \hat{D}_y)$ we have $m(x, y) \in U$, (2.18) yields

$$f(x, y) = h(x, m(x, y)) = w(m(x, y)) \quad \forall (x, y) \in \hat{D}_x \times \hat{D}_y.$$

Thus f can be evaluated on the basis of $m(x, y)$ alone over the set $m^{-1}(U) \cap (\hat{D}_x \times \hat{D}_y)$ and this can be done by processor P_2 before sending the last message. Thus (2.18) leads to a protocol with $r - 1$ messages for computing f over $m^{-1}(U) \cap (\hat{D}_x \times \hat{D}_y)$. This will contradict (2.7) once we show that $m^{-1}(U) \cap (\hat{D}_x \times \hat{D}_y)$ is a nonempty open set. To this effect, we notice that $\hat{S}(c)$ is nonempty and that

$$\hat{S}(c) \subset m^{-1}(U) \cap (\hat{D}_x \times \hat{D}_y) \quad \forall c \in U,$$

from which it follows that $m^{-1}(U) \cap (\hat{D}_x \times \hat{D}_y)$ is nonempty. Furthermore, $m^{-1}(U)$ is open since it is the inverse image of the open set U under a continuous mapping. Thus, $m^{-1}(U) \cap (\hat{D}_x \times \hat{D}_y)$ is open, since $\hat{D}_x \times \hat{D}_y$ is open by construction.

Since no function w can have the property (2.18), we conclude that there exists some $\hat{c} \in U$ such that $h(x, \hat{c})$ is a nonconstant function of x on the set $\hat{S}_x(\hat{c})$. Since h is a continuous function and the set $\hat{S}_x(\hat{c})$ is nonempty and connected, we see that $h(\hat{S}_x(\hat{c}), \hat{c})$ must contain an open interval in $\mathfrak{R}$. Using the fact $f(x, y) = h(x, \hat{c})$ for all $(x, y) \in \hat{S}_x(\hat{c}) \times \hat{S}_y(\hat{c})$, we have

$$f(\hat{S}_x(\hat{c}), y) = h(\hat{S}_x(\hat{c}), \hat{c}) \quad \forall y \in \hat{S}_y(\hat{c}).$$

Therefore, $f(\hat{S}_x(\hat{c}), y)$ contains an open interval, or equivalently, $\hat{S}_x(\hat{c}) \in \mathcal{S}^{(2)}(y)$ for all $y \in \hat{S}_y(\hat{c})$ (cf. definition 2.1). Let us fix some $\hat{y} \in \hat{S}_y(\hat{c})$. Then, using the definition of n_f (2.1), there exist $x^1, \cdots, x^{n_f} \in \hat{S}_x(\hat{c})$ such that $\nabla_y f(x^1, \hat{y}), \cdots, \nabla_y f(x^{n_f}, \hat{y})$ are linearly independent. Meanwhile, we observe that

$$\hat{S}_y(\hat{c}) = \{y \in \hat{D}_y \mid \hat{m}_i(y, \hat{c}^{i-1}) = \hat{c}_i \quad \forall i \in T_{2\to 1}\}$$

and that, for any fixed $x \in \hat{S}_x(\hat{c})$, $f(x, y) = h(x, \hat{c})$ is a constant function of y on the set $\hat{S}_y(\hat{c})$. Moreover, by Lemma 2.2, we have

$$\text{rank}\,[\nabla_y \hat{m}_i(y, \hat{c}^{i-1});\ i \in T_{2\to 1}] = r_2 \quad \forall y \in \hat{D}_y. \tag{2.19}$$

Thus we are now in a position to apply Theorem A.4 (with the correspondence $A \leftrightarrow \hat{S}_y(\hat{c})$, $F \leftrightarrow \{\hat{m}_i(y, \hat{c}^{i-1}) - \hat{c}_i;\ i \in T_{2\to 1}\}$) and conclude that

$$\nabla_y f(x, \hat{y}) \in \text{span}\,\{\nabla_y \hat{m}_i(\hat{y}, \hat{c}^{i-1}),\ i \in T_{2\to 1}\} \quad \forall x \in \hat{S}_x(\hat{c}).$$

Since each $x^j \in \hat{S}_x(\hat{c})$, we see that $\nabla_y f(x^j, \hat{y})$ is the span of the vectors $\{\nabla_y \hat{m}_i(\hat{y}, \hat{c}^{i-1}),\ i \in T_{2\to 1}\}$, for $j = 1, \cdots, n_f$. Using the fact that the vectors $\nabla_y f(x^j, \hat{y})$ are linearly independent, we conclude that $r \geqq r_2 \geqq n_f \geqq \min\{m_f, n_f\}$, which is the desired result, under the assumption that processor P_1 performs the final evaluation of f. A similar argument yields $r \geqq r_1 \geqq n_f \geqq \min\{m_f, n_f\}$ for the case where processor P_2 performs the final evaluation of f. This completes the proof of the theorem.

As a remark, we note that in the preceding proof we have actually shown that $r_2 \geqq n_f$ in the case where processor P_1 performs the final computation and $r_1 \geqq m_f$ if processor P_2 performs the final computation. Therefore, if $C_1(f; D_x \times D_y) = \min\{m_f, n_f\}$, then either $r_1 = m_f$ and $r_2 = 0$, or, $r_1 = 0$ and $r_2 = n_f$. This means that our

lower bound is tight only for those problems for which one-way communication protocols are optimal.

COROLLARY 2.1. *If* $C_1(f; D_x \times D_y) = \min\{n_f, m_f\}$, *then any optimal communication protocol for computing* f *over* $D_x \times D_y$ *is necessarily a one-way communication protocol.*

3. Computing a root of a polynomial. We now apply Theorem 2.1 to the distributed computation of a root of a polynomial. We shall demonstrate that in this case Abelson's result is far from optimal.

Let $x = (x_0, \cdots, x_{n-1}) \in \Re^n$ and $y = (y_0, \cdots, y_{n-1}) \in \Re^n$; let $F(z; x, y)$ be the polynomial in the scalar variable z defined by

$$F(z; x, y) = \sum_{i=0}^{n-1} (x_i + y_i) z^i. \tag{3.1}$$

Processor P_1 (respectively, P_2) has access to the vector x (respectively, y); and the objective is the computation of a particular root of the polynomial $F(z; x, y)$. In order for the problem to be well defined, we must specify which one of the $n-1$ roots of the polynomial is to be computed. This is accomplished as follows. We fix some $(x^*, y^*) \in \Re^{2n}$ such that one of the roots (call it z^*) of the polynomial $F(z; x^*, y^*)$ is real and simple. This root will vary continuously and will remain a real and simple root as x and y vary in some open set containing x^*, y^*. We formulate this discussion in the following result.

LEMMA 3.1. *Suppose that* z^* *is a real and simple root of* $F(z; x^*, y^*)$. *Then, there exist open sets* $D_x, D_y \subset \Re^n$ *such that* $(x^*, y^*) \in D_x \times D_y$ *and an infinitely differentiable function* $f: D_x \times D_y \mapsto \Re$ *such that* $f(x^*, y^*) = z^*$ *and*

$$F(f(x, y); x, y) = 0 \quad \forall (x, y) \in D_x \times D_y. \tag{3.2}$$

Proof. Note that $(\partial F/\partial z)(z^*; x^*, y^*) \neq 0$, since z^* is a simple root. By the implicit function theorem ([S65, p. 41]), we see that there exists an open set D containing (x^*, y^*) and an infinitely differentiable function $g: D \mapsto \Re$ such that $g(x^*, y^*) = z^*$ and $F(g(x, y); x, y) = 0$ for all $(x, y) \in D$. Now by the continuity of $(\partial F/\partial z)(z; x, y)\big|_{z=g(x,y)}$ at the point (x^*, y^*), there exist open sets D_x, D_y such that $(x^*, y^*) \in D_x \times D_y \subset D$ and such that $(\partial F/\partial z)(z; x, y)\big|_{z=g(x,y)} \neq 0$ for all $(x, y) \in D_x \times D_y$. As a result, $g(x, y)$ is a simple root of the polynomial equation $F(z; x, y) = 0$ for all $(x, y) \in D_x \times D_y$. Let f be the restriction of g on $D_x \times D_y$. Clearly, f has all the desired properties. □

By Lemma 3.1, we see that $f(x, y)$ is a root of $F(z; x, y)$ and is a well-defined smooth map from $D_x \times D_y$ to $\Re$. We are interested in the communication complexity $C_1(f; D_x \times D_y)$ of computing $f(x, y)$ as (x, y) varies in the set $D_x \times D_y$. We start by pointing out that Abelson's lower bound (Theorem 1.1) is rather weak.

LEMMA 3.2. *The rank of the matrix* $H_{xy}(f)$, *whose* (i, j)*th entry is equal to* $\partial^2 f/\partial x_i \partial y_j$, *is at most* 3, *for any* $(x, y) \in D_x \times D_y$.

Proof. We have

$$\sum_{i=0}^{n-1} (x_i + y_i)(f(x, y))^i = 0 \quad \forall (x, y) \in D_x \times D_y.$$

We differentiate both sides of the above equation, with respect to y_m, to obtain

$$\sum_{i=1}^{n-1} i(x_i + y_i)(f(x, y))^{i-1} \cdot \frac{\partial f(x, y)}{\partial y_m} + (f(x, y))^m = 0 \quad \forall (x, y) \in D_x \times D_y, \tag{3.3}$$

$$0 \leqq m \leqq n-1.$$

We differentiate (3.3) further, with respect to x_l, to obtain

$$\begin{aligned}&\sum_{i=1}^{n-1} i(i-1)(x_i+y_i)(f(x,y))^{i-2}\frac{\partial f(x,y)}{\partial x_l}\frac{\partial f(x,y)}{\partial y_m}\\&\quad+\sum_{i=1}^{n-1} i(x_i+y_i)(f(x,y))^{i-1}\frac{\partial^2 f(x,y)}{\partial x_l \partial y_m}\\&\quad+m(f(x,y))^{m-1}\frac{\partial f(x,y)}{\partial x_l}+l(f(x,y))^{l-1}\frac{\partial f(x,y)}{\partial y_m}=0,\end{aligned}\tag{3.4}$$

$$\forall (x,y)\in D_x\times D_y,\qquad 0\leqq m,\ l\leqq n-1.$$

Since $f(x,y)$ is a simple root, it follows that $\sum_{i=1}^{n-1} i(x_i+y_i)(f(x,y))^{i-1}\neq 0$. Equation (3.4) shows that $\partial^2 f(x,y)/\partial x_l\partial y_m$ is of the form $u_1(l)v_1(m)+u_2(l)v_2(m)+u_3(l)v_3(m)$, where $u_i(l)$, $v_i(m)$ are some real numbers depending on x, y. Therefore the rank of the matrix $H_{xy}(f)$ can be at most 3, for any point $(x,y)\in D_x\times D_y$. □

We now illustrate the power of our general results by deriving a lower bound that matches the obvious upper bound.

THEOREM 3.1. *Let D_x, D_y be as in Lemma 3.1. Then, $C_1(f(x,y); D_x\times D_y)=n$.*

Proof. The upper bound $C_1(f; D_x\times D_y)\leqq n$ is obvious, so we concentrate on the proof of the lower bound. To this effect, we will employ Theorem 2.1 and it suffices to verify that Assumption 2.1 holds with $n_f=m_f=n$. Since the roots of a polynomial equation cannot remain constant when the coefficients vary over an open set, it follows that the continuous function $f(x,y)$ given by Lemma 3.1 satisfies parts (a) and (b) of Assumption 2.1. Now we fix some $y\in D_y$ and some $S\in\mathcal{S}^{(2)}(y)$, that is, $S\subset D_x$ and $f(S,y)$ contains an open interval. Let $c_1,\cdots,c_n$ be some distinct real numbers in $f(S,y)$ and $x^1,\cdots,x^n\in S$ such that

$$f(x^i,y)=c_i,\qquad i=1,\cdots,n.\tag{3.5}$$

Let x_j^i be the jth coordinate of x^i. Using (3.3), we see that

$$a_i\nabla_y f(x^i,y)=-\begin{bmatrix}1\\ c_i\\ \vdots\\ c_i^{n-1}\end{bmatrix},\tag{3.6}$$

where $a_i=\sum_{j=1}^{n-1} j(x_j^i+y_j)c_i^{j-1}$. If we form a matrix whose columns are the vectors $(1, c_i,\cdots,c_i^{n-1})$, $i=1,\cdots,n$, this matrix is a Vandermonde matrix and is nonsingular, because the values $c_1,\cdots,c_n$ are chosen to be distinct. Then, (3.6) implies that the vectors $\nabla_y f(x^i,y)$, $i=1,\cdots,n$, are linearly independent. This proves that $n_f=n$. The proof that $m_f=n$ is similar. □

As a remark, we point out that Theorem 3.1 is in some sense the strongest result possible. The only assumptions we used in showing Theorem 3.1 are that (a) the message functions are continuously differentiable; (b) the final evaluation function is a continuous function; (c) the protocol computes a root of a polynomial on some open set. As discussed in [L89], assumption (a) is necessary since its removal could lead to unreasonable conclusions. Assumption (b) is basic and natural since the function to be computed, i.e., a particular real simple root of some polynomial, is continuous, while assumption (c) is minimal. Finally, we note that no truly two-way communication protocol can be optimal. In other words, if each processor transmits at least one message, then at least $n+1$ messages have to be exchanged. This is a simple consequence of Corollary 2.1 of § 2.

4. Comparison with Abelson's bound. In the previous section, we saw that Theorem 2.1 can yield a much better bound than Abelson's result (Theorem 1.1). However, it is not true, as we shall see next, that Theorem 2.1 always provides a stronger lower bound. The reason is, loosely speaking, that our result only places a constraint on the minimum number of messages that has to be sent by a single processor, while Abelson's result is a bound on the total number of messages sent by both processors. As pointed out at the end of § 2, any two-way communication protocol that attains the lower bound in Theorem 2.1 is necessarily a one-way protocol. Notice that our result makes use of information about the first-order derivatives of function f. This is in contrast to Abelson's result which uses only the second-order derivatives of f. In what follows, we provide an example where Abelson's bound is more effective than our bound.

Let $f(x, y) = x^T Q y$, where Q is some $m \times n$ matrix, $x \in \Re^m$ and $y \in \Re^n$. By Theorem 1.1, we see that $C_2(f; \Re^m \times \Re^n) \geqq \text{rank}\,(Q)$. Using the singular value decomposition of Q, one can construct a protocol that uses exactly rank(Q) messages (see [LT89]). Therefore, we conclude that $C_2(f; \Re^m \times \Re^n) = \text{rank}(Q)$. To see what lower bounds are provided by Theorem 2.1, we need to calculate the values of m_f and n_f.

Suppose that rank $(Q) = r > 0$. Let D_x, D_y be some convex open subsets of $\Re^m$ and $\Re^n$, respectively. We assume that $0 \notin D_x$ and $0 \notin D_y$, in which case $f(x, y)$ is nonconstant as x or y vary in an open subset of D_x or D_y, respectively. Thus parts (a) and (b) of Assumption 2.1 are satisfied. We now show that Assumption 2.1 can only hold with $\min\{m_f, n_f\} \leqq 2$. By the singular value decomposition, there exist two linearly independent families of vectors $u_1, \cdots, u_r$ in $\Re^m$ and $v_1, \cdots, v_r$ in $\Re^n$, such that

$$Q = u_1 v_1^T + u_2 v_2^T + \cdots + u_r v_r^T. \tag{4.1}$$

It follows that $x^T Q y = \sum_{i=1}^r (u_i^T x)(v_i^T y)$. Since $r > 0$, there exists some point $(x_0, y_0) \in D_x \times D_y$ such that $x_0^T Q y_0 \neq 0$. Hence, we can, without loss of generality, assume that $(u_r^T x_0)(v_r^T y_0) \neq 0$. Let $S = \{x \in D_x \mid u_i^T x = u_i^T x_0,\ 1 \leqq i \leqq r-1\}$. Clearly, S is nonempty since $x_0 \in S$. We claim that if $r > 1$, then $f(S, y_0)$ contains an open interval. In fact, (4.1) shows that

$$\begin{aligned} x^T Q y_0 &= \sum_{i=1}^{r} (u_i^T x)(v_i^T y_0) \\ &= \sum_{i=1}^{r-1} (u_i^T x_0)(v_i^T y_0) + (u_r^T x)(v_r^T y_0) \quad \forall x \in S. \end{aligned} \tag{4.2}$$

Since u_r is linearly independent from $u_1, \cdots, u_{r-1}$, we see that $u_r^T x$ is a nonconstant function of x on S. Using (4.2) and the fact that $v_r^T y_0 \neq 0$, we see that $x^T Q y_0$ is also a nonconstant function of x on the set S. Note that S is connected because D_x is assumed to be convex. It follows that $f(S, y_0)$ contains an open interval. To see that $n_f \leqq 2$, we note that

$$\nabla_y f(x, y_0) = \sum_{i=1}^{r-1} (u_i^T x_0) v_i + (u_r^T x) v_r \quad \forall x \in S.$$

Hence, $\dim [\nabla_y f(x, y_0);\ x \in S] \leqq 2$. Thus Assumption 2.1 can only hold with $n_f \leqq 2$. The relation $m_f \leqq 2$ can be established in a symmetrical fashion. As a result, we have shown that $\min\{m_f, n_f\} \leqq 2$.

Thus, for the problem $f(x, y) = x^T Q y$, Theorem 2.1 provides a lower bound of at most 2, as opposed to the lower bound of rank(Q) provided by Abelson's result. Hence, Theorem 2.1 can be quite far from optimal, in general. Furthermore, the above example and the results of § 3 illustrate that Theorems 1.1 and 2.1 are incomparable.

Appendix. This appendix contains some results concerning multivariable functions that are used in § 2.

Let $F: U \times V \mapsto \Re^s$ be a continuously differentiable mapping, where U and V are open subsets of $\Re^r$ and $\Re^t$, respectively. We assume that $r > s$ and that rank $[\nabla_u F(u^*, v^*)] = s$, for some $(u^*, v^*) \in U \times V$. Then, the matrix $\nabla F_u(u^*, v^*)$ has s linearly independent rows and we can find a set $J \subset \{1, \cdots, r\}$ of indices, of cardinality s, such that the vectors $(\partial F_1(u^*, v^*)/\partial u_i, \cdots, \partial F_s(u^*, v^*)/\partial u_i)$, $i \in J$ are linearly independent. We define the projection $\Pi: \Re^r \mapsto \Re^{r-s}$ by letting $\Pi(u)$ be the vector with coordinates u_i, $i \notin J$. We have the following lemma.

LEMMA A.1. *There exists a connected open subset R of $U \times V$, and a connected open set $S \subset \Re^{r+t}$, and a continuously differentiable function $g: S \mapsto R$ such that $(u^*, v^*) \in R$,*

$$S = \{(F(u, v), \Pi(u), v) \mid (u, v) \in R\},$$

and such that

$$(u, v) = g(F(u, v), \Pi(u), v) \quad \forall (u, v) \in R. \tag{A.1}$$

Proof. Consider the mapping $q: U \times V \mapsto \Re^{r+t}$ defined by $q(u, v) = (F(u, v), \Pi(u), v)$. We claim that $\nabla q(u^*, v^*)$ has full rank. To see this, let us permute the rows of $\nabla q(u^*, v^*)$ so that the last $r + t - s$ rows correspond to the partial derivatives with respect to the variables v and u_i, $i \notin J$. Then, $\nabla q(u^*, v^*)$ will have the structure

$$\nabla q(u^*, v^*) = \begin{bmatrix} A & 0 \\ B & I \end{bmatrix},$$

where A, B are suitable submatrices of $\nabla F(u^*, v^*)$ and I is the $(r+t-s) \times (r+t-s)$ identity matrix. Each one of the s rows of matrix A is a vector of the form $(\partial F_i(u^*, v^*)/\partial u_i, \cdots, \partial F_s(u^*, v^*)/\partial u_i)$, $i \in J$, and these vectors are linearly independent by construction. Thus $\det(\nabla q(u^*, v^*)) = \det(A) \neq 0$. The result then follows from the inverse function theorem [S65, p. 35]. □

THEOREM A.1. *Let Q be an open subset of $\Re^r$. Let $F: Q \mapsto \Re^s$ be a continuously differentiable mapping such that*

$$\max_{z \in Q} \operatorname{rank}(\nabla F(z)) = s. \tag{A.2}$$

Suppose that $f: Q \mapsto \Re$ is a continuously differentiable function with the property

$$\nabla f(z) \in \operatorname{span}\{\nabla F(z)\} \quad \forall z \in Q.$$

Then, there exists some continuously differentiable function h such that $f(z) = h(F(z))$ for all $z \in R$, where R is some open subset of Q.

Proof. Suppose that $z^* \in Q$ is a vector at which the maximum in (A.2) is attained. By taking $t = 0$ and dropping the set V, we see that all the assumptions of Lemma A.1 are satisfied[2], and thus Lemma A.1 applies. Let R, S, and g be as in Lemma A.1. By assumption, $\nabla f(z) \in \operatorname{span}\{\nabla F(z)\}$, for all $z \in R$. Thus, for every $z \in R$, there exists a vector $d(z) \in \Re^s$ such that

$$\nabla f(z) = \nabla F(z) d(z) \quad \forall z \in R. \tag{A.3}$$

[2] We have assumed that $r > s$ here. The proof for the case $r = s$ is essentially the same except that Π is redundant.

Using Lemma A.1, we have

$$F(z) = F(g(F(z), \Pi(z))) \quad \forall z \in R,$$

or

$$u = F(g(u, v)) \quad \forall (u, v) \in S. \tag{A.4}$$

Let $\nabla_v g$ be the $(r-s)\times r$ matrix of the partial derivatives of g, with respect to the components of v. Since the left hand side of (A.4) does not depend on v, the chain rule yields

$$0 = \nabla_v g(u, v) \cdot \nabla F(g(u, v)) \quad \forall (u, v) \in S. \tag{A.5}$$

We use Lemma A.1 once more to obtain

$$f(z) = f(g(F(z), \Pi(z))) \quad \forall z \in R.$$

We define a function $\bar{h}: S \mapsto \Re$ by letting

$$\bar{h}(u, v) = f(g(u, v)) \quad \forall (u, v) \in S. \tag{A.6}$$

Note that $\bar{h}$ is continuously differentiable. Using the chain rule,

$$\nabla_v \bar{h}(u, v) = \nabla_v g(u, v) \cdot \nabla f(g(u, v)) \quad \forall (u, v) \in S,$$

where $\nabla_v \bar{h}(u, v)$ is the vector of partial derivatives of $\bar{h}$ with respect to the components of v. Using (A.3) and (A.5), we conclude that $\nabla_v \bar{h}(u, v) = 0$, for all $(u, v) \in S$. Since S is open and connected, it is easily shown that $\bar{h}$ is independent of v and there exists a continuously differentiable function $h: V \mapsto \Re$ such that

$$\bar{h}(u, v) = h(u) \quad \forall (u, v) \in S.$$

Here $V = F(R)$, which is obviously open and connected. For any $z \in R$, we have

$$f(z) = f(g(F(z), \Pi(z))) = \bar{h}(F(z), \Pi(z)) = h(F(z)),$$

as desired. □

THEOREM A.2. *Let $F: Q \mapsto \Re^s$ be continuously differentiable, where $Q \subset \Re^r$ is open. We assume that* rank $(\nabla F(z)) < s$, *for all $z \in Q$, and that $\nabla F_1(z)$ (the first component mapping of F) is not equal to zero on the set Q. Then, there exists some positive integer i and some continuously differentiable function h such that*

$$F_{i+1}(z) = h(F_1(z), \cdots, F_i(z)) \quad \forall z \in R,$$

where R is some nonempty open subset of Q and F_i denotes the ith component mapping of F.

Proof. We let i be the largest index such that there exists some $\hat{z} \in Q$ with the property

$$\dim \text{span}\,\{\nabla F_1(\hat{z}), \cdots, \nabla F_i(\hat{z})\} = i.$$

Clearly, $1 \leqq i < s$. By continuity, there exists some open subset $\hat{Q}$ of Q containing $\hat{z}$ such that $\nabla F_1(z), \cdots, \nabla F_i(z)$ are linearly independent for all $z \in \hat{Q}$. By our choice of the index i, we have

$$\nabla F_{i+1}(z) \in \text{span}\,\{\nabla F_1(z), \cdots, \nabla F_i(z)\} \quad \forall z \in \hat{Q}.$$

By Theorem A.1, we see that there exists a continuously differentiable function $h: U \mapsto \Re$ such that

$$F_{i+1}(z) = h(F_1(z), \cdots, F_i(z)) \quad \forall z \in R$$

where R is some open subset of $\hat{Q}$ and $U = F(R)$. □

THEOREM A.3. *Let $F: U \times V \mapsto \Re^s$ be a continuously differentiable mapping, where U and V are open subsets of $\Re^r$ and $\Re^t$, respectively. Let $(u^*, v^*) \in U \times V$ be such that* $\operatorname{rank}[\nabla_u F(u^*, v^*)] = s$ *and* $F(u^*, v^*) = 0$. *Then, there exists some nonempty open subsets $W \subset U$, $\bar{V} \subset V$ such that $u^* \in W$, $v^* \in \bar{V}$, and*

$$\{u \mid F(u, v) = 0\} \cap W$$

is nonempty and connected for all $v \in \bar{V}$. Furthermore, if $\{v_i \in V;\ i = 1, 2, \cdots\}$ is a sequence of vectors such that $\lim_{i\to\infty} v_i = v$ and $v \in \bar{V}$, then there exists a sequence $\{u_i \in W\}$ such that $F(u_i, v_i) = 0$ and $\lim_{i\to\infty} u_i = u$ for some $u \in W$.

Proof. We are in a situation where the assumptions of Lemma A.1 hold.[3] Let q, g, R, S be given as in Lemma A.1. Thus $(u, v) = g(q(u, v)) = g(F(u, v), \Pi(u), v)$, for all $(u, v) \in R$. Let g_u, g_v be the corresponding component mappings of g such that $u = g_u(q(u, v))$ and $v = g_v(q(u, v))$. Since S is open, we can take a connected open subset of S with the form $W_1 \times W_2 \times \bar{V}$ such that $W_1 \subset \Re^s$, $W_2 \subset \Re^{r-s}$, and $q(u^*, v^*) \in W_1 \times W_2 \times \bar{V}$. It is easy to check that W_2 is nonempty and connected and that $v^* \in \bar{V}$. Since g is a diffeomorphism, it follows that the set $g(W_1 \times W_2 \times \bar{V})$ is open. Moreover, we claim that g has following properties:

(a) $g_v(w_1, w_2, v) = v$ for all $(w_1, w_2, v) \in W_1 \times W_2 \times \bar{V}$;

(b) $\Pi(g_u(w_1, w_2, v)) = w_2$ for all $(w_1, w_2, v) \in W_1 \times W_2 \times \bar{V}$.

To prove the first property, let us write $(w_1, w_2, v) = q(u, v')$ for some $(u, v') \in R$. This is possible since $(w_1, w_2, v) \in S$. Hence, $(w_1, w_2, v) = (F(u, v'), \Pi(u), v')$. It follows that $v = v'$ and $(w_1, w_2, v) = q(u, v)$. Thus, $g_v(w_1, w_2, v) = g_v(q(u, v)) = v$, which proves (a). We now show the second property. As we have just seen, there exists some u such that $(w_1, w_2, v) = q(u, v)$ and $(u, v) \in R$. Thus, $(w_1, w_2, v) = (F(u, v), \Pi(u), v)$, from which it follows that $w_2 = \Pi(u)$. On the other hand, we have

$$\Pi(g_u(w_1, w_2, v)) = \Pi(g_u(q(u, v))) = \Pi(u),$$

from which it follows that $w_2 = \Pi(g_u(w_1, w_2, v))$.

Now let $W = g_u(W_1 \times W_2 \times \bar{V})$ and $S_u(v) = \{u \in U \mid F(u, v) = 0\}$. Since W is the projection of the open set $g(W_1 \times W_2 \times \bar{V})$, it follows that W is open in $\Re^r$. Also, it can easily be seen that $W \subset U$ and $u^* \in W$. Furthermore, we claim that

$$S_u(v) \cap W = \{g_u(0, w_2, v) \mid w_2 \in W_2\} \quad \forall v \in \bar{V}. \tag{A.7}$$

In fact, let us fix some $v \in \bar{V}$ and let $E(v)$ be the set in the right-hand side of (A.7). We will show that $E(v) \subset S_u(v) \cap W$. Clearly, $E(v) \subset W$. Thus, we only need to show that $E(v) \subset S_u(v)$. Let u be an element of $E(v)$. Then, there exists some $w_2 \in W_2$ such that $u = g_u(0, w_2, v)$. Since $q(u^*, v^*) = (F(u^*, v^*), \Pi(u^*), v^*) = (0, \Pi(u^*), v^*)$ and $q(u^*, v^*) \in W_1 \times W_2 \times \bar{V}$, we see that $0 \in W_1$. Thus, $(0, w_2, v) \in W_1 \times W_2 \times \bar{V}$. In light of property (a), we see that $v = g_v(0, w_2, v)$. Consequently,

$$F(u, v) = F(g_u(0, w_2, v), g_v(0, w_2, v)) = F(g(0, w_2, v)) = 0.$$

It follows that $E(v) \subset S_u(v) \cap W$.

For the reverse inclusion, given any $u \in S_u(v) \cap W$, we have $F(u, v) = 0$. Furthermore, there exists some $(w_1, w_2, v') \in W_1 \times W_2 \times \bar{V}$ such that $u = g_u(w_1, w_2, v')$. By property (b), we see that $\Pi(u) = w_2$. Thus $(0, w_2, v) = (F(u, v), \Pi(u), v) = q(u, v)$. Hence, $u = g_u(q(u, v)) = g_u(0, w_2, v)$. This implies that $u \in E(v)$, and (A.7) has been established. As a result, the set $S_u(v) \cap W$ is connected because, according to (A.7),

[3] Here we have assumed that $r > s$. The same argument works for the case $r = s$ except that Π should be dropped in the remaining proof.

it is the image of the connected set W_2 under a continuous mapping. Since $E(v)$ is nonempty for each $v \in \bar{V}$, (A.7) also shows that $S_u(v) \cap W$ is nonempty.

Given a sequence of vectors $\{v_i \in \bar{V};\ i = 1, 2, \cdots\}$ such that $\lim_{i\to\infty} v_i = v$ and $v \in \bar{V}$, let us pick $u_i = g_u(0, w_2, v_i)$, $i = 1, 2, \cdots$, where w_2 is some fixed vector in W. Hence, $u_i \in E(v_i)$ for all i. According to (A.7), we see that $F(u_i, v_i) = 0$. Furthermore, by the continuity of g_u, we see that

$$\lim_{i\to\infty} u_i = \lim_{i\to\infty} g_u(0, w_2, v_i) = g_u(0, w_2, v),$$

which is clearly in W. □

THEOREM A.4. *Let Q be an open set in $\Re^t$. Let also $F: Q \mapsto \Re^s$ be a continuously differentiable mapping such that*

$$\text{rank}\,(\nabla F(z)) = s \quad \forall z \in A, \tag{A.8}$$

where $A = \{z \mid F(z) = 0\}$. Suppose that $f: Q \mapsto \Re$ is continuously differentiable and is a constant function of z on A. Then,

$$\nabla f(z) \in \text{span}\,\{\nabla F(z)\} \quad \forall z \in A. \tag{A.9}$$

Proof. Consider the following constrained optimization problem:

$$\min_{z\in A} f(z). \tag{A.10}$$

By assumption, each z in A is an optimal solution to (A.10). Since the regularity condition (A.8) ensures the existence of a set of Lagrange multipliers, the necessary condition for optimality gives the desired result ([L84, p. 300]). □

REFERENCES

[A80] H. ABELSON, *Lower bounds on information transfer in distributed computations*, J. Assoc. Comput. Mach., 27 (1980), pp. 384-392.

[AU83] A. V. AHO, J. D. ULLMAN, AND M. YANNAKAKIS, *On notions of information transfer in* VLSI *circuits*, in Proc. 15th ACM Symposium on Theory of Computing, 1983, pp. 133-139.

[BT89] D. P. BERTSEKAS, AND J. N. TSITSIKLIS, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[L84] D. G. LUENBERGER, *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1984.

[L89] Z. Q. LUO, *Communication complexity of some problems in distributed computation*, Ph.D. thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[LT89] Z. Q. LUO AND J. N. TSITSIKLIS, *On the communication complexity of distributed algebraic computation*, Tech. Report LIDS-P-1851, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, *J. Assoc. Comput. Mach.*, submitted, 1989.

[MS82] K. MEHLHORN AND E. M. SCHMIDT, *Las Vegas is better than determinism in* VLSI *and distributed computing*, in Proc. 14th ACM *Symposium on Theory of Computing*, 1982, pp. 330-337.

[PE86] K. F. PANG AND A. EL GAMAL, *Communication complexity of computing the Hamming distance*, SIAM J. Comput., 15 (1986), pp. 932-947.

[PS82] C. H. PAPADIMITRIOU AND M. SIPSER, *Communication complexity*, in Proc. 14th ACM Symposium on Theory of Computing, 1982, pp. 196-200.

[PT82] C. H. PAPADIMITRIOU AND J. N. TSITSIKLIS, *On the complexity of designing distributed protocols*, Inform. and Control, 53 (1982), pp. 211-218.

[S65] M. SPIVAK, *Calculus on Manifolds*, W. A. Benjamin, New York, 1965.

[TL87] J. N. TSITSIKLIS AND Z. Q. LUO, *Communication complexity of convex optimization*, J. Complexity, 3 (1987), pp. 231-243.

[U84] J. D. ULLMAN, *Computational Aspects of* VLSI, Computer Science Press, Rockville, MD, 1984.

[Y79] A. C. YAO, *Some complexity questions related to distributed computing*, in Proc. 11th ACM Symposium on Theory of Computing, 1979, pp. 209-213.