Theoretical Computer Science 6 (1978) 41–67.

TOWARDS A THEORY OF LOCAL AND GLOBAL IN COMPUTATION

Harold ABELSON

Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Communicated by Albert Meyer
Received February 1977
Revised July 1977

Abstract. We formulate the rudiments of a method for assessing the difficulty of dividing a computational problem into "independent simpler parts". This work illustrates measures of complexity which attempt to capture the distinction between "local" and "global" computational problems. One such measure is the *covering multiplicity*, or average number of partial computations which take account of a given piece of data. Another measure reflects the intuitive notion of a "highly interconnected" computational problem, for which subsets of the data cannot be processed "in isolation". These ideas are applied in the setting of computational geometry to show that the connectivity predicate has unbounded covering multiplicity and is highly interconnected; and in the setting of numerical computations to measure the complexity of evaluating polynomials and solving systems of linear equations.

1. Introduction

Many approaches to computational complexity focus on issues concerning the speed of computation: How many basic operations are required to compute a given function? How can computation time be decreased by performing operations in parallel rather than serially? What are the time-space tradeoffs for a given class of algorithms? There is, however, another kind of complexity. This is the *organizational* or *structural* complexity of processes realized by large numbers of interconnected elements. Structural complexity is an important concern in many of the difficult areas to which computational methods are just beginning to be applied. For example, in exploring computational models for vision [6], one is struck by the fact that for a contemporary digital computer, the ratio of connections to components is about three, whereas for the mammalian cortex it lies between 10 and 10 000. A comparison such as this raises a challenge for theoretical computer science: *Is it possible to characterize those computational problems whose solution is inherently better suited to a highly interconnected structure than to a weakly interconnected structure?*

A similar issue arises in the study of distributed computation and computer networks. In this setting, an entity such as a data base might be widely distributed among the nodes of a network. Analyses of sorting or searching the data base must be concerned not only with elementary operations but also with internodal communication. Consider, for example, solving a large system of linear equations, where each column of the matrix resides at a different node. If the network is far-flung, the time and cost of communication could dominate the solution process, and arithmetic operations performed at individual nodes might be viewed simply as overhead. In a case like this, one would be concerned with minimizing, not necessarily the number of operations, but rather the total amount of information which must be shipped across the network. What complexity measure is appropriate to this problem?

Many of the issues in structural complexity revolve around the notion of "local and global" or "parts and wholes". Loosely speaking, a computational problem is *inherently local* if it can be divided into small, weakly interacting modules. A computational problem is *inherently global* if any way of dividing it into pieces must entail substantial interaction among the pieces. Creating a useful theory of local and global is of course a formidable task, and this paper can be no more than an initial attempt. I introduce a measure, called the *covering multiplicity*, which reflects the organizational complexity of a problem in the sense hinted at above. Covering multiplicity is, roughly, the number of independent parts of a process which must take account of a given piece of data. In visual processing this might be, for example, the average number of "low level" elements influenced by a given patch of the retina. The concept of covering multiplicity surely does not capture all of what might be meant by "local and global in computation"; but it is at least a start.

This introduction continues with a review of the setting established by Minsky and Papert [7] in their analysis of the perceptron. We will make use of the same basic framework, although many of Minsky and Papert's techniques, relying fundamentally on the linearity of the perceptron's decision element, are unsuitable in the present, more general setting. Section 2 of the paper begins the formal presentation, providing definitions both of covering multiplicity and also another complexity measure based on the idea of a "highly interconnected" computation. We find that computations which determine whether or not a geometric figure is connected must exhibit arbitrarily high covering multiplicity, and must be highly interconnected, thus providing a justification of Minsky and Papert's intuitive guess [7]: "We chose [in studying perceptrons] to investigate *connectedness* because of a belief that this predicate is nonlocal in some very deep sense." Section 3 turns from geometry to the computation of real-valued functions and gives a necessary and sufficient condition for "computational decomposability" which is used to identify multivariate polynomials whose evaluation requires arbitrarily high covering multiplicity. We also discuss the matrix problem cited above. The conclusion notes some of the many questions which are left untouched in this initial treatment of local and global complexity.

1.1. The perceptron framework

Minsky and Papert's theory begins with an idealized *retina* R, which is simply a collection of n points. Figures on the retina are subsets $X \subset R$. We can think of R as the squares in a two-dimensional plane grid and "arbitrary geometric figures" as approximated by some collection of squares. A *predicate* on R is a function f from figures on R to $\{0, 1\}$. The *support* of f is the set of all points of R which affect the value of f, and the *order* of f is the size of support(f). A *perceptron* is a predicate which has the form

$$f(X) = \lceil \sum_i a_i f_i(X) > 0 \rceil,$$

where f_i are predicates and θ, a_1, a_2, ... are real numbers. (We follow Minsky and Papert in using the notation $\lceil$some condition$\rceil$ to signify the predicate whose value is 1 if the condition is true and 0 if the condition is false.) The *order of the perceptron* is the maximum order of any of the f_i. Minsky and Papert characterize the complexity of geometric predicates in terms of the order of the perceptrons which compute them. They demonstrate, for example, that the predicate $\lceil X$ is connected$\rceil$ has *infinite order*, that is, the order of any perceptron which computes connectivity must become arbitrarily large as the size of the retina becomes large.

1.2. Insufficiency of the perceptron analysis as a general theory of local and global

The notion of order has considerable appeal as a characterization of "local versus global": A low order perceptron is local—the partial predicates can make independent computations based on small patches of the retina. A high order computation is global—individual partial predicates must access large portions of the retina. Unfortunately, this characterization fails when we consider structures more general than the perceptron.

A perceptron can be viewed as a composition of functions $f = g(f_1, f_2, \ldots, f_r)$ where g is a predicate on Boolean r-tuples. In a perceptron, g must be a linear threshold function. We would like to consider more general computational schemes in which there are no restrictions on g. Extending Minsky and Papert's results to this more general setting, however, raises many problems. Consider their paradigm result: Collectivity is not of finite order. This follows from an analysis whose main step is: *Parity*, that is, the predicate

$$\lceil X \text{ contains an even number of points} \rceil$$

is not of finite order. But this can be true only in the linear threshold context. If we omit this restriction, then the determination of parity can be as "local" as we please: For any arbitrary division of the retina into disjoint sets S_i let

$$f_i(X) = \lceil X \cap S_i \text{ has even parity} \rceil.$$

Then X has even parity if and only if the product over all i of $2f_i(X) - 1$ is positive.

To obtain a hold on what makes the above parity computation "local" and why we suspect that any connectivity computation must be "global" notice that for parity the supports are *disjoint*—each point of the retina is examined by only one partial predicate. Moreover, *any* division of the retina into disjoint sets can serve as the partial predicate supports for computing parity. Let's examine in this light a predicate which we might agree is "local", the property of being *locally convex*, which can be determined by checking that X has non-negative curvature at each of its boundary points. Even though this determination is "local", it cannot be easily realized with disjoint supports. Suppose, for example, that we divide the retina in two disjoint halves and attempt to compute local convexity. Consider the shape in Fig. 1. Although it is not convex, the point of negative curvature will be undetected

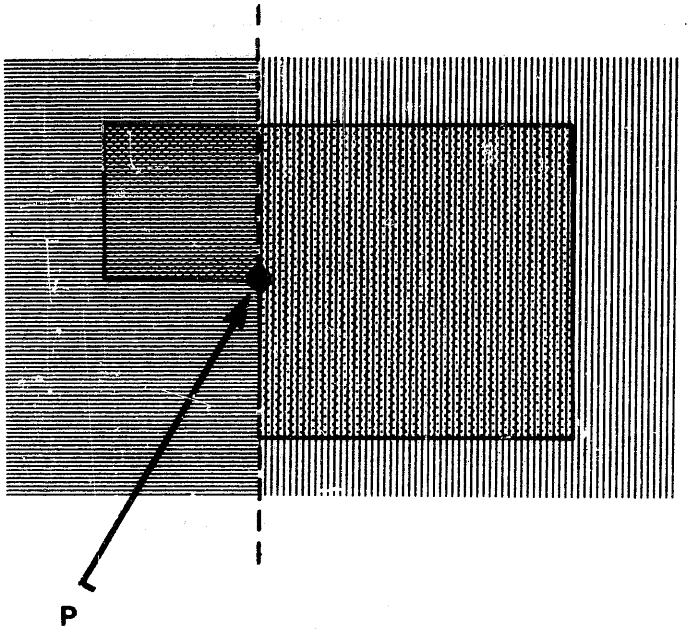


Fig. 1. With disjoint supports, the point of negative curvature is missed.

in either support. But this is hardly a fundamental problem. We merely need to allow a bit of overlap as in Fig. 2, so that a few points of the retina lie in more than one support, and our local computation can proceed without problem. We suspect, though, that no such simple scheme can work for connectivity. In determining connectivity, we would guess, points must in general be accessed by many partial predicates. This provides motivation for our definition of covering multiplicity: *The covering multiplicity μ is the average number of partial predicates which examine a given point on the retina.*

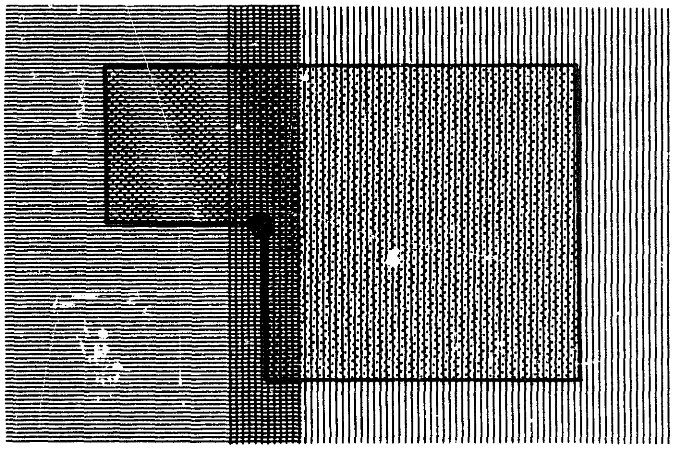


Fig. 2. Curvature is detected using overlapping supports.

2. Support structures for geometric predicates

2.1. *Basic definitions*

Definition 2.1.1. For any function of n variables $f(x_1, \ldots, x_n)$ the set $R = \{1, \ldots, n\}$ is called the *retina* of f. The *support* of f is the smallest set $S \subset R$ such that

$$f(x_1, \ldots, x_n) = f(y_1, \ldots, y_n),$$

whenever $x_i = y_i$ for all $i \in S$.

Throughout Section 2, we shall assume that the functions concerned are Boolean functions of Boolean variables, although later on we shall also consider real-valued functions of real variables. Note that if R is finite, then any function has a unique support.

Definition 2.1.2. A *support structure* on a retina R is a sequence of the form

$$H = \{S_1, \ldots, S_r\},$$

where $S_j \subset R$. Note that the S_j need not be distinct. The number r is called the *rank* of H. The *order* of H is the maximum over j of $|S_j|$. For any $A \subset R$ define $\text{cov}_H(A)$ to be the number of supports of the structure which intersect A:

$$\text{cov}_H(A) = |\{j \colon A \cap S_j \neq \emptyset\}|.$$

The *covering multiplicity* of H is the average over R of the number of supports containing a given point:

$$\mu(H) = (1/n) \sum_{i \in R} \text{cov}(\{i\}).$$

Note that the sum over R of $\text{cov}(\{i\})$ is precisely the number of pairs (i, S_j) where $i \in S_j$. Therefore, if all the S_j have the same size k, then the covering multiplicity, the order, the rank and the size of the retina are related by $\mu(H)n = kr$.

Definition 2.1.3. A support structure $H = (S_i)$ is said to *admit* a function $f(x_1, \ldots, x_n)$ if f can be represented in the form

$$f = g(f_1, \ldots, f_r),$$

where support $(f_i) = S_i$ for $i = 1, \ldots, r$.

If f is a function from $\{0, 1\}^R$ to $\{0, 1\}$, i.e., a "predicate on R", we might try to define the covering multiplicity of f to be the minimum covering multiplicity of any support structure which admits f. But this will not work, since any predicate is admitted by the multiplicity 1 structure $(\{1\}, \{2\}, \ldots, \{n\})$. Covering multiplicity, therefore, is not a useful measure of complexity when considering structures consisting of many small supports. We will concentrate on the opposite situation, in which we attempt to keep the ranks of the structures bounded for large retinas by using larger and larger supports. One example of this kind of structure is the *fractional support structure*, in which each support is some fixed fraction of the entire retina:

Definition 2.1.4. Let M be a positive integer. A support structure H on a retina R is said to be a $1/M$*-fractional support structure* if each support in H has size n/M.

(If M does not divide n evenly we suppose each support to have size within $\pm\frac{1}{2}$ of this value.)

Strictly speaking, of course, a predicate f is defined only for a particular retina, so it makes no formal sense to speak of "computing f on large retinas". On the other hand, we can think of properties like "parity" and "connectedness" as defining entire families of predicates, one predicate for each retina. We can now define the covering multiplicity of such a predicate (family) in the context of fractional supports.

Definition 2.1.5. For any predicate f on R and positive integer M we define $\mu(f, M, R)$, the *covering multiplicity of f for $1/M$-fractional supports* to be the minimum μ of any $1/M$-fractional support structure which admits f.

Definition 2.1.6. A predicate family f is said to have *covering multiplicity at most B for all fractional supports*, $\mu(f; \text{frac}) \leq B$, if, for all M, $\mu(f, M, R)$ is uniformly bounded by B on large retinas. That is, for any M there should exist a bound n_M such that $\mu(f, M, R) \leq B$ for any retina R with $|R| > n_M$. If such a finite value B exists, we say that f has *finite covering multiplicity for fractional supports.*

2.2. *Examples*

To illustrate the above definitions, we compute the covering multiplicity for $1/M$-fractional supports for the local convexity predicate, using the computation outlined in Section 1.2. We noted that local convexity can be determined by examining each boundary point of the figure in question, and that this can be done with supports that do not overlap, except for small "interfaces" along the edges.

Suppose that the retina is a square h units on a side, $n = h^2$. Using supports of order n/M we divide R into M strips, each strip a rectangle of size h by h/M. We must also assign additional partial functions to examine the interfaces between the strips. (See Fig. 3.) Each interface is itself a strip of width 2, and thus one additional support (of order h^2/M) can contain $h/2M$ interfaces (each of size $2h$). Since there are $M-1$ interfaces in all, we need $2M(M-1)/h$ additional "interface supports".

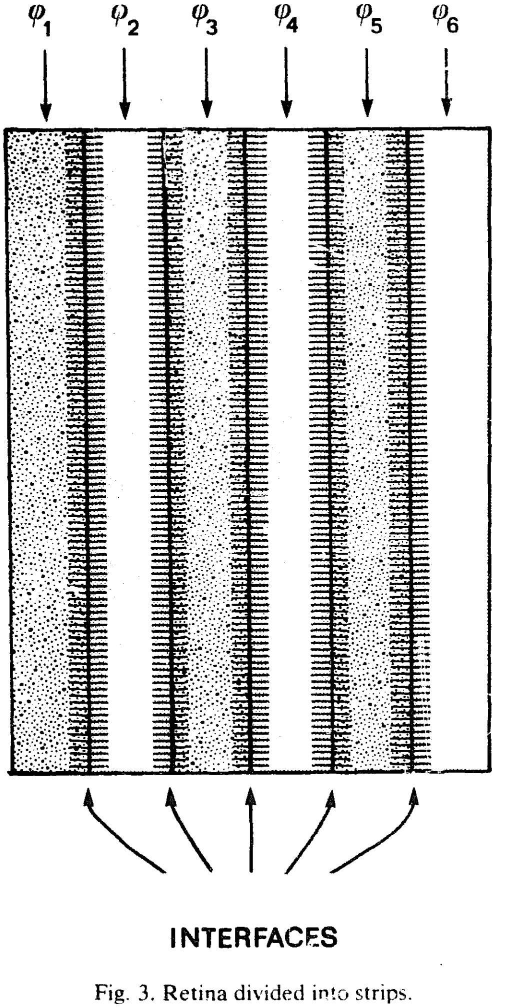


Fig. 3. Retina divided into strips.

So the entire structure for local convexity has

$$r = \text{rank} = \#(\text{strips}) + \#(\text{interface supports}) = M + 2M(M-1)/h;$$

$k = \text{order} = h^2/M$;

$$\mu = kr/n = 1 + 2(M-1)/h = 1 + 2(M-1)/\sqrt{n}.$$

Thus we have shown

Example 2.2.1. On a square retina R, predicate $f = \lceil X$ is locally convex$\rceil$ has

$$\mu(f, M, R) \leq 1 + 2(M-1)/\sqrt{n}.$$

Consequently, for any $\varepsilon > 0$, we have $\mu(f; \text{frac}) < 1 + \varepsilon$.

As a second example, consider the predicate $\lceil X$ is a single, solid rectangle$\rceil$. This can be computed by partial functions which "count the corners of X". That is, a figure X is a single solid rectangle if and only if its boundary contains precisely four "convex corners" (Fig. 4a), no "concave corners" (Fig. 4b) and no "double corners" (Fig. 4c). This computation can be performed with almost the same support structure as used above, i.e., by dividing the retina into disjoint strips together with "interface" strips of width 2. This time, however, we assign *three*

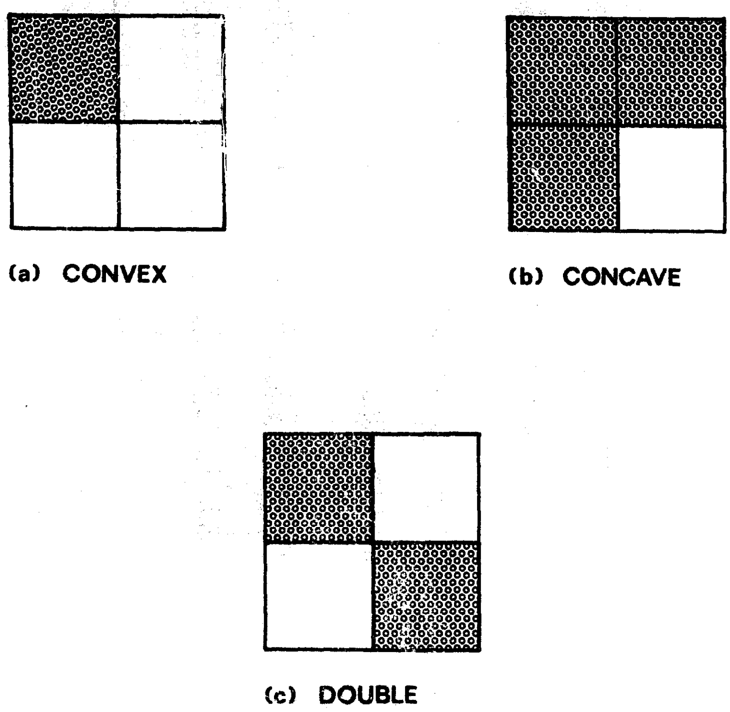


Fig. 4. Corner clusters for detecting rectangles.

partial functions, hence three supports, to each of the strips. In each support A, the three corresponding partial functions f_1, f_2, f_3 output 0 or 1 as follows:

(1) If $X \cap A$ contains concave corners, double corners, or more than four convex corners, then all f_i output 1.

(2) Otherwise the three functions output the number (from 0 to 4, counting in binary) of convex corners of $X \cap A$.

Using this information, the function g can determine whether or not X is a rectangle. The covering multiplicity here is three times as great as for local convexity:

Example 2.2.2. If f is the predicate $\lceil X$ is a single solid rectangle$\rceil$ then for any $\varepsilon > 0$, $\mu(f; \text{frac}) < 3 + \varepsilon$.

These examples illustrate predicates that can be computed with small covering multiplicity. On the other hand, we find that no fixed bound on covering multiplicity can suffice for computing arbitrary predicates on large retinas:

Proposition 2.2.3. *Let M be a positive integer, and ε a positive real number. Then any $1/M$-fractional support structure H which admits all predicates on a retina R must have*

$$\mu(H) > (n - \varepsilon)/M$$

provided r is sufficiently large with respect to M.

Proof. There are 2^{2^n} predicates on a retina of size n. Consider, on the other hand, the number of predicates admitted by support structures of rank r and order k. For each of the r partial predicates f_i there are $\binom{n}{k}$ ways of choosing support(f_i); and having selected a support, there are then 2^{2^n} functions f_i with that support. In addition, there are 2^{2^r} possibilities for g. So if the structure admits any function of n variables, we must have

$$2^{2^n} \leqslant \binom{n}{k}^r 2^{2^r + r2^k}.$$

Using the fact that $\binom{n}{k} < 2^n$, and taking logarithms, gives

$$2^n < rn + 2^r + r2^k$$

or, since $k = n/M$ and $r = n\mu/k = M\mu$

$$2^n \leqslant 2^{M\mu} + M\mu(n + 2^{n/M}). \quad \square$$

2.3. Highly interconnected computations

Covering multiplicity is a measure of the "globalness" of a computation, which can provide meaningful results about families of supports in which the order grows large along with the size of the retina. For arbitrary supports we can consider instead the question of how much "interconnection" among elements of the retina is required for a given computation.

Definition 2.3.1. If $H=(S_1,\ldots,S_r)$ is a support structure and $A\subset R$ define $\mathrm{Con}_H(A)\subset R$ to be the set

$$\mathrm{Con}_H(A)=\bigcup\{S_i\colon S_i\cap A\neq\emptyset\}.$$

Intuitively, imagine that H is "wired" by connecting together all pairs of elements in each support. Then Con(A) consists of all those points of R with "direct connections" to points of A. By way of analogy with 2.2.3, we will show that any computational scheme which can compute arbitrary Boolean functions must be "highly interconnected" in the sense that, for any $A\subset R$, either A must intersect many supports or else Con(A) must contain essentially all of R.

Definition 2.3.2. A family of support structures H on retinas R is said to be *highly interconnected* if, given any positive integer B and positive real number ε, one has that if n is sufficiently large with respect to B, then for any subset $A\subset R$ with $|A|>\varepsilon n$, either $\mathrm{cov}_H(A)>B$ or else $|\mathrm{Con}_H(A)|>(1-\varepsilon)n$. A predicate f is said to be *highly interconnected* if any family of support structures which admits f on large retinas must be highly interconnected.

Proposition 2.3.3. *Any family of support structures which admits arbitrary predicates must be highly interconnected. More precisely, if H is a support structure which admits arbitrary predicates on a retina R, then, for any $A\subset R$, either $\mathrm{cov}_H(A)\geq|A|$ or else*

$$|\mathrm{Con}_H(A)|>n-\log_2(\mathrm{cov}_H(A)).$$

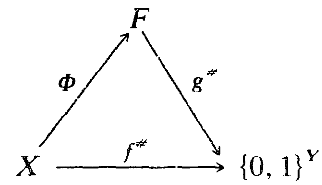
This will follow from another simple counting argument:

Lemma 2.3.4. *Given collections of Boolean variables $X=(x_1,\ldots,x_a)$ and $Y=(y_1,\ldots,y_b)$, let t be any integer such that $t<a$ and $\log_2 t<b$. Then there exist functions $f\colon X\times Y\to\{0,1\}$ which cannot be represented in the form*

$$f(X,Y)=g(f_1(X),\ldots,f_t(X),Y).$$

Proof. Functions $f\colon X\times Y\to\{0,1\}$ are equivalent to functions $f^{\#}$ from X into $\{0,1\}^Y$, the set of Boolean functions from Y into $\{0,1\}$, via the correspondence

$[f^{\#}(X)](Y)=f(X, Y)$. Representing f in the required form is equivalent to finding a factorization of $f^{\#}$:



where F is a set of t Boolean variables; $\Phi(X)=(f_1(X), \ldots, f_t(X))$ and $[g^{\#}(F)](Y)=g(F, Y)$. To demonstrate that not all functions factor in this way, we need only consider the sizes of the sets involved: An element of X can take on 2^a possible values, and there are 2^{2^b} possible values for an element of $\{0, Y\}^Y$. To construct a function $f^{\#}$ which does not factor, first enumerate the elements of X:

$$v_1, v_2, \ldots, v_{2^a}$$

and the elements of $\{0, 1\}^Y$:

$$\xi_1, \xi_2, \ldots, \xi_{2^{2^b}}.$$

There are two cases, depending on the relative sizes of a and b:

(1) If $2^b > a > t$ let $f^{\#}(v_i)=\xi_i$ for $i=1, \ldots, 2^a$. Then the image of $f^{\#}$ contains 2^a distinct elements, so that $f^{\#}$ cannot factor through F, which has size $2^t < 2^a$.

(2) If $a > 2^b > t$, let $f^{\#}(v_i)=\xi_i$ for $i=1, \ldots, 2^{2^b}$. Then the image of $f^{\#}$ contains 2^{2^b} distinct elements, so that $f^{\#}$ cannot factor through F, which has size $2^t < 2^{2^b}$. □

Proof of 2.3.3. For any $A \subset R$ define $R(A) \subset R$ to be the set $(R - \mathrm{Con}(A)) \cup A$, i.e., delete from R all variables which lie in $\mathrm{Con}(A)$ but not in A. The inclusion of $R(A)$ in R induces a surjection $\{0, 1\}^R \to \{0, 1\}^{R(A)}$. Set $t=\mathrm{cov}(A)$, X to be the variables in A, and Y to be the $n-|\mathrm{Con}(A)|$ variables in $R(A)-A$. Then any predicate on R which is admitted by H induces a predicate on $R(A)$ of the form $g(f_1(X), f_2(X), \ldots, f_t(X), Y)$. Now apply 2.3.4. □

In the case of $1/M$-fractional supports, the following combinatorial argument shows that interconnectedness implies high covering multiplicity:

Lemma 2.3.5. *If H is any family of highly interconnected $1/M$-fractional support structures then $\lim_{n\to\infty} \mu(H) \geq M$. Consequently, a highly interconnected predicate cannot have finite multiplicity for all fractional supports.*

Proof. First note that if $A \subset R$ with $\mathrm{cov}(A) \leq M-1$ then $|\mathrm{Con}(A)| \leq n(M-1)/M$. So if the family of support structures is highly interconnected, we can select any number ε and be sure that, for R large enough, we have $\mathrm{cov}(A) \leq M-1$ implies $|A| < \varepsilon n$.

For any of the retinas R in the family, let R_M be the subset of R consisting of points lying in fewer than M supports:

$$R_M = \{i \in R : \text{cov}(\{i\}) < M\}.$$

Note, in particular, that R_1 consists of all points with covering multiplicity 0. Then

$$\mu n = \sum_{i \in R_M} \text{cov}(\{i\}) + \sum_{i \in R - R_M} \text{cov}(\{i\}).$$

But

$$\sum_{i \in R_M} \text{cov}(\{i\}) \geq |R_M| - |R_1|$$

and

$$\sum_{i \in R - R_M} \text{cov}(\{i\}) \geq M(n - |R_M|).$$

Therefore, we have

$$\mu n \geq Mn - (M-1)|R_M| - |R_1|$$

or

$$\mu \geq M - (M-1)|R_M|/n - (|R_1|/n). \tag{1}$$

If $\sigma \subset \{S_j\}$ is some collection of the supports in H, define R_σ to be the subset of R consisting of those points which are contained in only those supports which lie in σ:

$$R_\sigma = \{i \in R : i \notin S_j \text{ for } S_j \notin \sigma\}.$$

Consider now the union of the R_σ over all sets σ of $M-1$ partial predicates. This union is precisely the subset of R consisting of points i for which $\text{cov}(\{i\}) \leq M-1$. That is,

$$R_M = \bigcup_{|\sigma| = M-1} R_\sigma. \tag{2}$$

If r is the rank of H then the number of sets σ containing precisely $M-1$ supports is equal to the binomial coefficient $\binom{r}{M-1}$. Notice that for a given σ, $\text{cov}(R_\sigma)$ cannot be greater than $|\sigma|$. Thus, according to the remark noted at the beginning of the proof, we have $|R_\sigma| < \varepsilon n$ on large retinas whenever $|\sigma| = M-1$. Eq. (2) therefore implies that

$$|R_M| \leq \binom{r}{M-1} \varepsilon n.$$

Combining this with $|R_1| < \varepsilon n$ and Eq. (1) yields:

$$\mu \geq M - (M-1)\varepsilon \binom{r}{M-1} - \varepsilon.$$

Since the order of H is equal to n/M, we have that $r=\mu M$. Substituting this into the above inequality and using the fact that $\binom{\mu M}{M-1}<2^{\mu M}$ gives

$$\mu > M-\varepsilon[(M-1)2^{\mu M}+1]. \quad (3)$$

Now, given any $\delta>0$, whoose ε small enough so that $\varepsilon[(M-1)2^{M^2}+1]<\delta$ and consider retinas with n large enough so that (3) holds. Suppose, for n this large, we had $\mu \leq M-\delta$. Then $2^{\mu M}<2^{M^2}$, so $\varepsilon[(M-1)2^{\mu M}+1]<\delta$, and

$$\mu+\varepsilon[(M-1)2^{\mu M}+1]<\mu+\delta\leq M-\delta+\delta\leq M,$$

which would contradict (3). Therefore we must have $\mu > M-\delta$. □

2.4. Interconnectedness of the connectivity predicate

Theorem 2.4.1. *The predicate* ⌈*X is connected*⌉ *is highly interconnected.*

The proof arises as a generalization of the simple observation that connectivity cannot be admitted by a support structure of rank 2 in which the retina is partitioned into disjoint halves, as can be seen immediately by considering the connectivity of the figures formed by the various combinations $S_i \cup T_j$ shown in Fig. 5. This construction has been generalized by Papert [9] to show that, if we allow only those structures in which no support intersects both the left and right halves of the retina, then the rank required for determining connectivity must grow arbitrarily large on large retinas:

Proposition 2.4.2. *Suppose H is a family of support structures which admits* ⌈*X is connected*⌉, *and that no support in H intersects both the left and right halves of the retina. Then, as n increases, the rank of H must grow at least as rapidly as* $\sqrt{n}$.

Proof. Consider the family of figures illustrated in Fig. 6, each consisting of a square, with m horizontal lines meeting the sides of the square at contact points $x_1, \ldots, x_m, y_1, \ldots, y_m$. Each pair of Boolean m-tuples $X=(x_1, \ldots, x_m)$ and $Y=(y_1, \ldots, y_m)$ gives a figure $F(X, Y)$; and it is easy to see that one of these figures is connected if and only if each horizontal line is connected to the square, either on the left or on the right. In other words, $F(X, Y)$ is connected if and only if $X \vee Y$. (Recall that $X \vee Y$, "X or Y", means that for each i we have $x_i \vee y_i$.)

Let $f_1, \ldots, f_a$ be those partial predicates whose support lies in the left half of the retina. Then, for any figure $F=F(X, Y)$, the a-tuple $f_L(F)=(f_1(F), \ldots, f_a(F))$ depends only on the m-tuple X, i.e., $f_L(F(X, Y))=f_L(X)$. We claim, therefore, that $a \geq m$. For, if not, then there are two distinct m-tuples X_1 and X_2 with $f_L(X_1)=f_L(X_2)$. But then, for any m-tuple Y, we have $F(X_1, Y)$ is connected if and only if $F(X_2, Y)$ is connected. On the other hand, taking $Y=\sim X_1$ and noting that

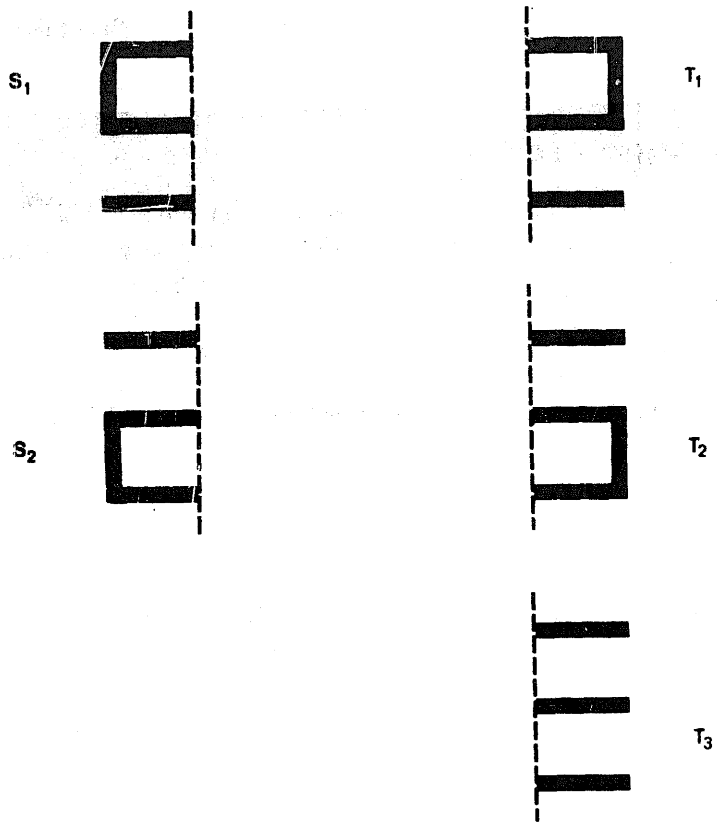


Fig. 5. A simple rank 2 structure cannot detect connectivity for the figures $S_i \cup T_j$.

$F(X_1, \sim X_1)$ is connected, shows that we must have $X_2 \vee \sim X_1$. Similarly, taking $Y = \sim X_2$ gives $X_1 \vee \sim X_2$. Therefore $X_1 = X_2$, and so $a \geq m$. Finally, observe that on a retina of size n, we can choose the number of horizontal strips to be proportional to $\sqrt{n}$. □

The next step in the proof of 2.4.1 is to extend the above construction so that the "contact points" x_i and y_i can be distributed throughout the retina:

Lemma 2.4.3. *For any integer m there is a constant* $K(m)$ *such that if H admits connectivity on a retina R, and* $A \subset R$ *with* $|A| > K\sqrt{n}$, *then either* $\text{cov}(A) > m$ *or else* $|\text{Con}(A)| > R - K\sqrt{n}$.

Proof. Let $R_1 \subset R$ be the subset of R consisting of all points which do not lie within distance $2m$ of the boundary of R. Choose points $x_1, \ldots, x_m$ lying in A such that (1) each x_i is contained in R_1; (2) for $i \neq j$, x_i and x_j do not lie within the same horizontal row of R, and, moreover, the horizontal rows containing x_i and x_j are at least six squares apart. (See Fig. 7.) Notice that we can do this so long as $|A|$ is

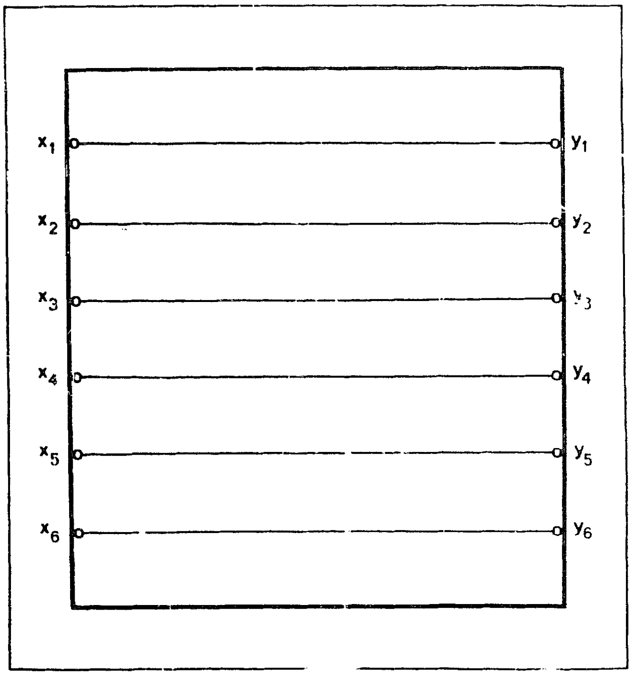


Fig. 6. Connectivity figures for Proposition 2.4.2.

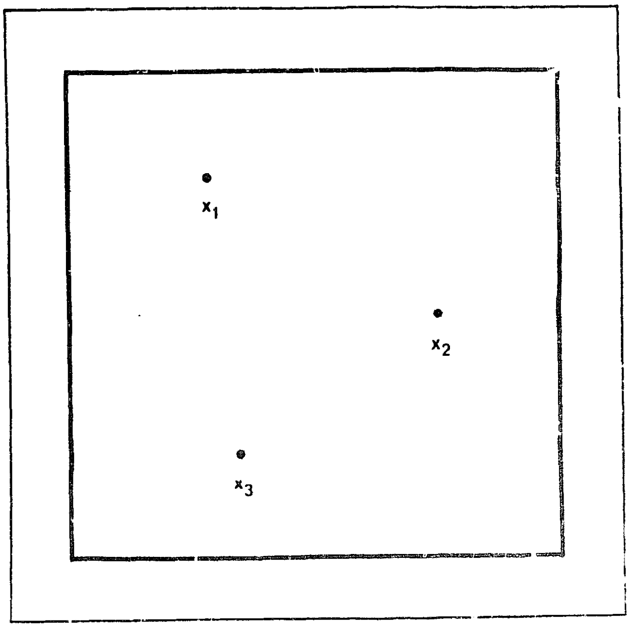


Fig. 7. Selecting the points x_i in Lemma 2.4.3.

greater than $|R_1|$ plus the size of $6m$ horizontal strips, i.e., $|A| > 14\sqrt{n}$. Next, choose points $y_1, \ldots, y_m$ in $R - \mathrm{Con}(A)$ such that (1) each y_i lies in R_1; (2) the horizontal row containing any y_j lies at least 6 units from the row containing any y_k $(j \neq k)$ or any x_i. We can do this so long as $|R - \mathrm{Con}(A)| > 18m\sqrt{n}$. To prove the Lemma we show that these x's and y's can be used as "contacts" in a family of figures equivalent to the figures of 2.4.2. This will imply at once that $\mathrm{cov}(A) \geqslant m$.

We construct the figures as follows: Begin by renumbering the x's from top to bottom, i.e., so that x_i is above x_j for $i < j$. Next arrange each x_i to be a contact point for three horizontal "wires" as shown in Fig. 8. Extend each of these wires on the left to meet the boundary of R_1. Now do the same thing for the y's, only this time working to the right. Next, in the boundary of R_1, connect the bottom wire of each x_i to the top wire of x_{i+1}, and similarly for the y's. Connect the top wires for x_1 and y_1, and the bottom wires for x_m and y_m. Finally, make a connection "around the bottom of R" to join the middle wire of x_i to the middle of the corresponding y_i, as follows: for (x_1, y_1) work in the boundary of R; for (x_2, y_2), in the strip 3 units in from the boundary; and so on. These latter connections all lie in $R - R_1$, and so do not interfere with the previous wires. (The final figure is illustrated in Fig. 9 for the case $m = 3$.) As in 2.4.2, these are connected if and only if $X \vee Y$. □

Proof of 2.4.1. Given M and ε, let K be the constant $K(M+1)$ given by Lemma 2.4.3, and choose R large enough so that $K\sqrt{n} < \varepsilon n$. Then, if $|A| > \varepsilon n$ and $|\mathrm{Con}(A)| < (1-\varepsilon)n$, we have that $|A|$ and $|R - \mathrm{Con}(A)$ are both greater than $K\sqrt{n}$. Hence $\mathrm{cov}(A) \geqslant M+1$ by 2.4.3, which shows that the computation is highly interconnected. □

2.5. Covering multiplicity of the connectivity predicate

Applying Lemma 2.3.5 to Theorem 2.4.1, we see immediately that connectivity, unlike the "local" predicates for local convexity and ⌈X is a single solid rectangle⌉

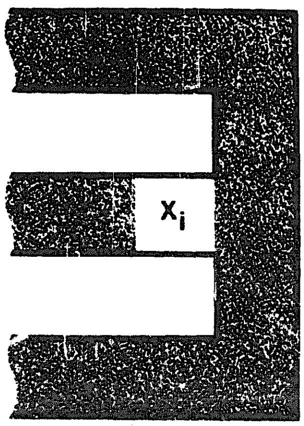


Fig. 8. Details of the contact x_i in Lemma 2.4.3.

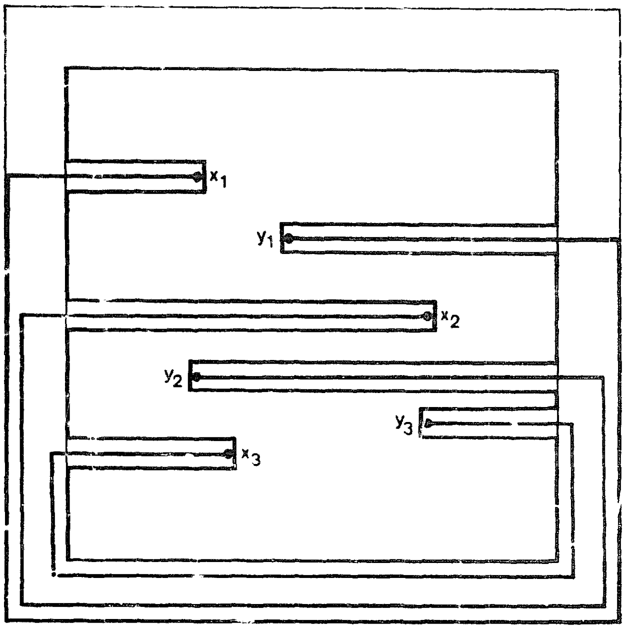


Fig. 9. Connectivity figure for Lemma 2.4.3.

discussed in Section 2.2, does not have finite covering multiplicity over all fractional supports. In particular, $1/M$-fractional support schemes which compute connectivity on large retinas must have $\lim_{n\to\infty} \mu \geq M$. But connectivity, we expect, should be "even more global" than that—more like computing arbitrary Boolean functions, where, for any fixed M, we have $\lim_{n\to\infty} \mu = \infty$ by 2.2.3. Indeed, the same reasoning as in 2.4.2. shows that this must be true in connectivity computations, and that μ must grow as rapidly as $\sqrt{n}$, so long as we assume that the supports of the partial predicates are disjoint. In this section we show that connectivity has unbounded covering multiplicity for $1/M$-fractional supports, even if the supports are allowed to overlap. We shall not, however, consider the case of arbitrary overlap. Rather, we restrict attention to schemes in which the supports overlap "regularly" according to the following prescription:

Definition 2.5.1. Let R be a square retina which is partitioned into D square blocks of equal size. Then a *d/D-regular support structure* is a structure in which each support consists of some fixed number $d < D$ of these blocks.

Note that this is a special case of d/D- fractional supports. Supports may overlap, but they must overlap "regularly"—two intersecting supports must necessarily have an entire block in common. We remark also that, although we phrase our results here only for square retinas, the same sort of thing will be true for all families of retinas with a sufficiently large "interior", i.e., for sequences of retinas R in which the perimeter of R grows no faster than $\sqrt{|R|}$.

Theorem 2.5.2. *For any family of d/D-regular support structures which computes connectivity, the covering multiplicity must satisfy* $\lim_{n\to\infty} \mu = \infty$.

The proof uses the switching network construction to translate the problem of determining connectivity into that of computing an arbitrary Boolean function. Recall [7] that if $X = (x_1, \ldots, x_a)$ is a set of Boolean variables, and $f: X \to \{0, 1\}$ is a predicate, then a *switching network for* f is given by a function F which associates to each Boolean a-tuple a figure on a retina R such that $F(x_1, \ldots, x_a)$ is connected if and only if $f(x_1, \ldots, x_a) = 1$. We can construct such a network by writing f in conjunctive normal form and translating the Boolean expression into a network, interpreting conjunction as series coupling and disjunction as parallel coupling. For example, the predicate

$$f(x_1, x_2, x_3) = \lceil \text{at least 2 of the variables are equal to 1} \rceil$$

has conjunctive normal form

$$(x_1 \vee x_2 \vee x_3) \wedge (\sim x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \sim x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee \sim x_3)$$

and the corresponding network is shown in Fig. 10. The figure is interpreted as follows: when the variable x_i is equal to 1, the squares marked "x_i" are filled in and the squares marked "$\sim x_i$" are left empty. Conversely, when $x_i = 0$, the squares marked "$\sim x_i$" are filled in and the squares marked "x_i" are left empty. These squares are then called the *contacts* of the network.

Notice that a function of a variables can have at most $a2^a$ terms in its conjunctive normal form, and hence a network with this many contacts can realize any Boolean function of a variables. This is the sense in which computations for determining connectivity must also be able to compute arbitrary Boolean functions.

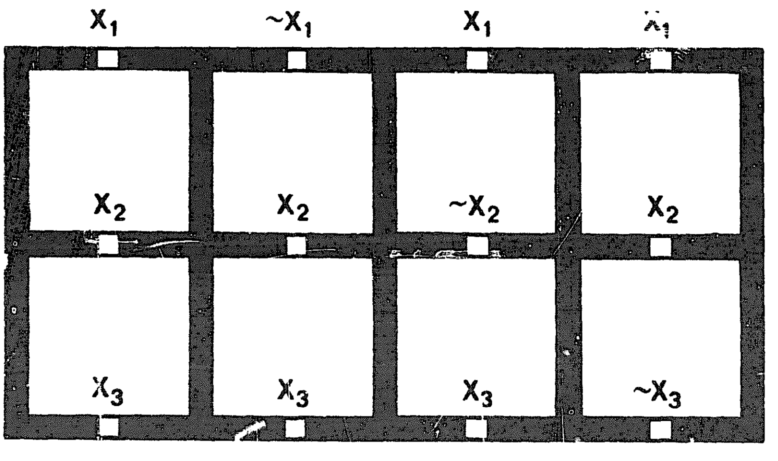


Fig. 10. Switching network for the function $(x_1 \vee x_2 \vee x_3) \wedge (\sim x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \sim x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee \sim x_3)$.

The proof of 2.5.2 now rests on the following:

Lemma 2.5.3. *Suppose we have D sets of Boolean variables*

$$X_1 = (x_{11}, \ldots, x_{1m}), \ldots, X_D = (x_{D1}, \ldots, x_{Dm}).$$

Consider support structures of order dm in which each support consists of some collection of $d < D$ of the X_i. Then for any fixed values of d and D, and any bound B, we can choose m large enough so that there are Boolean functions of the X's which cannot be computed by structures of this type having covering multiplicity $\mu < B$.

This lemma follows at once from 2.2.3. Moreover, by examining the proof of 2.2.3, we see that μ would have to grow large with the same order as m.

Proof of 2.5.2. Construct in R a switching network that can realize any function of Dm variables, such that the contacts corresponding to the variables in the i^{th} set X_i all lie within the i^{th} block. Then any regular support structure which admits connectivity on R must also admit arbitrary Boolean functions as in 2.5.3. As R becomes large we can choose larger and larger values for m. Therefore μ must also increase without bound. □

Remark 2.5.4. The above proof shows that μ must grow at least as rapidly as log n. This logarithmic factor arises from the use of the general switching network to realize arbitrary Boolean functions. It is natural to ask if there is a bound more in keeping with the $\sqrt{n}$ growth observed for disjoint supports. Also, is there some way to eliminate the "regularity" assumption, and so establish 2.5.2 for arbitrary $1/M$-fractional supports? What is the order of growth in this case?

3. Local and global in real-valued computations

The techniques presented above for analyzing Boolean functions are also applicable to the study of real-valued functions. The "retina" in this setting is an index set for n real variables $(x_1, \ldots, x_n)$, and the support of a function $f : \mathbb{R}^n \to \mathbb{R}$ consists of all indices i such that f depends on x_i. This extension to the real-valued domain is analogous to Uesaka's extension of Minsky and Papert's work on the theory of "analog perceptrons" [10, 11]. The analog perceptron formulation, however, deals only with functions of the form $f(x_1, \ldots, x_n) = \Sigma_i f_i$ where each f_i is a function of (hopefully) fewer than n of the x's. In keeping with the comments in Section 1.2, we find that this linearity requirement is too restrictive to serve as a basis for a general study of structural complexity. For example, the function

$$\text{mult} : (x_1, \ldots, x_n) \to x_1 \cdots x_n$$

cannot be written as a sum of functions f_i of fewer than n of the x's [10] and is therefore "global" from the perceptron point of view; but allowing multiplication rather than merely summation, at the "output stage" would enable us to compute mult using any partition of the $(x_1, \ldots, x_n)$ as supports for the f_i. (This is exactly analogous to the comments on the parity predicate in Section 1.2.)

In the sections below we shall work in the category of real-valued differentiable functions. From this point of view we analyze, in Section 3.2, the covering multiplicity necessary compute arbitrary polynomials. Section 3.3 applies the same ideas to computing the determinant of a matrix and solving systems of linear equations. All of these results are based on a theorem on functional decomposition proved in Section 3.1.

3.1. The decomposition theorem

Following the framework of Section 2 we will say that a support structure $(S_1, \ldots, S_r)$ *admits differentiably* a function $f: \mathbf{R}^n \to \mathbf{R}$ if f can be represented in the form $f = g(f_1, \ldots, f_r)$ where support$(f_i) \subset S_i$ and $g, f_1, \ldots, f_r$ are differentiable. We will also consider situations in which f is defined only locally in some neighborhood $U \subset \mathbf{R}^n$ in which case we require that g and the f_i be locally defined. To investigate conditions under which this can be done, consider first a simple kind of support structure in which the variables x_i are partitioned into two disjoint sets $X = (x_1, \ldots, x_a)$ and $Y = (x_{a+1}, \ldots, x_n)$. Let Diff (X) denote the algebra of real-valued differentiable functions of the variables in X. For any $f: \mathbf{R}^n \to \mathbf{R}$ let $\Delta(X, f)$ denote the module generated over Diff(X) by the a functions $\partial f/\partial x_1, \ldots, \partial f/\partial x_a$.

Theorem 3.1.1 (The decomposition theorem). *If $f: U \subset \mathbf{R}^a \times \mathbf{R}^{n-a} \to \mathbf{R}$ is a differentiable function which can be represented as*

$$f(X, Y) = g(f_1(X), \ldots, f_r(X), Y),$$

where $g: \mathbf{R}^{r+(n-a)} \to \mathbf{R}$ and *$f_i: \mathbf{R}^a \to \mathbf{R}$, $i = 1, \ldots, r$ are differentiable, then the module $\Delta(X, f)$ has rank at most r throughout U. Conversely, if f is continuously differentiable, and $\Delta(X, f)$ has rank at most r, then there is an open subset $V \subset U$ on which f can be so represented.*

Proof. Differentiating the equation $f = g(f_1, \ldots, f_r, Y)$ with respect to any $x_i \in X$ gives

$$\partial f/\partial x_i = \sum_j (\partial g/\partial f_j)(\partial f_j/\partial x_i).$$

Each f_j, and therefore each $\partial f_j/\partial x_i$, lies in Diff$(X)$. Hence $\partial f/\partial x_i$ lies in the module generated over Diff(X) by the r functions $\partial g/\partial x_j$. Since this is true for all i, we have that all of $\Delta(X, f)$ is contained within this module and is therefore of rank at most r.

To show that the condition is also sufficient, begin by choosing any $p = (p_X, p_Y) \in U$ and $r+1$ sets of particular values $Y_1, \ldots, Y_{r+1}$ for the variables Y such that the pairs (p_X, p_Y) all lie in U, and consider the functions $f_i(X) = f(X, Y_i)$. We claim that for any values Y_i the matrix

$$\|\partial f_i/\partial x_j\|, \quad i=1,\ldots,r+1, \quad j=1,\ldots,a$$

has rank at most r.

To prove this, pick any $r+1$ columns, say, for notational simplicity, columns 1 through $r+1$. According to the hypothesis on $\Delta(X,f)$ we have

$$\partial f(X,Y)/\partial x_{r+1} = \sum_{j=1,\ldots,r} \varphi_{r+1,j}(X)(\partial f(X,Y)/\partial x_j),$$

where the $\varphi_{r+1,j}(X)$ are functions of X alone. Thus we can substitute any Y_i for Y in the above equation to get

$$\partial f_i/\partial x_{r+1} = \sum_j \varphi_{r+1,j}(X)(\partial f_i/\partial x_j).$$

Since the $\varphi_{r+1,j}(X)$ are independent of i this shows that

$$((r+1)\text{st column of matrix}) = \sum_j \varphi_{r+1,j}(X)(j\text{th column of matrix})$$

which proves the claim.

Now to complete the proof of the theorem, choose $Y_1,\ldots,Y_{r+1}$ to maximize the rank of $\|\partial f_i/\partial x_j\|$. Denoting this maximum rank by m, $m \leq r$ by the claim. Without loss of generality we may assume that the first m rows of the matrix are linearly independent. Consider now the function

$$F: U \to \mathbf{R}^{m+1} \times \mathbf{R}^{n-a},$$

where

$$F(X,Y) = (f(X,Y), f_1(X),\ldots,f_m(X), Y).$$

By choice of the $Y_1,\ldots,Y_{r+1}$ we have that the rank of F is at most $m+n-a$. Therefore the functions describing the image of F are functionally dependent [1]. (This is where one needs that f is continuously differentiable.) Moreover, the set of points at which the $f_1,\ldots,f_m$ have maximal rank forms an open set $V \subset U$, and at any such point $p=(p_X, p_Y) \in V$ we can apply the implicit function theorem to solve the functional dependence relation for $f(X,Y)$, i.e., there exists a continuously differentiable function g defined in a neighborhood of p_X such that

$$f(X,Y) = g(f_1(X),\ldots,f_m(X), Y). \quad \square$$

Note: In the case $r=1$ this theorem reduces to a result of Leontief [4].

To relate 3.1.1 to the framework of Section 2, we define Diff(B) for any $B \subset \{1,\ldots,n\}$ to be the algebra of differentiable functions in all the variables $(x_i)_{i\in B}$.

Corollary 3.1.2. *Suppose H is a support structure which admits differentiably the function $f:\mathbf{R}^n \to \mathbf{R}$, and $A \subset \{1, \dots, n\}$. Then among the derivatives $(\partial f/\partial x_i)_{i\in A}$, at most $\mathrm{cov}_H(A)$ are linearly independent over* $\mathrm{Diff}(\mathrm{Con}_H(A))$.

Proof. Proceed just as above, noting that, for any $i \in A$

$$\partial f/\partial x_i = \sum_j (\partial g/\partial f_j)(\partial f_j/\partial x_i),$$

where the summation ranges over $j \in \mathrm{cov}(\{i\})$; and that for any such j

$$\partial f_j/\partial x_i \in \mathrm{Diff}(\mathrm{Con}(\{i\})) \subset \mathrm{Diff}(\mathrm{Con}(A)). \quad \square$$

3.2. Polynomial evaluation

We now apply the results of 3.1 to prove that support structures which admit arbitrary polynomials must be highly interconnected, and of infinite covering multiplicity for fractional supports:

Proposition 3.2.1. *If H is a support structure which admits differentiably the polynomial $P(x_1, \dots, x_m) = \sum_{i,j} x_j x_i^j$ then for any $A \subset \{1, \dots, n\}$, either $\mathrm{cov}(A) \geq |A|$ or else $\mathrm{Con}(A)$ is the entire index set $\{1, \dots, n\}$.*

Proof. Suppose that there is some index i not contained in $\mathrm{Con}(A)$. By renumbering the x's we may assume that this index is n, and that $A = \{1, \dots, |A|\}$. (This renumbering does not change P, which is symmetric in the x_i.) Then $\mathrm{Diff}(\mathrm{Con}(A)) \subset \mathrm{Diff}(x_1, \dots, x_{n-1})$. Moreover

$$P(x_1, \dots, x_n) = \sum_{i,j} x_j x_i^j = \sum_j x_j x_n^j + \sum_{i \neq n} \sum_j x_j x_i^j$$

$$= x_n^n + \sum_{j\neq n} x_j x_n^j + \sum_{i \neq n} x_n x_i^n, \quad \mathrm{mod}\ \mathrm{Diff}(x_1, \dots, x_{n-1})$$

and so, taking the derivative with respect to x_k $(k \neq n)$ gives

$$\partial P/\partial x_k = x_n^k + n x_n x_k^{n-1}, \quad \mathrm{mod}\ \mathrm{Diff}(x_1, \dots, x_{n-1}).$$

As the reader can verify, this equation implies that the module generated by $(\partial P/\partial x_k)_{k\in A}$ contains the $|A|$ elements $x_n, x_n^2, \dots, x_n^{|A|}$, which are linearly independent over $\mathrm{Diff}(x_1, \dots, x_{n-1})$ and hence over $\mathrm{Diff}(\mathrm{Con}(A))$. Therefore, by 3.1.2 we have $\mathrm{cov}(A) \geq |A|$. $\square$

Corollary 3.2.2. *Any family of support structures which differentiably admits arbitrary polynomials must be highly interconnected. Consequently in any family of such structures with $1/M$-fractional supports the convering multiplicity satisfies $\lim_{n\to\infty} \mu \geq M$.*

This follows at once from 3.2.1 and 2.3.5. $\square$

3.3. *Matrices and linear equations*

We next turn from the computation of arbitrary polynomials to consider the specific polynomial which expresses the determinant of a matrix in terms of the matrix elements. The retina here is an index set of n^2 real variables $(x_{11}, \ldots, x_{nn})$ which we view as a matrix $X = \|x_{ij}\|$. The function $\det : (x_{ij}) \to \det(X)$ is a polynomial of degree n. Expanding det by the ith row shows that

$$(\partial/\partial x_{ij}) \det = \pm X^{ij},$$

where X_{ij} is the cofactor of x_{ij}, itself a determinant of order $n-1$.

Lemma 3.3.1. *Let $S = (s_1, \ldots, s_n)$ be any collection of n of the variables x_{ij}. (That is, each s_k represents some x_{ij}.) Let X^k be the cofactor of s_k. Then the cofactors $X^1, \ldots, X^n$ are linearly independent over* $\mathrm{Diff}(s_1, \ldots, s_n)$.

Proof. By permuting the rows and columns of X and reordering the s_k's we may assume that $s_1 = x_{11}$. The cofactors $X^1, \ldots, X^k$ are sums of monomials of degree $n-1$. The key to the lemma is the claim that, in the expansion of X^{11} there is at least one monomial which does not contain any of the x's lying in S.

To prove this claim we note that, since X^{11} is itself a determinant of order $n-1$, the monomials α in the expansion for X^{11} are the products $\alpha = \alpha_2 \cdots \alpha_n$ where α_j lies in the jth column of X and no two α_j's lie in the same row. So we must show that there is at least one such set of α_j's, none of which lie in S. Let c_j be the number of elements of S lying in the jth column of X. By permuting the columns, we may assume that $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_n$. There are two cases to consider;

Case 1: $c_1 > 1$. We want to choose α_2 from the second column. There are $n-1-c_2$ possible choices. Likewise, in choosing α_3 from the third column, and not lying in the same row as α_2, there are $n-2-c_3$ choices. In general, there are $n-(j-1)-c_j$ choices for α_j. We need to know, then, that

$$n > c_j + (j-1), \quad j = 2, \ldots, n \tag{4}$$

in order to be sure that there are choices possible at each step. But $\Sigma c_j = n$ and $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_n$ so that $c_j \leqslant n/j$. Also, if $c_1 > 1$ we can be sure that $c_j = 0$ for $j > n/2$. So we need only verify (4) for $j < n$:

$$c_j + (j-1) < (n/j) + (j-1) = [(j-n)(j-1)/j] + n < n.$$

Therefore we can construct α as required.

Case 2: $c_1 = 1$. In this case each column of X contains one element of S. Now, if some row contains more than one element of S we can prove the result by applying Case 1 to the transpose of X (which has the same determinant as X). Otherwise, if each row and column contains one element of S, we may permute rows and columns so that the s_k are precisely the diagonal elements of X. If so we can take the monomial α to be, e.g., $X_{23}X_{34} \cdots X_{n,n-1}X_{n-1,1}$. This proves the claim.

Now to complete the proof of the lemma, let $F = \mathrm{Diff}(s_1, \ldots, s_n)$, and suppose that the X^k were linearly dependent over F, i.e.,

$$\sum_k f_k X^k = 0, \quad \text{where } f_k \in F.$$

Now let α be the monomial in X^1 whose existence is assured by the above claim, and set $X^1 = \alpha + \beta$. We would then have

$$f_1\alpha = -f_1\beta - \sum_{k=2,\ldots,n} f_k X^k. \tag{5}$$

Let $t_1, \ldots, t_{n^2-n}$ be those x_{ij} which do not lie in S and regard (5) as an equation in the polynomial ring $F[t_1, \ldots, t_{n^2-n}]$. The construction of α guarantees that, as an element of this ring, α has degree $n-1$. Also, the terms on the right side of (5) have degree at most $n-1$. Moreover, none of these terms can be equal to α, since a monomial in the expansion of the cofactor of an element of a matrix cannot also lie in the expansion of a cofactor of a different element of the matrix. Therefore Eq. (5) is impossible. □

As a consequence we deduce a theorem about the distributed computation of determinants: Suppose we have an $n \times n$ matrix X, whose entries we partition into n sets S_i of order n. (The S_i can be the rows, columns, $\sqrt{n} \times \sqrt{n}$ submatrices or whatever.) Consider computing $\det(X)$ by first evaluating separate "preprocessing" functions f_i of the S_i and then combining the results of the f_i by some function g. We find that computing a determinant is a worst case for this kind of distributed computation:

Theorem 3.3.2. *Let $H = (S_1, \ldots, S_n)$ be a support structure for functions of n^2 variables where the S_j are pairwise disjoint and of size n. If H differentiably admits the determinant function, then* $\mathrm{cov}(S_j) \geq n$ *for all j.*

Proof. Since the S_j are disjoint, we have $\mathrm{Con}(S_j) = S_j$ and the theorem follows immediately from 3.3.1 and 3.1.2. □

In other words, no matter what "local" computations we make based on the n elements of each S_i, we still must transmit n numbers to be combined by g. From the communication point of view, there is no point in doing any local computations at all. We may as well transmit the n elements of S_i to g directly.

As a final application of these ideas, consider the problem of solving an $n \times n$ system of linear equations $Xy = b$ based on dividing the matrix X into columns. That is to say, each of the elements y_j should be computed by a structure in which each support can intersect at most one column of X (and we'll also allow each partial function to access any entry of b). As with determinants, this turns out to be a worst case: the information needed from each column of X cannot be transmitted by fewer than n differentiable functions:

Theorem 3.3.3. *Let b be a fixed non-zero n-vector, and let $y_j(X)$ be the jth entry of the solution to the system of linear equations $Xy = b$. Suppose that H is a support structure which differentiably admits y_j, and that each support in H is contained within a single column of X. Then, for the jth column X_j of X, we have* $\text{cov}(X_j) \geq n$.

Proof. Let b^jX denote the "augmented matrix" in which the jth column of X is replaced by b. By Cramer's rule, $y_j = |b^jX|/|X|$. Hence, for any element x_{ik}

$$\partial y_j/\partial x_{ik} = 1/|X|^2[(|X|)(\partial|b^jX|/\partial x_{ik}) - (|b^jX|)(\partial|X|/\partial x_{ik})].$$

For any j, if b is non-zero, there are matrices X for which $|X| \neq 0$ and $|b^jX| \neq 0$. So consider the elements in the jth column of X, i.e., take $j = k$ in the above equation. Since $|b^jX|$ does not involve the variable x_{ij}, $\partial|b^jX|/\partial x_{ij} = 0$ and so

$$\partial y_i/\partial x_{ij} = -(|b^jX|/|X|^2)\partial|X|/\partial x_{ij} = \pm(|b^jX|/|X|^2)X^{ij}.$$

But, by Lemma 3.3.1 the $X^{1j}, X^{2j}, \ldots, X^{nj}$ are linearly independent over $\text{Diff}(x_{1j}, \ldots, x_{nj})$. Hence the multiples of these polynomials by $|b^jX|/|X|^2$ are also linearly independent. Corollary 3.1.2 therefore implies that $\text{cov}(X_j) \geq n$. □

3.4. *Relations with Hilbert's 13th problem*

The preceding perspective on the complexity of functions is related to investigations growing out of Hilbert's 13th problem, which concerns the possibility of representing functions of several variables as superpositions of functions of a smaller number of variables. (For a survey see [5].) Notable among these is the result due to Kolmogorov and Arnol'd [3] that any continuous function can always be expressed as a superposition of continuous functions of two variables. On the other hand, it is known that this cannot be done if the functions in the decomposition are required to satisfy differentiability constraints. This qualitative difference between differentiable and non-differentiable decompositions suggests there is no straightforward extension of the techniques of the previous sections to allow for non-differentiable partial functions.

4. Conclusion; Questions for further research

This paper has suggested precise computational formulations which interpret such vague notions as "global", "gestalt" or "the difficulty of dividing a computation into independent simpler parts". The goal is to develop meaningful measures of complexity which reflect only how the pieces of a computation are interrelated.

and are independent of the specific operations performed by each piece. Hopefully, such an effort could lead to a unified perspective for discussing problems of "parts and wholes" in computational geometry, in numerical computation, in distributed data processing, and perhaps even in artificial intelligence and cognitive theory as indicated in [8].

The theory is still at an embryonic stage, and the reader will no doubt recognize numerous ways in which the above results can be improved. What can one say about the order of growth of covering multiplicity for connectivity or other geometric predicates? Are there general techniques for establishing lower bounds for the covering multiplicity of specific Boolean functions? The decomposition theorem of Section 3.1 is a differentiable analogue of results of Ashenhurst on "disjoint decompositions" in switching theory [2]. Can this be extended to more general decompositions? We have only hinted at applications to the study of distributed data bases. Developing covering multiplicity criteria in this context is surely a major area left untouched by the present investigation. Other extensions of the theory could deal with "continuous retinas", in which the subsets $S \subset R$ become measures defined on the plane, and the predicates f become functions on the Hilbert space of measures. It is also important to develop alternative measures of "local and global" complexity and contrast these with covering multiplicity. For example, there should be a whole spectrum of "interconnectedness" running from non-overlapping support structures to the highly interconnected structures defined in Section 2.3, and one should be able to measure precisely the "intrinsic interconnectedness" required for computing a function. Certainly, much remains to be done in this area.

Acknowledgments

Marvin Minsky and Seymour Papert provided support and encouragement for developing the above ideas. In particular, Papert's note [9] inspired proofs of the connectivity results which are much simpler than those appearing in an earlier draft of this paper. I would like to express special appreciation to Jan Mycielski for his many helpful suggestions, including use of the term "covering multiplicity" and phrasing the above results in terms of support structures; and to Suzin Jabari of the MIT Artificial Intelligence Laboratory for preparing the illustrations for the manuscript.

References

[1] R.C. Buck, *Advanced Calculus* (McGraw-Hill, New York, 1956).
[2] H.A. Curtis, *A New Approach to the Design of Switching Circuits* (Van Nostrand, New York, 1962).

[3] A.N. Kolmogorov, On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables, *Dokl. Akad. Nauk* **114** (1957) 953–956 (in Russian); English transl.: *Am. Math. Soc. Transl.* (2) **17** (1961) 369–373.
[4] W. Leontief, A note on the interrelation of subsets of independent variables of a continuous function with continuous first derivatives, *Bull. Am. Math. Soc.* **53** (1947) 343–350.
[5] G.G. Lorentz, The 13th problem of Hilbert, in: *AMS Proc. Symposia in Pure Mathematics* **28** (1976) 419–430.
[6] D. Marr and T. Poggio, Cooperative computation of stereo disparity, *Science* **194** (October 1976).
[7] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry* (MIT Press, Cambridge, MA, 1969).
[8] M. Minsky and S. Papert, *Artificial Intelligence, Condon Lectures, Oregon State System of Higher Education* (Oregon Univ. Press, 1974).
[9] S. Papert, On an extension of one of Abelson's extensions of perceptrons, unpublished note (1977).
[10] Y. Uesaka, Analog perceptrons: On the additive representations of functions, *Information and Control* **19** (1971) 41–65.
[11] Y. Uesaka, Analog perceptron: its decomposition and order, *Information and Control* **27** (1975) 199–217.