4.1. Write down all the linear threshold functions $L(\Phi)$ and the corresponding Boolean functions for the following case, where $\mathbf{\Phi}$ is the family of predicates defined on the two-variable space **R** (*Hint:* There are 16 Boolean functions of two variables):

$$R = \{x, y\}, \qquad \Phi = \{\phi_1 = x,\ \phi_2 = y\}.$$

Does $L(\Phi)$ include all the Boolean functions of x and y? If not, what functions are not included?

4.2. (*a*) Consider a rectangular grid. The coordinates of the grid points, ordered lexicographically, by a row-by-row scan from left to right and bottom to top are (0,0), (0,1), ..., $(0, M)$, (1,0), (1,1), ..., $(1, M)$, ..., $(N, 0)$, $(N, 1)$, ..., (N, M). The edges of the grid are traversed either horizontally or vertically. Let $T(N, M)$ denote the number of shortest paths from (0,0) to (N, M). Then justify that $T(N, M)$ is the solution to the two-dimensional difference equation,

$$T(N, M) = T(N, M-1) + T(N-1, M),$$

with boundary conditions

$$T(0, M) = 1 \qquad \text{and} \qquad T(N, 0) = 1.$$

Subsequently, show that the solution for $T(N, M)$ is

$$T(N, M) = \binom{N+M}{M} = \binom{N+M}{N}.$$

(*b*) The two-dimensional difference equation,

$$L(N, M) = L(N-1, M) + L(N-1, M-1),$$

with boundary conditions

$$L(1, M) = 2 \qquad \text{and} \qquad L(N, 1) = 2N,$$

occurs in the solution of $L(N, M)$ given in Eq. (4.3). Verify the solution by any method you know.

4.3. Given is a perceptron with analog nonlinear preprocessing units as shown in Fig. 4.5. What types of decision surfaces can be realized? What is the capacity of the machine?

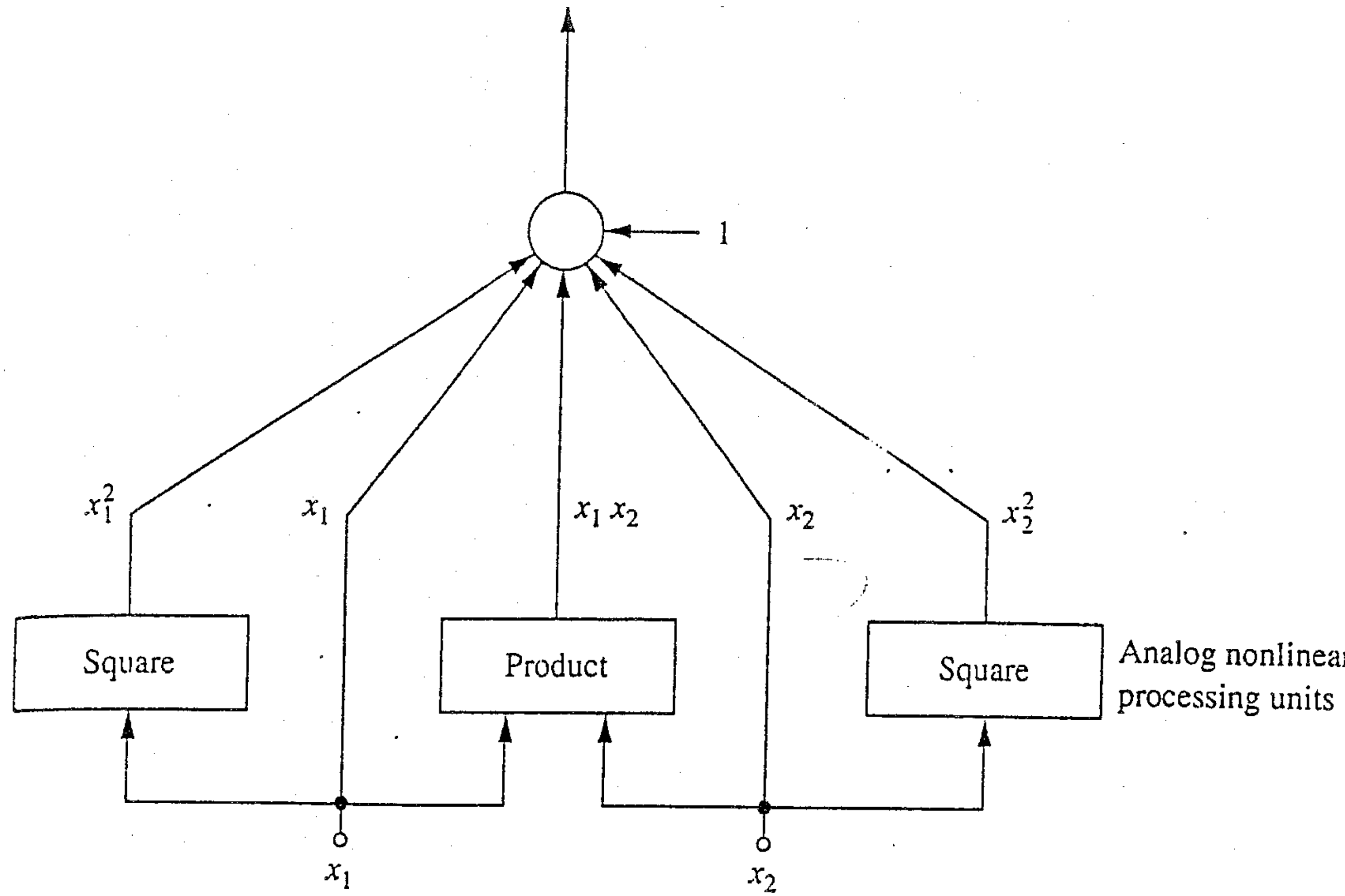

FIGURE 4.5
Analog network for realizing any quadratic decision surface in the two input variables, x_1 and x_2. The connection weights are not shown.

4.4. Design a perceptron with one layer of analog preprocessing units to realize any circular decision surfaces. What is the capacity of the machine?

4.5. Before applying the perceptron learning algorithm, a pattern vector of the form $(y_1\ y_2\ \cdots\ y_M)^T$ is extended by appending a 1 so that the augmented vector is $(y_1\ y_2\ \cdots\ y_M\ 1)^T$. Why is this done? Can a different constant be used instead of 1, for example, 0, 5, -10? In case a nonzero constant different from 1 is used, then after convergence, what is the equivalent threshold of the TLU without the connection to the input set at this constant value? Can different constants be used for different patterns?

4.6. Apply the perceptron learning algorithm to classify the following three-dimensional unipolar binary patterns before augmentation:

$$\text{Class A: } \{x\} = \{(0,0,0),\ (1,1,1)\}$$
$$\text{Class B: } \{x\} = \{(0,0,1),\ (0,1,1)\}$$

Draw a figure of the perceptron obtained, with its connections, weights, threshold, an transfer characteristic specified.

4.7. Apply the absolute correction rule to the following unaugmented unipolar binary pat terns. After augmentation and adjustment, a ternary set results.

$$C_+: \{(0,0),\ (0,1)\}, \qquad C_-: \{(1,0),\ (1,1)\}.$$

4.8. Apply the fixed-increment rule with $c = 1$ to the following three-dimensional unipola binary patterns before augmentation:

$$C_+: \{(0,0,0),\quad (1,0,0),\quad (1,0,1),\quad (1,1,0)\}$$

$$C_-: \{(0,0,1),\quad (0,1,1),\quad (0,1,0),\quad (1,1,1)\}$$

Let $\mathbf{w}_0 = (-1\ \ -2\ \ -2\ 0)^T$ denote the initial weight vector.

4.9. Show that the pseudoinverse solution $\mathbf{w}^* = S^+\mathbf{d}$ minimizes the mean squared error

$$e^2 = \frac{1}{N}\|\mathbf{d} - S\mathbf{w}\|^2 = \frac{1}{N}(\mathbf{d} - S\mathbf{w})^T(\mathbf{d} - S\mathbf{w})$$

in the case when all entries are real-valued.

4.10. (*a*) Use the general gradient descent update equation

$$\mathbf{w}(k+1) = \mathbf{w}(k) - c\left\{\frac{\partial E(\mathbf{w}, \mathbf{x})}{\partial \mathbf{w}}\right\}_{\mathbf{w}=\mathbf{w}(k)}$$

and the error function

$$E(\mathbf{w}, \mathbf{y}, T) = \frac{1}{8\|\mathbf{y}\|^2}\left[(\mathbf{w}^T\mathbf{y} - T) - |\mathbf{w}^T\mathbf{y} - T|\right]^2,$$

where $T > 0$, to derive a perceptron error correction algorithm.

(*b*) Let $c = T = 1$. Apply the algorithm obtained in (*a*) to the patterns in Problem 4.7.

(*c*) Discuss the effects of increasing T on the convergence of the algorithm for linearly separable patterns.

4.11. Consider an augmented pattern vector $\mathbf{y}(k) = (y_1(k) \cdots y_n(k)\ 1)^T$ and a weight vector $\mathbf{w} = (w_1 \cdots w_n\ w_{n+1})^T$. Assume that the augmented pattern vectors belonging to class C_- have been multiplied by -1. Call the totality of patterns after this modification the resulting set or the adjusted augmented pattern set.

(*a*) Show that if the patterns are linearly separable, then a solution $\hat{\mathbf{w}}$ exists such that $\hat{\mathbf{w}}^T\mathbf{y} > T$ for each pattern vector $\mathbf{y}$ in the resulting set where T is a nonnegative real number. What is the geometric interpretation in the weight space?

(*b*) The normal distance from an extended pattern vector $\mathbf{y}(k)$ to a decision hyperplane defined by $\mathbf{w}^T\mathbf{y}(k) = 0$ is

$$d = \frac{|\mathbf{w}^T\mathbf{y}(k)|}{\|\mathbf{w}\|}$$

Let

$$\alpha = \max_{\mathbf{w}} \min_{\{\mathbf{y}(k)\}_{k=1}^{N}} \frac{|\mathbf{w}^T\mathbf{y}(k)|}{\|\mathbf{w}\|}$$

where N is the number of patterns. Then the condition

$$\frac{|\mathbf{w}^{\mathrm{T}}\mathbf{y}(k)|}{\|\mathbf{w}\|} \geq \alpha$$

gives the optimal decision hyperplane in the sense that the minimum distance from all the extended training patterns to it is the largest possible.

Find the minimum distance from all training patterns to the decision hyperplane obtained after convergence of the following modified fixed-increment rule:

$$\mathbf{w}(k+1) = \begin{cases} \mathbf{w}(k) + \mathbf{y}(k) & \text{if } \mathbf{w}^{\mathrm{T}}\mathbf{y}(k) \leq T \\ \mathbf{w}(k) & \text{if } \mathbf{w}^{\mathrm{T}}\mathbf{y}(k) > T. \end{cases}$$

4.12. You are given the following two classes of patterns, before augmentation, with only one pattern in each class: $C_+ = \{(57.595722, -99.759033)\}$ and $C_- = \{(41.887859, -72.551994)\}$. Surprisingly, it may take many thousands of iterations for any of the perceptron error correction rules to converge to a solution. Augment the patterns in the standard way to $C_+^a = \{(57.595722, -99.759033, 1)\}$, and $C_-^a = \{(41.887859, -72.551994, 1)\}$.

(*a*) Write a computer program implementing any of the perceptron error correction rules to verify this.

(*b*) Provide an explanation as to why it takes so long to solve such a simple problem, and suggest a general method to speed up the convergence of the perceptron algorithms that will reduce the iterations required for convergence to well below 50 in this case.

(*c*) Repeat (*a*) and (*b*) using the Widrow-Hoff LMS algorithm.

(*d*) Generalize your method obtained in (*b*) to patterns of higher dimensions.

4.13. In many applications, some components of the training vectors are not specified, either because they are not available or because of too much noise in the measurement. These components are called *uncertain* components. In different patterns, the uncertain components may be different. Generalize the perceptron algorithms to patterns with uncertain components.

4.14. Derive Eq. (4.43) in text.

$$\nabla_k = \frac{\partial E(e_k^2)}{\partial \mathbf{w}(k)} = 2(\mathbf{Q}\mathbf{w}(k) - \mathbf{P}), \tag{4.43}$$

4.15. Consider the adaptive linear element (combiner) shown in Fig. P4.15. The kth augmented pattern vector and the weight vector are, respectively,

$$\mathbf{y}(k) = [y_1\ y_2\ \cdots\ y_n\ 1]^{\mathrm{T}} \qquad \text{and} \qquad \mathbf{w}(k) = [w_1\ w_2 \cdots w_n\ w_{n+1}]^{\mathrm{T}}.$$

The desired response to $\mathbf{y}(k)$ is $d(k)$, and the present error of the linear combiner is

$$e_k = d(k) - \mathbf{y}(k) \bullet \mathbf{w}(k).$$

(*a*) In the μ-LMS algorithm, an instantaneous gradient $\hat{\nabla}_k$ is calculated to be

$$\hat{\nabla}_k = \frac{\partial e_k^2}{\partial \mathbf{w}(k)}.$$

The weight update based on this estimate instead of the true gradient is

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \hat{\nabla}_k, \qquad \mu > 0.$$

Show that the weight update simplifies to

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu e_k \mathbf{y}(k).$$

(*b*) Comment on the use of the instantaneous gradient, the size of μ, and the step size corresponding to the period during which a small finite number of exemplars may be presented without appreciable change in the weight vector.

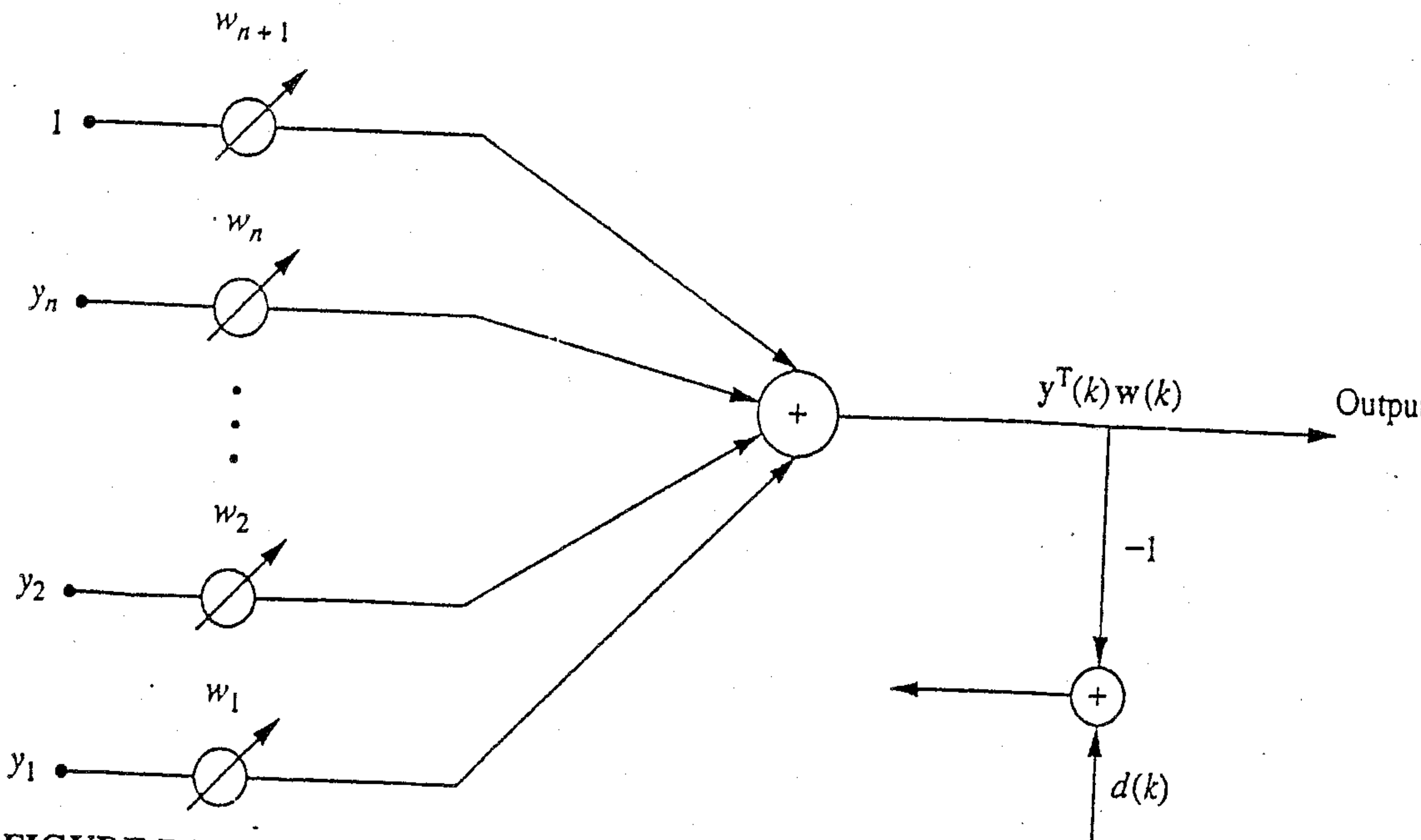

FIGURE P4.15

4.16. In a certain training scheme, the weights $w_1, w_2, \ldots, w_n$ were ordered as follows:

$$w_1 > w_2 > w_3 > \cdots > w_{n-1} > w_n > 0.$$

The elements $x_1, x_2, \ldots, x_n$ of a training pattern vector were also ordered as follows:

$$0 < x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n.$$

The weight w_i need not be associated with x_i. A neural network designer remembers only that it is possible to obtain the correct association from a particular one-to-one correspondence between the weight w_i and the signal $x_{p(i)}$ for $i = 1, 2, \ldots, n$, which maximizes the value of the expression

$$E_p = \sum_{i=1}^{n} w_i x_{p(i)}$$

over all possible permutations p of the integer index set $\{1\ 2\ \cdots\ n\}$. Help the neural network designer by identifying the expression that produces that maximum.

4.17. Consider two vectors:

$$\mathbf{v} = (v_1\ v_2\ \cdots\ v_m)$$

$$\mathbf{w} = (w_1\ w_2\ \cdots\ w_m).$$

It is specified that

$$v_1 > v_2 > \cdots > v_m > 0$$

$$0 > w_1 > w_2 > \cdots > w_m.$$

It is required to find a permutation p for $1, 2, \ldots, m$ corresponding to each of the following cases. Identify each permutation by writing the expression for the optimization desired.

(a) $S_1 = \sum v_i w_{p(i)}$ has to be a minimum.

(b) $S_2 = \sum v_i^2 w_{p(i)}^2$ has to be a maximum.

4.18. Suppose there are eight binary inputs $\mathbf{x} = (x_7 x_6 \cdots x_0)$, where $x_7 x_6 \cdots x_0$ can be considered as a binary number $\mathbf{x}$. Let $B(\mathbf{x})$ be the decimal representation of the binary number. For example, 35 is the decimal (radix 10) representation of the binary (radix 2) number 00100011. Design the simplest perceptron to compute the predicate function,

$$\Psi_1 = \begin{cases} 1 & B(\mathbf{x}) \geq 8 \\ 0 & \text{otherwise.} \end{cases}$$

What is the support of Ψ_1?

Let Ψ be a linear threshold function with respect to the family of predicates $\mathbf{\Phi}$ defined on the retina $\mathbf{R} = (x_7, x_6, \ldots, x_0)$. Then Ψ has the representation

$$\Psi(\mathbf{x}) = P_B\left[\sum_{\phi_i \in \Phi} \alpha_i \phi_i(\mathbf{x}) - \theta\right],$$

where $\mathbf{x}$ is a subset of $\mathbf{R}$, θ is a real number denoting threshold, Φ_i is a predicate that belongs to the family $\mathbf{\Phi}$, and

$$P_B[y] = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{if } y < 0. \end{cases}$$

What are the supports of the partial predicates ϕ_i? What is the order of the perceptron you have designed?

4.19. Repeat Problem 4.18 with Ψ_1 replaced by the predicate function,

$$\Psi_2 = \begin{cases} 1 & B(\mathbf{x}) \geq 8 \quad \text{and} \quad B(\mathbf{x}) \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

4.20. Consider the predicate function,

$$\Psi(\mathbf{x}) = P_B\left[a \sum_{x_i \in \mathrm{R}} x_i + b \sum_{x_i} \sum_{\substack{x_j \\ i<j}} x_i x_j - c\right],$$

where a, b, and c are real constants. What is the order of $\Psi(\mathbf{x})$? If $\Psi(\mathbf{x})$ is expressed in the form,

$$\Psi(\mathbf{x}) = P_B\left[\alpha_i \phi_i(\mathbf{x}) - \theta\right],$$

write α_i, $\phi_i(\mathbf{x})$, and θ in terms of a, b, c, and the x_i's. If a perceptron is to implement $\Psi(\mathbf{x})$, how many connections to the output threshold logic unit exist, given that the cardinality of the retina is $|\mathrm{R}|$?

4.21. What is the order of the predicate $\varphi(|\mathbf{x}|) = [(|\mathbf{x}| = M_1 \vee |\mathbf{x}| = M_2)]$, where $M_1 < M_2$ and M_1, M_2 are arbitrary positive integers?

4.22. In a *diameter-limited* perceptron, for each $\phi_i(\mathbf{x}) \in \Phi$, the set of points (defined by specializations of the set of variables $\{x_1, x_2, \ldots, x_n\}$ in $\mathbf{x}$) on which $\phi_i(\mathbf{x})$ depends is restricted so as not to exceed a certain *fixed diameter* (measured, say, according to the Euclidean metric) in the Euclidean space. Can a diameter-limited perceptron compute Ψ_{convex}? Why?

4.23. Answer whether each of the following statements is true or false.

(*a*) A diameter-limited perceptron can compute Ψ_{convex}.

(*b*) The order of the predicate $\Psi(|X| < M)$ is 1, where M is an arbitrary but fixed positive integer and $|X|$ denotes the number of points in the finite set X.

(*c*) The predicate $\Psi_{\text{connected}}$ cannot have an order greater than 3.

(*d*) The predicate Ψ_{convex} is of order 3.

(*e*) All Boolean functions of two variables have order 1.

(*f*) The conjunction

$$y_1 \wedge y_2 \wedge \cdots \wedge y_n \triangleq y_1\, y_2 \cdots y_n$$

of the Boolean variables $y_1, y_2, \ldots, y_n$ (each either assumes the value 1 or 0) is of order 1.

(*g*) The disjunction

$$y_1 \vee y_2 \vee \cdots \vee y_n \triangleq y_1 + y_2 + \cdots + y_n$$

of the Boolean variables $y_1, y_2, \ldots, y_n$ (each assumes either the value 1 or 0) is of order 1.

(*h*) The counting predicate that recognizes that a finite set X has M points is of order 4.

4.24. Each accompanying part has one correct answer. Identify this correct answer by circling the appropriate letter (capital A, B, C, etc.).

(a) The Widrow-Hoff learning law is obtained after calculating the gradient $\nabla F(\mathbf{w})$ of the expression

$$F(\mathbf{w}) = p - 2\mathbf{w}^{\mathrm{T}}\mathbf{q} + \mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w},$$

where p is a scalar, $\mathbf{w}$ and $\mathbf{q}$ are vectors, $\mathbf{R}$ is a matrix, and the superscript T denotes transpose operation. The gradient $\nabla F(\mathbf{w})$ is

A. $-2\mathbf{q} + \mathbf{R}\mathbf{w}$.
B. $-2\mathbf{q} + 2\mathbf{R}\mathbf{w}$.
C. $-\mathbf{q} + \mathbf{R}\mathbf{w}$.
D. $-\mathbf{q} + 2\mathbf{R}\mathbf{w}$.
E. none of the above.

(b) The order of the predicate $\Psi(|X| = M_1 \vee |X| = M_2 \vee |X| = M_3)$, where $0 < M_1 < M_2 < M_3$ and M_1, M_2, M_3 are integers, is

A. 2.
B. 4.
C. 6.
D. 8.
E. none of the above.

(c) The orders of the predicates $\Psi_1 = T_b(x_7 + x_6 + x_5 + x_4 + x_3 + 5x_0 - 6)$ and $\Psi_2 = T_b(\sum_{i=3}^{7}(x_0 x_i - 1))$, where x_k for $k = 0, 3, \ldots, 7$ are unipolar binary variables, are

A. both equal to 2.

B. both equal to 1.

C. different from each other.

D. none of the above.