Pergamon

Neural Networks, Vol. 9, No. 3, pp. 483–501, 1996

Printed in Great Britain
0893–6080/96 $15.00 + .00

0893-6080(95)00114-X

CONTRIBUTED ARTICLE

Asymmetric Hopfield-type Networks: Theory and Applications

ZONG-BEN XU,[1] GUO-QING HU[1] AND CHUNG-PING KWONG[2]

[1] Xi'an Jiaotong University and [2] The Chinese University of Hong Kong

(*Received* 27 *September* 1993; *revised and accepted* 12 *September* 1995)

Abstract—*The Hopfield-type networks with asymmetric interconnections are studied from the standpoint of taking them as computational models. Two fundamental properties, feasibility and reliability, of the networks related to their use are established with a newly-developed convergence principle and a classification theory on energy functions. The convergence principle generalizes that previously known for symmetric networks and underlies the feasibility. The classification theory, which categorizes the traditional energy functions into regular, normal and complete ones according to their roles played in connection with the corresponding networks, implies that the reliability and high efficiency of the networks can follow respectively from the regularity and the normality of the corresponding energy functions. The theories developed have been applied to solve a classical NP-hard graph theory problem: finding the maximal independent set of a graph. Simulations demonstrate that the algorithms deduced from the asymmetric theories outperform those deduced from the symmetric theory. Copyright © 1996 Elsevier Science Ltd.*

Keywords—Asymmetric Hopfield-type networks, Convergence principle, Classification theory on energy functions, Regular and normal correspondence, Maximal independent set problem, Combinatorial optimization.

1. INTRODUCTION

The neural network model we will consider is the binary Hopfield-type network represented by the weighted and undirected graph $N = (W, T)$. This type of network has been intensively studied and applied in the literature and with a common assumption that the interconnection weight W is symmetric (e.g., Kohonen, 1974, 1989; Little, 1974; Hopfield, 1982; Goles et al., 1985; Hopfield & Tank, 1985; Personnaz et al., 1985; Kanter & Sompolinsky, 1987; McEliece et al., 1987; Bruck, 1990; Bruck & Goodman, 1988; Bruck & Blaum, 1989; Bruck & Roychowdhury,1989; Cottrel, 1988; Aiyer et al., 1990). In this paper our aim is to provide detailed studies on the Hopfield-type networks with asymmetric interconnections.

This research has been inspired by several facts: first, the Hopfield network (Hopfield, 1982) is known to be one of the most successful and influential models in both biological and artificial neural systems. Many other models, say, bidirectional associative memories (BAM) (Kosko, 1988), Boltzmann machines (Hinton and Sejnowski, 1983), Q-state attractor neural networks (Kohring, 1993) are either direct variants or generalizations of this type. Therefore, any more deep understanding of the Hopfield-type networks can naturally shed some light on those models. Second, the symmetry assumption is somewhat unnatural and unacceptable for physiological reasons. It precludes many applications that are biologically or practically important (e.g., see Porat, 1989; Hertz et al., 1991; Coolen & Sherrington, 1993; Xu et al., 1994). For example, when used as knowledge representation, the symmetric connection of the BAM, viewed as a Hopfield-type network for storing pattern pairs $Z^{(i)} = X^{(i)}, Y^{(i)})^{\mathrm{T}}$ in product space, implies that the logical implication "$X^{(i)}$ if and only if $Y^{(i)}$" must exist between $X^{(i)}$ and $Y^{(i)}$. Consequently it is impossible to apply this type of network in developing connectionist expert systems in which asymmetric logical relations ordinarily hold (Xu et al., 1994). Thus it is imperative to apply the Hopfield networks with asymmetric connections. Finally, even if a theoretically well-designed symmetric Hopfield

Acknowledgements: The authors wish to thank the referees for their valuable suggestions towards improvement of presentation of this paper. This work was supported by the Research Grants Council of Hong Kong, and was carried out when the first author worked as a Research Fellow at Department of Information Engineering, the Chinese University of Hong Kong.

Requests for reprints should be sent to Profesor Zong-Ben Xu, Research Center for Applied Mathematics, Xi'an Jiaotong University, Xi'an, 710049, The People's Republic of China. E-mail:zbxu@xjtu.edu.cn

network is justified to be an effective and appropriate model for a specific application task, it is almost impossible to implement the hardware network precisely conserving symmetry connections (Hertz et al., 1991). So the physically realized network (an appliance) actually is asymmetric. This raises the importance of developing an asymmetric Hopfield network theory based on which the actual performance of the network can be evaluated.

There have been quite a few papers on asymmetric neutral networks (e.g., Hopfield, 1982; Sompolinsky, 1987, 1988; Hertz et al., 1986, 1991; Goles, 1986; Parisi, 1986; van Hemmen & Kühn, 1986; Crisanti & Sompolinsky, 1987; Derrida et al., 1987; Coolen et al., 1993; Xu et al., 1994, 1995). However all these previous studies were either restricted only to a capacity analysis aspect, or less general (only dealing with the very weak asymmetry cases), or on somewhat different models. In this paper some general theories of asymmetric Hopfield-type networks will be developed from the standpoint of taking them as computational models. Two fundamental issues, feasibility and reliability, related to the use of the networks will be studied in particular.

1.1. The Hopfield-type Networks and Related Symmetric Theory

A Hopfield network $N = (W, T)$ of order n comprises n computational units, called neurons. In the pair (W, T), $W = (w_{ij})$ is an $n \times n$ real matrix with w_{ij} representing the interconnection weight from neurons j to i; $\mathbf{T} = (t_i)$ is an n-dimensional real vector with t_i representing the threshold attached to neuron i. There are two possible values for the *state* of each neuron: $+1$ or -1. Denote the state of neuron i at time t as $v_i(t)$, the vector $\mathbf{V}(t) = (v_1(t), v_2(t), \ldots, v_n(t))^{\mathrm{T}}$ is then the state of the whole network at time t.

The state of a neuron is updated according to the following equation:

$$v_i(t+1) = \operatorname{sgn}\{H_i(t)\} = \begin{cases} 1, & H_i(t) \geqslant 0 \\ -1, & H_i(t) < 0 \end{cases} \tag{1.1}$$

where

$$H_i(t) = \sum_{j=1}^{n} w_{ij} v_j(t) - t_i. \tag{1.2}$$

If only one neuron is allowed to change state at any time instant, the network is said to be operating in a *serial mode*. Otherwise, the network is operating in a *parallel mode*, or *fully parallel mode* when all the neurons are updated simultaneously.

The Hopfield networks have been employed both as associative memories and as methods of solving optimization problems. The examples are those related to pattern recognition and image recovery (Hopfield, 1982; Cottrel, 1988; Kosko, 1988; Aiyer et al., 1990), and those related to computations of various NP-hard problems such as the traveling salesman problem (TSP) (Hopfield & Tank, 1985), minimal vertex covers of undirected graphs (Shrivastava et al., 1992), repair of RAMs (Mazumadar & Yih, 1989), error correcting codes (Bruck & Blaum, 1989; Bruck & Roychowdhury, 1989) and map coloring problems (Dahl, 1988).

The performance of any tasks listed above is accomplished by Hopfield networks through the search for certain bipolar vectors known as stable state(s) of these networks. A bipolar vector $\mathbf{V}^*$ is said to be a *stable state* of the network $N = (W, T)$ if it satisfies

$$\mathbf{V}^* = \operatorname{sgn}(WV^* - T). \tag{1.3}$$

In associative memory applications, for example, these stable states are treated as memorized patterns. Likewise in optimization applications the stable states correspond to possible suboptima of the objective function, and one of the stable states is the expected global optimum.

The success of Hopfield networks thus rests on the assured convergence of the state sequence of the networks. In the symmetric case, this is well studied. The basic results for instance are as follows:

- The network $N = (W, T)$ will always converge to a stable state when operating in the serial mode, if the diagonal elements of W are nonnegative (Hopfield, 1982).
- The network $N = (W, T)$ will either converge to a stable state or encounter a two-cycle when operating in the fully parallel mode. In particular, N will converge to a stable state if W is nonnegative definite (Goles et al., 1985).

No general result exists for assured convergence of the networks with asymmetric interconnections (but, see Goles, 1986; Xu & Kwong, 1995a).

In this paper, we will be mainly concerned with computation applications of asymmetric Hopfield-type networks (the Hopfield networks with asymmetric interconnections). In this case, as pointed out by Porat (1989), the assured convergence property has then been considered as a prerequisite for accepting such networks as a feasible computational model. We will resolve this *feasibility* problem by developing a very general convergence principle for asymmetric Hopfield-type networks.

Besides the feasibility, another issue also critical to success of the optimization applications is the *reliability* of the networks. When a particular computation problem, say TSP, is prepared to be solved by using a Hopfield network, the first step of one's effort then is to transform the problem into a minimization of an *energy* function of the form

$$E(V) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} v_i v_j + \sum_{i=1}^{n} t_i v_i + C. \qquad (1.4)$$

Then, the deduced network $N = (W, T)$ is implemented to minimize the energy. In this case, the reliability of N is concerned with the problem whether or not the optimum of the energy (equivalently, the optimal solution of the original computation problem) can be assuredly found by searching for stable states of the network. We will conduct a detailed study on this reliability issue.

1.2. Contributions

We first resolve the feasibility problem by developing a generalized convergence principle for asymmetric Hopfield-type networks. We prove that any Hopfield network $N = (W, T)$ will converge to a stable state when operating in the serial mode, whenever W is *diagonally dominant* in certain mild sense such as

(i)

$$w_{ii} \geqslant \frac{1}{2}\sum_{j=1, j\neq i}^{n} |w_{ij} - w_{ji}|, \forall i \in \{1, 2, \ldots, n\}$$

or

(ii)

$$w_{ii} \geqslant \sum_{j=1, j\neq i}^{n} |w_{ji}| + |t_i|,$$

or

(iii)

$$w_{ii} \geqslant \sum_{j=1, j\neq i} |w_{ij}| + 1t_i|.$$

The convergence of N in the parallel mode is also verified under corresponding nonnegative definiteness conditions. The results established not only generalize the global convergence theorems previously known for symmetric Hopfield networks, but also result in various new, easily verifiable tests for global convergence of asymmetric networks. They therefore underlie the feasibility of accepting the asymmetric Hopfield networks as computational models (Porat, 1989).

We also develop a classification theory on the energy functions associated with the Hopfield networks, based on which a satisfactory settlement of the reliability problem is made. Let $\Omega(N)$ denote the set of all stable states of a Hopfield network $N = (W, T)$, and $\Omega(E)$ the set of all local minimizers of the energy function $E(V)$ defined by (1.4). We classify the energy functions into three categories: *complete energies*, *regular energies* or *normal energies*, according to $\Omega(N) = \Omega(E)$, $\Omega(N) \supseteq \Omega(E)$ or $\Omega(N) \subseteq \Omega(E)$. The conditions for an energy function to be complete, regular or normal are characterized, and are found to be inversely related, as illustrated in the following specializations:

(i) E is a regular energy if

$$w_{ii} \geqslant \frac{1}{2}\sum_{j=1, j\neq i}^{n} |w_{ij} - w_{ji}| (i = 1, 2, \ldots, n);$$

(ii) E is a normal energy if

$$w_{ii} \leqslant -\frac{1}{2}\sum_{j=1, j\neq i}^{n} |w_{ij} - w_{ji}| (i = 1, 2, \ldots, n);$$

(iii) E is a complete energy if W is symmetric and with $w_{ii} = 0$.

We show that the regularity of an energy function implies the reliability of the deduced Hopfield network, and that the normality implies the high efficiency. The theory thus provides a solid foundation on reliability and high efficiency of both symmetric and asymmetric Hopfield networks in performing optimization computations. Furthermore, the theory reveals also some new, exclusive properties of symmetric networks. For instance, we conclude that the network with zero diagonals in the weight matrix has the least spurious stable states among all symmetric networks with weight matrices having nonnegative diagonals. This finding has led to a general strategy of improving the performance of existing symmetric networks of the Hopfield-type.

To demonstrate the power of the theories developed, we have applied the theories to a well-known graph theory problem: finding the *maximal independent set* of a graph (MIS problem), which is known to model wide-ranging applications in many fields (see, e.g. Bondy & Murty, 1976; Mazumadar & Yih, 1989; Pramanick, 1991) but be NP-hard in computation. By making use of a particular asymmetric Hopfield network structure (precisely, a triangular network), we propose two novel algorithms for

finding the best quality suboptimal solution(s) of the problem within computer memory permission. Unlike most of the existing algorithms that can only result in suboptimal solution(s), one of our algorithms can find global optimal solutions whenever the memory capacity is permissible. Another algorithm provides an elaborated trade-off strategy between the memory capacity and the computation time, which can always give the best quality suboptimal solution(s) within the memory capacity permission and by using less computation time. Simulations reveal that the algorithm outperforms the Hopfield's algorithm (Shrivastava et al., 1992) deduced directly from symmetric Hopfield networks.

The generalized convergence principle of asymmetric Hopfield networks is developed in Section 2. The classification theory on energy functions is enunciated in Section 3. The application to MIS problem and a series of simulations are presented respectively in Sections 4 and 5 to demonstrate the power of the established theoretical results. The paper then concludes with a summary in Section 6.

2. GENERALIZED CONVERGENCE PRINCIPLE

In this section we derive general conditions under which any Hopfield-type network converges to a stable state. The difficulty with this derivation was well recognized (e.g., see Goles, 1986; Porat, 1989). Our basic trick in this paper then is to develop a *parameter matrix skill combined with a carefully modified energy function reasoning*. A diagonal $n \times n$ matrix, say $R = \mathrm{diag}\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, is said to be a parameter if $\alpha_i > -1$ for any $i \in \{1, 2, \ldots, n\}$. Thus, in our derivation we will, instead of (1.1) and (1.2), consider the Hopfield network N in the serial and the fully parallel modes of operation respectively as

$$v_i(t+1) = \begin{cases} \mathrm{sgn}\left\{(1+\alpha_i)\left(\sum_{j=1}^{n} w_{ij} v_j(t) - t_i\right)\right\}, & \text{if } i = k \\ v_i(t), & \text{if } i \neq k \end{cases} \tag{2.1}$$

and

$$V(t+1) = \mathrm{sgn}\{(R+I)(WV(t) - T)\}, \; t \geqslant 0. \tag{2.2}$$

It should be noted that (2.1) and (2.2) really are no different in trajectories from the original updated models. However, the representations of (2.1) and (2.2) have their key benefit in uniformly deriving the general convergence conditions. In particular, this enables us to state and prove the convergence principle in a very general and uniform framework.

For any parameter matrix $\mathbf{R}$ and any small real number $\varepsilon > 0$, we denote

$$k_i(\alpha_i) = (1+\alpha_i)^{-1}\left(\sum_{j=1, j\neq i}^{n} |w_{ji} = \alpha_i w_{ij}| + (1+\varepsilon)|\alpha_i - 1||t_i|\right) \tag{2.3}$$

$$K(R) = \begin{pmatrix} k_1(\alpha_1) & 0 & \cdots & 0 \\ 0 & k_2(\alpha_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_n(\alpha_n) \end{pmatrix} \tag{2.4}$$

and

$$W^*(R) = W - K(R). \tag{2.5}$$

The basic result of this section then is stated as follows.

THEOREM 1. *Let $N = (W, T)$ be a Hopfield network, W not necessarily symmetric. Then N will converge to a stable state from any initial state $V(0)$, if it satisfies either* **(P1)** *or* **(P2)** *listed below:*

(P1): *N is operating in serial mode and W is weakly diagonal dominant in the sense that there exists a parameter matrix R such that*

$$w_{ii} \geqslant k_i(\alpha_i), \; \forall i = 1, 2, \ldots, n. \tag{2.6}$$

(P2): *N is operating in parallel mode and there exists a parameter matrix R such that $W^*(R)$ is nonnegative definite.*

Proof. See Appendix A.

Theorem 1 provides us with a very general convergence test for Hopfield-type networks. In particular, owing to existence of the parameter R in the listed conditions, which can be chosen arbitrarily, a series of specific tests for global convergence of the networks can be derived. For examples, we have:

COROLLARY 1. *Any one of the conditions* **(C1)**–**(C5)** *listed below is sufficient for the network $N = (W, T)$ to globally converge to a stable state when operating in the serial mode.*

For any $i = 1, 2, \ldots, n$,

(C1): W is symmetric and $w_{ii} \geqslant 0$;
(C2): W is β-symmetric,[1] with $\beta_i > -1$ and $w_{ii} \geqslant (1+\beta_i)^{-1}|\beta_i - 1||t_i|$;

(C3): $w_{ii} \geqslant \frac{1}{2}\sum_{j=1, j\neq i}^{n} |w_{ij} - w_{ji}|$;

(C4): $w_{ii} \geqslant \sum_{j=1, j\neq i}^{n} |w_{ji}| + (1+\varepsilon)|t_i|$;

(C5): $w_{ii} \geqslant \sum_{j=1, j\neq i}^{n} |w_{ij}| + (1+\varepsilon)|t_i|$.

Proof. This follows directly from Theorem 1, provided we notice that, when α_i is chosen to be 1, β_i, 1, 0, and tend to positive infinity, the condition (2.6) then reduces to **(C1)**, **(C2)**, **(C3)**, **(C4)** and **(C5)**, respectively.

Similarly, we have the following corollary:

COROLLARY 2. *Any one of the conditions* **(C1)′–(C5)′** *listed below is sufficient for the network* $N = (W, T)$ *to converge globally to a stable state when operating in parallel mode.*

(C1)′: W is symmetric, nonnegative definite;
(C2)′: W is β-symmetric with $\beta_i > -1$, and the matrix

$$W - \text{diag}\{(1+\beta_i)^{-1}|\beta_i - 1||t_i| : i = 1, 2, \ldots, n\}$$

is nonnegative definite;
(C3)′: $W - \text{diag}$

$$\left\{\frac{1}{2}\sum_{j=1, j\neq i}^{n} |w_{ij} - w_{ji}| : i = 1, 2, \ldots, n\right\}$$

is nonnegative definite;
(C4)′: $W - \text{diag}$

$$\left\{\sum_{j=1, j\neq i}^{n} |w_{ji}| + (1+\varepsilon)|t_i| : i = 1, 2, \ldots, n\right\}$$

is nonnegative definite;
(C5)′: $W - \text{diag}$

$$\left\{\sum_{j=1, j\neq i}^{n} |w_{ij}| + (1+\varepsilon)|t_i| : i = 1, 2, \ldots, ; n\right\}$$

is nonnegative definite.

[1] A matrix W is β-symmetric if there is an n-vector $\beta = (\beta_1, \beta_2 \ldots \beta_n)$ such that $\beta_i = w_{ij}/w_{ji}$ for any $j \neq i$ and $w_{ji} \neq 0$. In this term, clearly a symmetric matrix is 1-symmetric and an anti-symmetric matrix is (-1)-symmetric.

Observe that the conditions **(C1)**, **(C1)′** and **(C3)**, **(C3)′** are exactly those derived by Hopfield (1982), Goles et al. (1985) and Xu et al. (1995). However, in Corollaries 1 and 2, the conditions **(C2)**, **(C4)** and **(C5)**, **(C2)′** and **(C4)′** and **(C5)′** all are new findings for convergence of asymmetric Hopfield networks.

The following example shows that the conditions **(C3)–(C5)** [correspondingly, **(C3)′–(C5)′**] may be very different. Given a Hopfield network $N = (W, 0)$ with

$$W = \begin{pmatrix} a & 0 & 0 & -0.5 \\ 0 & b & -0.4 & 0 \\ 0.3 & 0.1 & c & 0 \\ 0.1 & 0.1 & 0 & d \end{pmatrix} \tag{2.7}$$

then **(C3)** implies global convergence of N when $a \geqslant 0.45$, $b \geqslant 0.35$, $c \geqslant 0.45$ and $d \geqslant 0.35$, but **(C4)** and **(C5)** imply the same respectively when $a \geqslant 0.4$, $b \geqslant 0.3$, $c \geqslant 0.4$ and $d \geqslant 0.5$ as well as $a \geqslant 0.5$, $b \geqslant 0.4$, $c \geqslant 0.5$ and $d \geqslant 0.2$.

The existence of the so many different conditions listed can not only increase one's ability to evaluate global convergence of various existing Hopfield-type networks, but also often make it possible to modify a given network or to develop a network so as to achieve the most desirable performance. In Section 4, we will explain this in more detail by considering its application in a classical graph theory problem. There, one will see that using the weakest convergence condition (if available) will be the most expected.

Thus, in Theorem 1, to use an optimal choice of R such that every $k_i(\alpha_i)$ is minimized will be of particular importance. We state this as Corollary 3.

COROLLARY 3. *Let* $R^* = \text{diag}\{\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*\}$ *be such that*

$$k_i(\alpha_i^*) = \inf_{-1<\alpha<\infty}\left\{(1+\alpha)^{-1}\left(\sum_{j=1, j\neq i}^{n} |w_{ji} - \alpha w_{ij}| + (1+\varepsilon)|\alpha - 1||t_i|\right)\right\}. \tag{2.8}$$

If

(C6): $w_{ii} > k_i(\alpha_i^*)$ *for any* $i \in \{1, 2, \ldots, n\}$,

then the network $N = (W, T)$ *globally converges to a stable state when operating in the serial mode; if*

(C6)′: $W - K(R^*)$ *is positive definite,*

then N converges to a stable state when operating in the parallel model.

In general, it is difficult (but not impossible) to

obtain the optimal values $k_i(\alpha_i^*)$. But, we can find the values in some particular cases. For example, we calculate in example (2.7) considered above that the optimal values are

$$k_1(\alpha_1^*) = 0.375,\ k_2(\alpha_2^*) = 0.2,\ k_3(\alpha_3^*) = 0.4,\ k_4(\alpha_4^*) = 0.2.$$

More generally, if $N = (B, T)$ is a lower triangular network [namely, $B = (b_{ij})_{n\times n}$ with $b_{ij} = 0$ for any $j > i$], just as will be encountered in our application in Section 4, we can easily find

$$k_i(\alpha_i^*) = \begin{cases} (a_i + b_i)/2; & \text{if } c_i \geqslant |a_i - b_i|/2 \\ b_i + c_i; & \text{if } c_i < |a_i - b_i|/2 \text{ and } a_i \geqslant b_i \\ a_i + c_i; & \text{if } c_i < |a_i - b_i|/2 \text{ and } a_i < b_i \end{cases} \tag{2.9}$$

where

$$a_i = \sum_{j=1}^{i-1} |b_{ij}|,\ b_i = \sum_{j=i+1}^{n} |b_{ij}|, \text{ and } c_i = |t_i|. \tag{2.10}$$

The global convergence property (or total stability, as also used in some other literature) of a connectionist model has been considered as a prerequisite for accepting the model as a feasible computational one (Porat, 1989). In this sense, the analysis conducted in this section has underlain the *feasibility* of asymmetric Hopfield-type networks as computational models. In the next section we will further consider the *reliability* aspect of the networks when used as computational models.

3. CLASSIFICATION THEORY ON ENERGY FUNCTIONS

We now study the reliability problem of Hopfield networks by establishing a classification theory on energy functions. This investigation, as observed in the Introduction to the paper, is original in the context of Hopfield-type networks both in symmetric and asymmetric cases.

3.1. The Classification Theory

Consider the problem

$$\text{minimize}\{E(V) : V = (v_1, v_2, \ldots, v_n) \in \{-1, +1\}^n\} \tag{3.1}$$

where $E(V)$ is a quadratic function of the form (1.4), which is modeling, for example, the TSP or the MIS problem. Associated with (3.1), we consider the Hopfield network $N = (WT)$ as an algorithm (i.e., a computational model) of solving the problem (3.1). Then, in this case, the classification theory to be developed, in essence, will characterize the reliability of such applications of N. To specify this more clearly, let us start with introducing a definition of local minimizer of E.

DEFINITION 1. *A vector* $\mathbf{V}^*$ *is said to be a local minimizer of* E *if* $E(\mathbf{V}^*) \leqslant E(V)$ *for any* $V \in B_H(\mathbf{V}^*, 1) = \{V : d_H(V, \mathbf{V}^*) \leqslant 1\}$, *where* $d_H(V, \mathbf{V}^*)$ *denotes the Hamming distance between* V *and* $\mathbf{V}^*$; a local minimizer $\mathbf{V}^*$ is said to be a global minimizer if $E(\mathbf{V}^*) \leqslant E(V)$ for any $V \in \{-1, +1\}^n$.

It is clear that any global minimizer of E is an optimal solution of (3.1). Consequently, to find such a global minimizer will be ideal for solving problem (3.1). However, (3.1) is clearly NP-hard. Its optimal solution can in general only be found after performing an exponential number of computations, which very often is an infeasible proposition. Thus, instead of finding optimal solution, one usually solves (3.1) by obtaining certain suboptimal solutions through a polynomial number of computations. There is no exact definition on what a suboptimal solution means. However the local minimizer notion introduced in Definition 1 can safely serve as a candidate for suboptimal solutions in any case.

Thus, solving the problem (3.1) may be based on two different purposes: either to find an optimal solution or to find a suboptimal solution. Since any Hopfield network finds such solutions through searching its certain stable states, it requires that, for the first purpose, any local minimizer of E must be in the stable states of N, in case the optimal solution (global minimizer) is lost. Similarly it should be required for the second purpose that any stable state of N is a local minimizer of E, assuring that any solution generated by N is a reasonable candidate of suboptimal solutions.

In characterizing the different requirements above, the following Definition 2 will be shown to be useful. We let

- $\Omega(E)$ be the set of all local minimizers of E, and
- $\Omega(N)$ be the set of all stable states of N.

DEFINITION 2. (i) *E is called a regular energy of N if $\Omega(E) \subseteq \Omega(N)$ (i.e, any local minimizer of E is a stable state of N); in this case, N is said to have regular correspondence property;*

(ii) *E is called a normal energy of N if $\Omega(E) \supseteq \Omega(N)$ (i.e., any stable state of N is a local minimizer of E). In this case, N is said to have normal correspondence property;*

(iii) *E is called a complete energy of N if $\Omega(E) = \Omega(N)$ (i.e., there is a one-to-one correspondence between the local minimizers of E and the stable*

states of N). In this case, N is said to have complete correspondence property.

From Definition 2, E is both regular and normal if it is complete. Nevertheless, it will be seen that the regular energy property (regular correspondence property) and normal energy property (normal correspondence property) are in general conversely related.

THEOREM 2. *For any parameter matrix R, let*

$$\sigma(\alpha_i -) = (1+\alpha_i)^{-1}\left(\sum_{j=1, j\neq i}^{n} |w_{ji} - \alpha_i w_{ij}| + |\alpha_i - 1||t_i|\right). \tag{3.2}$$

(i) *If there is an R such that*

$$w_{ii} \geqslant \sigma(\alpha_i), \forall i \in \{1, 2, \ldots, n\} \tag{3.3}$$

then E is a regular energy of N;

(ii) *If there is an R such that*

$$w_{ii} \leqslant -\sigma(\alpha_i), \forall_i \in \{1, 2, \ldots, n\} \tag{3.4}$$

then E is a normal energy of N;

(iii) *If there is an R such that*

$$w_{ii} = \sigma(\alpha_i) = 0, \forall_i \in \{1, 2, \ldots, n\} \tag{3.5}$$

then E is a complete energy of N.

Proof. See Appendix B.

The inequalities (3.3) and (3.4) can be shown to be most possibly tight conditions for E to be a regular, normal or complete energy. To see this, let us look through the Hopfield network $N = (W, T)$ and the corresponding energy function E, specified by

$$\begin{pmatrix} \eta & 0 \\ 1 & 0 \end{pmatrix} \text{ and } T = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

It is easy to show that, in this case, (3.3) implies $\eta > 0$ is a condition under which E becomes regular (in fact, $\Omega(E) = \Omega(N) = \{(1, 1)^T, (-1, -1)^T\}$). By loosening this condition very slightly, e.g., to $\eta = 0$, we then find $\Omega(N) = \{(1, 1)^T\}$, $\Omega(E) = \{(1, 1)^T, (-1, -1)^T\}$. That is, $\Omega(E) \not\subset \Omega(N)$, E is no longer regular. This shows that $\eta > 0$ is the weakest condition for regularity of E. In the same way, we can show that $\eta \leqslant 0$, implied by (3.4), is also the weakest condition for normality of E.

It is interesting to observe that in Theorem 2, the regular energy condition (3.3) and the normal energy condition (3.4) are exactly symmetric about zero. Therefore, once (3.3) is satisfied, that E is a regular energy of $N = (W, T)$ is equivalent to $-E$ is a normal energy of $-N = (-W, -T)$ and vice versa.

Observe also that the regular energy condition (3.3) differs from the global convergence condition (2.6) in Theorem 1 only because of the existence of the infinitesimal, ε, in (2.3) (that is, they are almost the same). This pleasantly surprising relationship can greatly facilitate applications of the networks. For instance, the same criteria **(C1)**–**(C6)** can then be used both to test global convergence and to test the regular correspondence property of a network.

Thus, we give the following corollary.

COROLLARY 4. *If any one of the conditions* **(C1)**–**(C6)** *in Section 2 is satisfied, then N has regular correspondence property.*

In particular, as in **(C1)**, we take $R = \text{diag}\{1, 1, \ldots, 1\}$ in Theorem 2 (iii), yielding the following:

COROLLARY 5. *If W is symmetric, with all diagonal elements being zero, then N has a complete correspondence property.*

Theorem 2 (Corollaries 4 and 5) will be of great significance in applications of Hopfield-type networks. We will further explain this in the next subsection.

3.2. Significance

(1) When the Hopfield network N is applied as an algorithm for solving (3.1), as mentioned above, no solid foundation has been known with regard to its reliability. Theorem 2 has provided an answer to this question: if an optimal solution of (3.1) is needed, the necessary condition for its reliability is that N has a regular correspondence property. That is, the regularity of N implies the reliability.

On the other hand, if N has a normal correspondence property, the network N searches for the solution(s) of (3.1) only among its local minimizers, i.e., in the cheapest way. We can then also conclude that the normality of N implies high efficiency of the network.

Thus, a network N having a complete correspondence property will be most ideal for solving (3.1), because it is both regular and normal (hence is both reliable and efficient). By Corollary 5, this ideal case can always happen whenever W is symmetric [in fact, as the basic variable V in (1.4) is bipolar, we can adjust the diagonal elements of W, which is

equivalent to adding an appropriate constant into E, such that E becomes the complete energy of a network N]. But this is unlikely in general for asymmetric cases.

However, in every case, we can easily construct (e.g., by adjusting the diagonal elements of W appropriately) a Hopfield network which has the expected regular correspondence property, and hence solve (3.1) reliably. To give an example, let us consider the function E specified by

$$W = \begin{pmatrix} -1 & 2 & 1 \\ 3 & 0 & 2 \\ 1 & 2 & 3 \end{pmatrix}, \quad T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } C = 0. \tag{3.6}$$

The function E itself is not a regular energy of the network $N = (W, T)$ [hence, when used as an algorithm to minimize E, $N = (W, T)$ is then not reliable]. However, adjusting the diagonal elements of W into W_1, as shown in (3.7), which functions as adding a constant $-2 = (-3/4)(v_1)^2 + (1/4)(v_2)^2 + (-3/2)(v_3)^2$ into the E (note that $v_i \in \{-1, 1\}$), we then see that the function $E + 1/2$ is a regular energy of the network $N_1 = (W_1, T)$ [Corollary 4 with **(C3)**], where

$$W_1 = W + \begin{pmatrix} -\frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & -\frac{3}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 2 & 1 \\ 3 & \frac{1}{2} & 2 \\ 1 & 2 & 0 \end{pmatrix}. \tag{3.7}$$

Thus N_1 has the expected regular correspondence property, and (3.1) in this case can be solved reliably by the asymmetric network $N_1 = (W_1, T)$.

To summarize the above analysis, we then propose the following tactical strategies for constructing Hopfield networks in performing the computation of (3.1):

- If a symmetric network is expected to be used, construct the network such that it has the complete correspondence property;
- If an asymmetric network is expected to be used, construct the network such that it has the regular correspondence property.

These strategies should underlie the applications of Hopfield networks in solving (3.1). We will particularly use them in the next section.

(2) When the Hopfield network N serves as a content addressable memory, its error-correction capability is known to be a fundamental feature. It has been known (Kanter & Sompolinsky, 1987; Xu and Kwong, 1995a; Xu et al., 1995) that this capability is essentially determined by how many spurious stable states N contains. The less spurious stable states N contains, the higher the error-correction capability of N. Therefore, for a given memory patterns set, say, $\bar{V} = \{V^{(1)}, V^{(2)}, \ldots, V^{(M)}\}$, it is of particular importance to recognize the network that has the least spurious stable states among all the applicable networks of Hopfield type. By saying a Hopfield network N is applicable we mean that N is globally convergent and $\bar{V} \subseteq \Omega(N)$.

Theorem 2 (Corollary 4) can play an important part in recognizing such a network. For instance, taking $\{V^{(1)}, V^{(2)}, \ldots, V^{(P)}\}$ to be the 26 English letters shown as in Figure 1, each of which is identified as a 49-dimensional vector, and let $N = (W^*, T^*)$ be the given applicable network specified by the pseudo-inverse rule:

$$W^* = \bar{V}\bar{V}^+ = \bar{V}(\bar{V}^T\bar{V})^{-1}\bar{V}^T, \quad T^* = 0$$

where $\bar{V} = [V^{(1)} | V^{(2)} | \ldots | V^{(M)}]$, $\bar{V}^+$ is the Moore–Penrose inverse of $\bar{V}$ and $\mathbf{0}$ is the n-dimensional zero vector. Then we can show that the network $N(D^*) = (W^* - D^*, T)$, where $D^* = \text{diag}\{w^*_{11},$

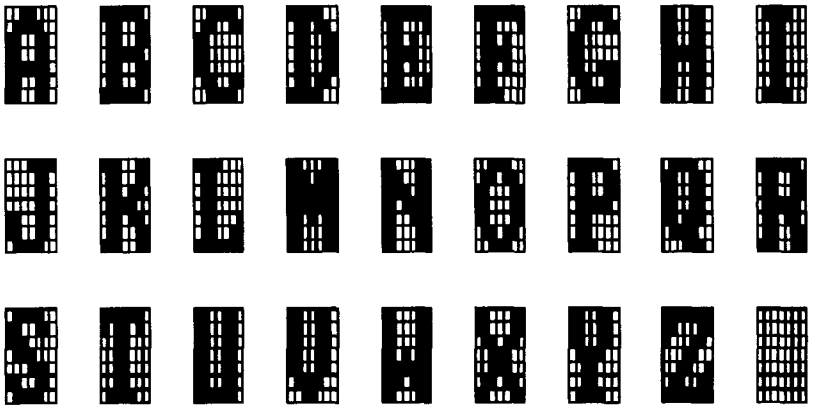


FIGURE 1. Sample memory patterns.

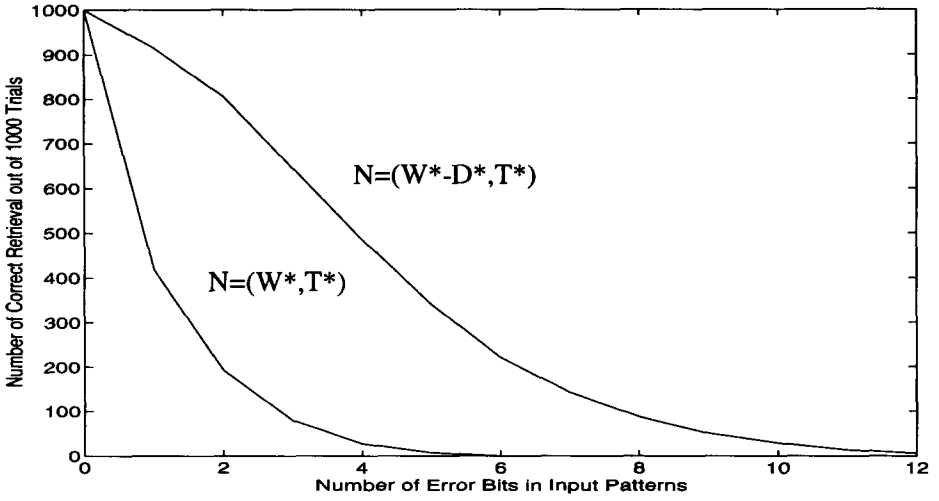


FIGURE 2. Comparison of error-correction capability between the network $N = (W^*, T^*)$ with pseudo-inverse rule and its normal-correspondence based modification $N = (W^* - D^*, T^*)$, as applied to 26-English letters recognition task.

$w^*_{22}, \ldots, w^*_{nn}\}$, is the one with least spurious stable states among the family of applicable networks:

$$\{N(D) = (W^* + D, T^*) : D = \text{diag}\{d_1, d_2, \ldots, d_n\} \text{ is a parameter matrix}\}.$$

It is based on finding the network with a complete correspondence property in the family that $N(D^*)$ is found. Simulations conducted reveal that $N(D^*)$ has yielded a dramatic improvement in both the storage capacity and the error-correction capability of N (cf. Figure 2). The details, together with the applications of the same principle to various other Hopfield-type networks, can be found in Xu et al. (1996).

4. APPLICATION TO MIS PROBLEMS

In this section we demonstrate an important application of the developed asymmetric theory of Hopfield networks in solving various NP-hard problems modeled by (3.1). The MIS problem in classical graph theory will be analyzed in detail as a typical example.

4.1. Maximal Independent Set (MIS) Problem

Consider an n-vertex undirected graph G in which no vertex has a selfloop. We will assume that the vertices of G are labeled $1, 2, \ldots, n$ (hence $V = \{1, 2, \ldots, n\}$ is the set of vertices of G).

Recall that a subset V' of V is said to be independent if no two vertices in V' are adjacent. An independent subset V' of V is maximal if every vertex in $V \setminus V'$ is connected to at least one vertex in V'. As examples, the black vertices in graphs 1–7 (Figure 4 later), are all maximal independent sets.

To find a maximal independent set (MIS) of a graph is a fundamental problem in graph theory. It is equivalent to that known as the vertex cover problem (Shrivastava et al., 1992), and has wide-ranging applications in many fields such as information retrieval, signal transmission, computer vision, computer networking, and repair of RAMs (Bondy & Murty, 1976; Mazumadar & Yih, 1989; Pramanick, 1991; Shrivastava et al., 1992).

Shrivastava and associates (1992) have shown that a class of symmetric Hopfield networks can be applied to finding the MISs of a graph more efficiently than most known algorithms. In this section, we will demonstrate that a particular asymmetric Hopfield network can in fact be applied to the MIS problem even much more efficiently than the symmetric network algorithms. Our approach differs completely from that of Shrivastava et al. (1992).

4.2. A Minimization Model for Finding MISs

Unlike the approach of Shrivastava et al. (1992), we will transform the problem of finding the MISs of a graph into a minimization problem of the form (3.1). To this end, for any given graph G, we let $A = (a_{ij})_{n\times n}$ be its adjacency matrix, which is defined by

$$a_{ij} = \begin{cases} 1, & \text{if vertices } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases}$$

We also associate any subset V' of G with an n-vector, $U = (u_1, u_2, \ldots, u_n)$, defined by

$$u_i = \begin{cases} 1, & \text{if vertex } i \text{ is in } V' \\ 0, & \text{otherwise.} \end{cases}$$

Then, the problem of finding a MIS of G is equivalent to that of finding an n-vector $U^* = (u_1^*, u_2^*, \ldots, u_n^*)^T$ in $\{0, 1\}^n$ such that

- U^* has as many as possible non-zero elements; and
- if u_i^* and u_j^* are both nonzero, then $a_{ij} = 0$ (i.e., vertices i and j are not connected).

Formulating these two requirements, we thus obtain the following optimization model for finding MISs in the graph G:

$$\text{Min}\{E(U) : U \in \{0, 1\}^n\}$$

with

$$E(U) = -\sum_{i=1}^{n}\sum_{j=1}^{n} u_i u_j + \frac{K}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} u_i u_j \tag{4.1}$$

where K is any sufficiently large positive constant, understood as a Lagrange multiplier. The relation between the MISs of the graph G and the solutions of (4.1) is presented in the following theorem.

THEOREM 3. *U^* is a MIS of the graph G if and only if U^* solves the problem (4.1) for any $K \geqslant 2(2n+1)$.*

Proof. See Appendix C.

In order that (4.1) can be solved by using the Hopfield networks discussed in the present paper, we rewrite (4.1), through the transformation $U : \to V = 2U - e$, where $e = (1, 1, \ldots, 1)^T$, as a typical form (3.1) with

$$\begin{aligned} E(V) = &-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{2} - \frac{K}{4}a_{ij}\right) v_i v_j \\ &+ \frac{K}{2}\sum_{i=1}^{n}\left(\frac{K}{8}\left(\sum_{j=1}^{n} a_{ij} + \sum_{j=1}^{n} a_{ji}\right) - \frac{n}{2}\right) v_i \\ &+ \left(\frac{K}{8}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} - \frac{n}{4}n^2\right). \end{aligned} \tag{4.2}$$

4.3. An Asymmetric Network Model for Solving (3.1)

The problem (3.1) with E specified by (4.2) can be solved, of course, by suitable symmetric Hopfield networks. We will show, however, that a much simpler, but asymmetric, Hopfield network can be applied, even much more effectively.

To be precise, we will construct a lower triangular Hopfield network $N = (B, T)$ to solve (3.1). Based on the discussions in Sections 2 and 3, this then requires the following:

(i) $B = (b_{ij})_{n\times n}$ which satisfies $b_{ij} = 0$ for any $j > i$;
(ii) $N = (B, T)$ has the global convergence property;
(iii) there is a constant C such that $E(V) + C$ is a regular energy of $N = (B, T)$ (hence N has regular correspondence property);
(iv) b_{ii} are as small as possible.

Such a network $N = (B, T)$ will be constructed as follows: T is the same as that in (3.1), but B is defined by

$$b_{ij} = \begin{cases} b_{ii}^*, & i = j \\ w_{ij} + w_{ji}, & j < i \\ 0, & j > i \end{cases} \tag{4.3}$$

where $b_{ii}^* = K_i(\alpha_i^*)$, calculated according to (2.9) and (2.10) in Section 2.

With such specified B and T, (i) is clearly satisfied. By Corollary 3, (4.3) and (2.9) and (2.10), (ii) and (iv) are also met. Furthermore, let

$$E_1(V) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} v_i v_j + \sum_{i=1}^{n} t_i v_i.$$

We observe from (4.3) that

$$\begin{aligned} E_1(V) &= -\frac{1}{2}\sum_{i=1}^{n}\left(\sum_{j=1}^{i-1} b_{ij} v_i v_j + b_{ii}\right) + \sum_{i=1}^{n} t_i v_i \\ &= -\frac{1}{2}\sum_{i=1}^{n}\left(\sum_{j=1}^{i-1} (w_{ij} + w_{ji}) v_i v_j + w_{ii} + d_i\right) + \sum_{i=1}^{n} t_i v_i \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} v_i v_j + \sum_{i=1}^{n} t_i v_i + \frac{1}{2}\sum_{i=1}^{n} d_i \\ &= E(V) + C_1 \end{aligned} \tag{4.4}$$

where

$$C_1 = \frac{1}{2}\sum_{i=1}^{n} d_i - C. \tag{4.5}$$

This implies (iii), also.

The network $N = (B, T)$, specified by (4.3)–(4.5), is the asymmetric Hopfield network model based on which our algorithms for solving (3.1) [MIS problem (4.1) in particular] will be developed.

4.4. Two Algorithms

Now, it is clear that any possible optimal solution(s) of (3.1) are in the stable states of the network $N = (B, T)$. Our algorithms then will provide two elaborate strategies for searching for stable states of N.

4.4.1. *Global Algorithm (GA).* This algorithm is designed to directly make use of the exclusive feature of the triangular network N: any stable state of N, say, $V^* = (v_1^*, v_2^*, \ldots, v_n^*)^T$, satisfies

$$\begin{cases} v_1^* = \operatorname{sgn}(b_{11}v_1^* - t_1) \\ v_i^* = \operatorname{sgn}\left[b_{ii}v_i^* - \left(t_i - \sum_{j=1}^{i-1} b_{ij}v_j^*\right)\right], \\ \forall i \in \{1, 2, \ldots, n\}. \end{cases} \tag{4.6}$$

The solutions of (4.6) are explicitly given by

$$v_i^* = \begin{cases} 1, & \delta_i \leqslant -b_{ii} \\ -1, & \delta_i > b_{ii} \\ 1 \text{ or } -1, & \delta_i \in (-b_{ii}, b_{ii}] \end{cases} \quad \forall i \in \{1, 2, \ldots, n\}. \tag{4.7}$$

where

$$\delta_i = t_i - \sum_{j=1}^{i-1} b_{ij}v_j^* \text{ for any } i \geqslant j. \tag{4.8}$$

Thus, we can find the global optimum of (3.1) through comparing all the stable states of N, which are easily obtainable from (4.7). This is stated as a global algorithm for finding optimal solution(s) for (3.1), as follows:

STEP 1. *Find all the stable states, say,* $\{V^*(1), V^*(2), \ldots, V^*(m)\}$, *of the system* (4.6) *according to the formula* (4.7). *In doing so, it is suggested that the "Binary Search Tree" algorithm is applied and, in each step of the binary search, the following* $\eta_i(k)$*'s are computed and stored, where*

$$\eta_i(k) = [b_{ii}V_i^*(k) - \delta_i(k)]\,V_i^*(k) \forall k \in \{1, 2, \ldots, m\}, \quad \forall i \in \{1, 2, \ldots, n\}. \tag{4.9}$$

STEP 2. *Set*

$$F(V^*(k)) = -\frac{1}{2}\sum_{j=1}^{n} \eta_i(k), \forall k \in \{1, 2, \ldots, m\} \tag{4.10}$$

and let the optimal solution(s) of (3.1) *be*

$$\{V^* \in \{V^*(1), \ldots, V^*(m)\} : F(V^*) \leqslant F(V^*(k)), k = 1, 2, \ldots, m\}.$$

From (4.4), (4.5), (4.9) and (4.10), the values for $F(V^*(k))$ differ from $E(V^*(k))$ with only the constant C, defined as in (4.5), which has no effect obviously on the location of the optimal solution(s) of (3.1) and hence the validity of the algorithm.

The global algorithm can definitely lead to all the optimal solutions of (3.1), provided the memory capacity of the computer used is suitable (note that a large amount of memory is needed in order to calculate all possible stable states for N). Simulations in the next section show that the algorithm indeed performs very efficiently as long as the memory capacity is sufficient.

The next algorithm is designed to offer a tactical trade-off strategy between memory capacity requirement and computation time.

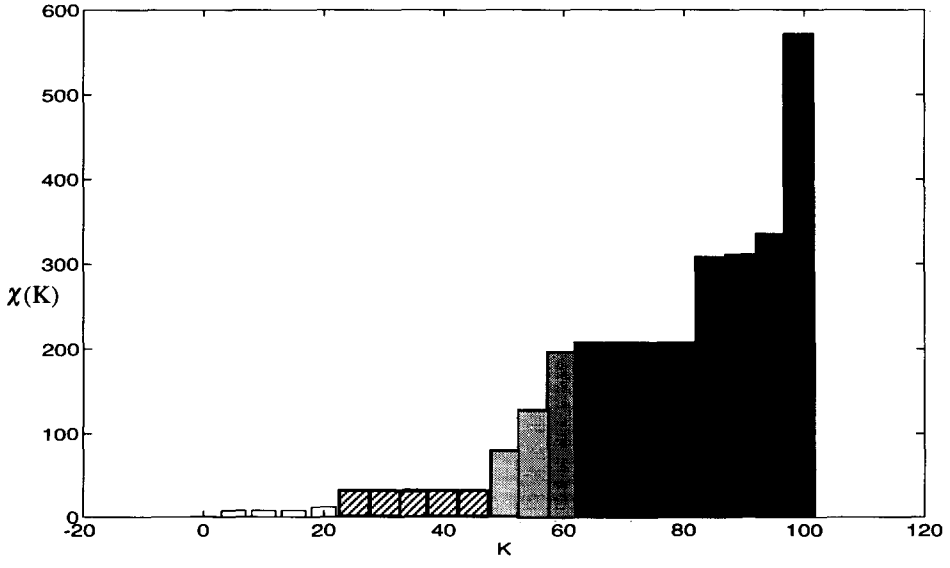


FIGURE 3. An illustration for monotonically increasing property of stable states $\chi(b_{11}, b_{22}, \ldots, b_{nn})$ of the network (B, T) with its diagonal elements b_{ii}. In the figure, $\chi(K) = \chi(0.01Kb_{11}^*, 0.01Kb_{22}^*, \ldots, 0.01Kb_{nn}^*)$ with fixed b_{ij}^*.

4.4.2. *Compromise Algorithm (CA).* This is an attempt to find suboptimal solutions of (3.1) as much as possible as allowed by the memory capacity. The basic idea is as follows:

We regard the stable states of network $N = (B, T)$ as a set-valued function of the diagonals b_{ii} $(i = 1, 2, \ldots, n)$. Denote this function by $\chi(b_{11}, b_{22}, \ldots, b_{nn})$. Then it is easily seen (cf. Figure 3 for an illustration) that $\chi(b_{11}, b_{22}, \ldots, B_{nn})$ is monotonically increasing with b_{ii}, in the sense of a set inclusion. [Here we should observe the fact that for any two Hopfield networks $N_1 = (W_1, T)$ and $N_2 = (W_2, T)$ with $W_1 = (w_{ij}^{(1)})$ and $W_2 = (w_{ij}^{(2)})$. If $w_{ij}^{(1)} = w_{ij}^{(2)}$ for any $i \neq j$, and $w_{ii}^{(1)} \leqslant w_{ii}^{(2)}$ for any i, then $\Omega(N_1) \subseteq \Omega(N_2)$, where $\Omega(N)$ denotes the set of stable states of N. This fact can be very easily proved, if we notice that a vector $\mathbf{V} = (v_1, v_2, \ldots, v_n) \in \Omega(N_k)$ $(k = 1, 2)$ if and only if

$$v_i\left(\sum_{j=1}^{n} w_{ij}^{(k)} v_j\right) = w_{ii}^{(k)} + v_i\left(\sum_{j \neq i} w_{ij}^{(k)} v_j\right) > 0, \; i = 1, 2, \ldots, n).$$

As a direct consequence, in the sense of set-inclusion, we can thus conclude that $\Omega(N)$ is an increasing function of the diagonal entries w_{ii} of W, i.e., the set $\Omega(N)$ will get larger when the diagonal entries w_{ii} of W get larger. In particular, with b_{ii}^* defined by (4.3), we can calculate that

$$\chi(0, 0, \ldots, 0) = \{\mathbf{V}^*\} \text{ and } \chi(b_{11}^*, b_{22}^*, \ldots, b_{nn}^*) = \Omega(N)$$

where $\mathbf{V}^{\#}$ is the unique vector defined by

$$v_1^{\#} = \operatorname{sgn}(t_1), \tag{4.11}$$

$$v_i^{\#} = \operatorname{sgn}\left\{\sum_{j=1}^{i-1} b_{ij} v_j^{\#} - t_i\right\}, \; i = 1, 2, \ldots, n \tag{4.12}$$

and $\Omega(N)$, as before, is the set of all stable states of $N = (B, T)$, which contains in particular the optimal solutions of (3.1) due to the regular correspondence property of N. Since $\chi(b_{11}, b_{22}, \ldots, b_{nn})$ is included in $\chi(b_{11}^*, b_{22}^*, \ldots, b_{nn}^*)$ for every $b_{ii} \in [0, b_{ii}^*]$, the optimal solution(s) of (3.1) is also possible in $\chi(b_{11}, b_{22}, \ldots, b_{nn})$. Thus, we propose, instead of finding $\chi(b_{11}^*, b_{22}^*, \ldots, b_{nn}^*)$, to find $\chi(b_{11}, b_{22}, \ldots, b_{nn})$ for searching for a suboptimal solution of (3.1) with a suitable $\{b_{11}, b_{22}, \ldots, b_{nn}\}$. Note that computing $\chi(0, 0, \ldots, 0)$ requires no memory space at all, yielding however, a most possibly lower quality suboptimal solution $\mathbf{V}^{\#}$, but, as b_{ii} increase, computing $\chi(b_{11}, b_{22}, \ldots, b_{nn})$ requires larger and larger memory capacity, resulting accordingly in higher and higher quality suboptimum. This provides a *continuation* strategy of compromising the memory capacity and the quality of the suboptimal solutions. In particular, the algorithm can then always find the highest quality suboptimal solutions of (3.1) within the limits of the memory capacity.

The algorithm deduced from the above idea is formally stated as follows:

STEP 1. *Set* $t_0 = \frac{1}{2}$, $k = 0$;
STEP 2. Let

$$b_{ii}(k) = t_k b_{ii}^*, \; i = 1, 2, \ldots, n; \tag{4.13}$$

and then search for $\chi(b_{11}(k), b_{22}(k), \ldots, b_{nn}(k))$ *and find the minimizer for* $E(V)$ *in* $\chi(b_{11}(k), b_{22}(k), \ldots, b_{nn}(k))$ *(by using a developed GA) if the memory capacity allows. Otherwise, let* $t_{k+1} = t_k/2$ *and go to Step 4;*
STEP 3. *If the minimizer found is desirable, stop the algorithm; otherwise, let* $t_{k+1} = (k+1)/(k+2)$;
STEP 4. *Go to Step 2 with* $k := k + 1$.

We will show through simulations in the next section that the proposed compromise algorithm is very promising in practice. We also state the following known Hopfield algorithm (Hopfield, 1982; Shrivastava et al., 1992), to facilitate cross-comparisons in the next section.

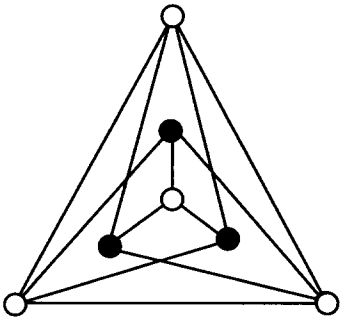
4.5. Symmetric Hopfield Algorithm (SHA)

STEP 1. *Let* $k = 1$, $V(0)$ *be any initial state;*
STEP 2. *Execute the Hopfield network* $N = (W, T)$ *in serial mode, starting from* $V(0)$*; Find a stable state* V_1 *of* N;
STEP 3. *If* V_1 *is a desirable suboptimal solution of* (3.1), *then stop; otherwise, go to Step 2 with any other randomly chosen* V (0).

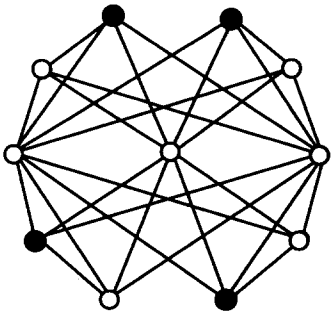
5. SIMULATIONS

We now give simulation results comparing the performance of the global algorithm (GA), the compromise algorithm (CA) and the symmetric Hopfield algorithm (SHA), as applied to MIS problems.

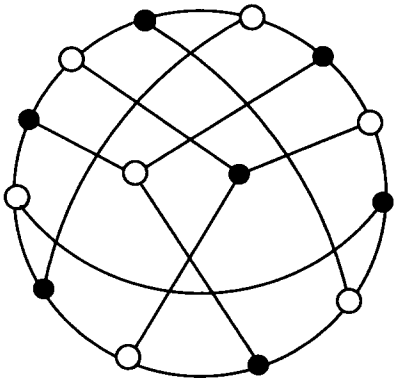
The example graphs used in the simulations are taken from the listed famous graphs in Bondy and Murty (1976), as shown in Figures 4a and 4b. In the simulations, whenever the memory capacity of the computer in use is suitable, all MISs of a graph were first calculated by using GA. The known cardinality of the MIS was then used as a measure of judging if a local optimal solution is desirable in executing CA and SHA. For comparison purposes, SHA has been run 20 times (tests), and in each time, 1000 randomly



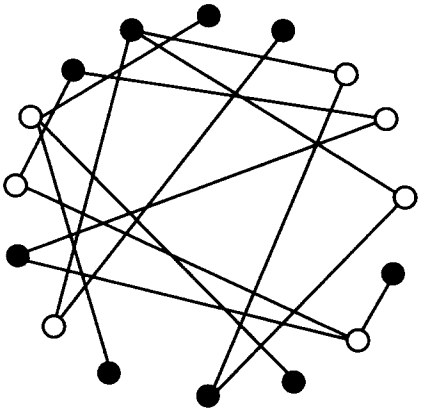
graph 1



graph 2

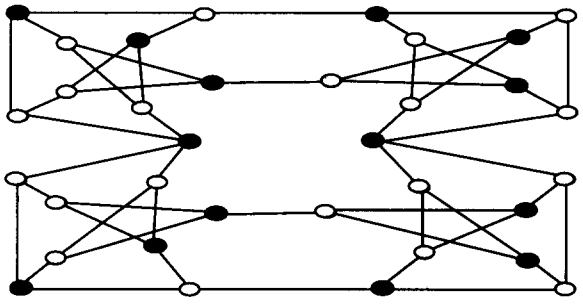


graph 3

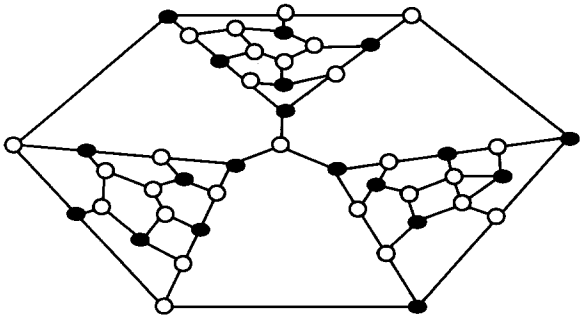


graph 4

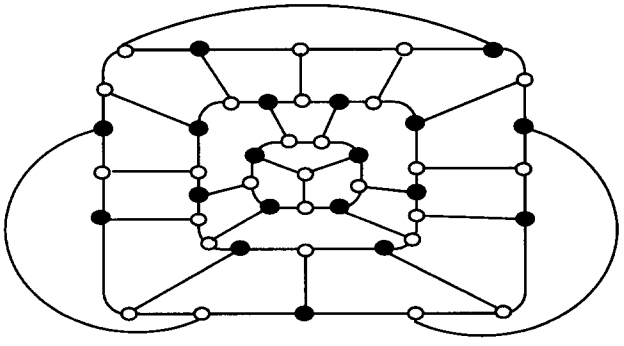
FIGURE 4a. The sample graphs (graphs 1–4).



graph 5



graph 6



graph 7

FIGURE 4b. The sample graphs (graphs 5–7).

TABLE 1
Comparison of Three Algorithms for MIS Problems

Example No.	|V|	|E|	GA |ST.S|	GA No.MIS	GA |MIS|	CA k	CA |ST.S|	CA No.MDS	CA |MDS|	SHA MDS %	SHA |MDS|	Optimal Solution
1	7	12	40	1	3	4	11	1	3	10.93	3	3
2	11	28	562	12	4	3	28	4	4	43.9	4	4
3	14	21	904	2	7	2	5	1	7	30.6	7	7
4	16	15	404	9	9	4	4	2	9	61.64	9	9
5	34	56				3	545	8	14	2.6	14	
6	46	66				5	1079	4	19	4.6	19	
7	46	69				5	3795	2	19	0.085	19	

chosen initial values were used. One then calculates the percentage of appearance of an MIS (or a same sized local MIS with that found by CA) among the total 20,000 runs.

The simulation results are summarized in Table 1. The first column gives $|V|$ and $|E|$, the cardinality of the vertices and edges of the corresponding graphs. The optimal solutions of graphs 1–4 are known and, therefore, their cardinality reported in the last column. In the "global algorithm" column we list the cardinality of optimal solution, |MIS|, the number of optimal solutions found, No.MIS, and the number of stable states, |ST.S|, from which the optimal solutions had been selected. Correspondingly, in the "compromise algorithm" column, except for |ST.S|, also listed are k, the number of steps the algorithm had run, No.MDS, the number of the most desirable solutions found, and |MDS|, the cardinality of the most desirable solutions obtained. In the "symmetric Hopfield algorithm" column we also list MDS%, the percentage of appearance of the most desirable solution in the total 20,000 runs, which provides a measure of reliability and efficiency of the Hopfield algorithm for solving MIS problems.

The global algorithm is capable of assuredly finding all of the optimal solutions in principle. Its applicability, however, is severely limited by the available memory capacity of the computer used. This is supported by the simulation results shown in Table 1. (Note that GA does not work for graphs 5–7, owing to the memory capacity limitation.)

The Hopfield algorithm, on the other hand, can be applied to any graphs without any memory requirement. Furthermore, it can, in principle, find the optimal solution in sufficient time. However, as indicated by the simulations, this algorithm actually finds the optimal or a most desirable solution with a high probability, within a reasonable polynomial time, only for small graphs. Indeed, from Table 1 and Figure 5, it is seen that the Hopfield algorithm finds the optimal solution with a probability ranging from 10.93% to 61.64% for graphs 1–4, but it finds the most desirable solution of graphs 5–7 only with a probability ranging from 0.085% to 2.6%. This performance implies that the high efficiency of the Hopfield algorithm, as it is applied to the MIS problem, is still questionable.

The global algorithm can be regarded as an elaborated safety method for finding the optimal solutions of MIS problems at the expense of memory space. The Hopfield algorithm, on the other hand, can be regarded as a novel method for the same purpose at the cost of computation time. The compromise algorithm, as its name implies, is a strategy of compromising between the space expense and the time expense. It can be seen from Table 1 that this algorithm always finds the optimal or the most desirable solution from very few stable states. This is a very attractive performance. In particular, it requires much less memory space than the global algorithm does (comparing |ST.S|s listed in "global algorithm" column and "compromise algorithm column"), and, at the same time, yields the solution without need for long iterations as needed by the Hopfield algorithm.

Since the Hopfield algorithm generally performs as well as or better than some other algorithms in the literature (Shrivastava et al., 1992), the compromise algorithm then deserves recommendation for use in MIS problems.

6. SUMMARY AND CONCLUSIONS

We have studied in detail the asymmetric Hopfield-type networks and shown how more powerfully they perform than the symmetric networks do in solving a wide-ranging class of combinatorial optimization problems.

The theoretical foundations of the networks are developed from the standpoint of taking the networks as computational models. Two key issues, feasibility and reliability, related to the use of networks are thoughtfully studied. Using a parameter matrix skill combined with a carefully modified energy function reasoning, a very general convergence theorem is proven, which underlies the feasibility of the net-

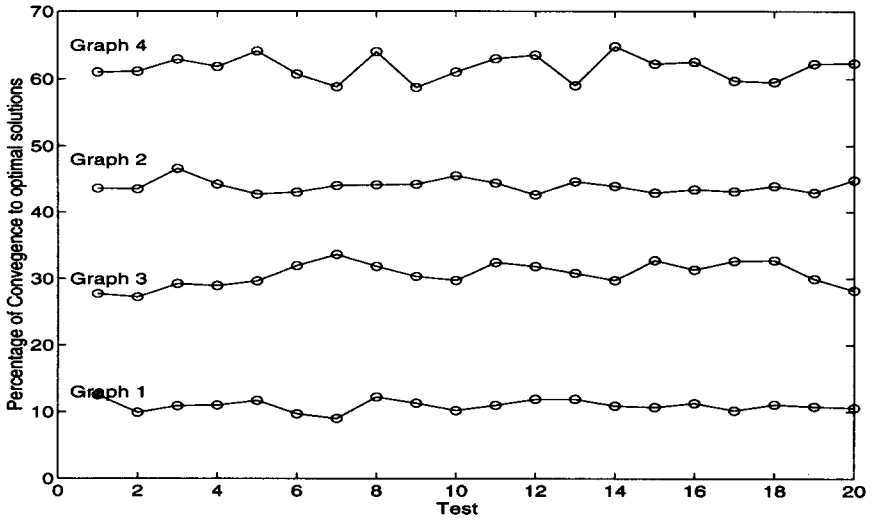


FIGURE 5a. Efficiency of symmetric Hopfield algorithm applied to find MISs in graphs 1–4.

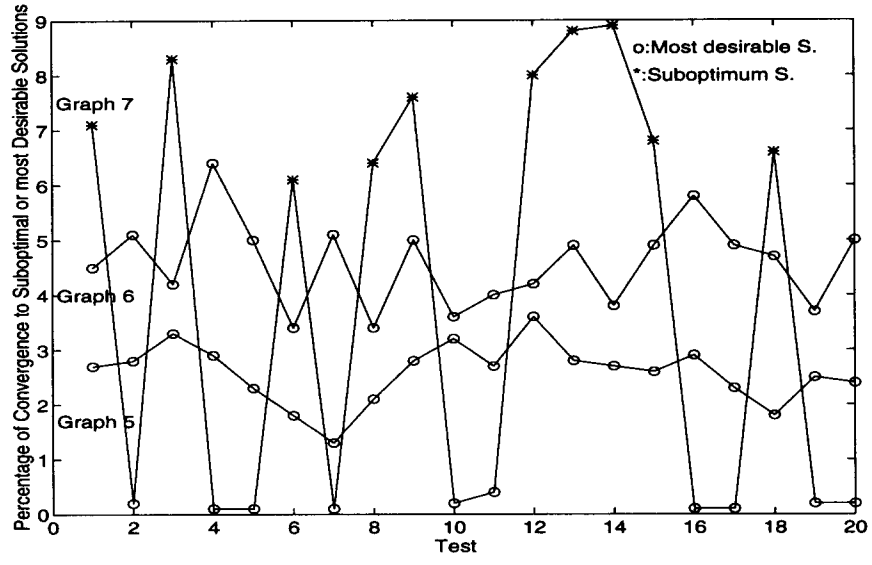


FIGURE 5b. Efficiency of symmetric Hopfield algorithm applied to find MISs in graphs 5–7.

works. The theorem not only generalizes the known results for symmetric networks, but also yields various new, easily verifiable tests for global convergence of asymmetric networks of the Hopfield type. Perhaps the most important contribution of this paper is in introducing a classification theory of the energy functions, based on which the problem of how a combinatorial optimization problem can be solved both reliably and efficiently by Hopfield networks is resolved.

The paper has mostly concentrated on studying asymmetric Hopfield-type networks performing optimization tasks, the theories developed, however, are by no means only significant for computation use. In a forthcoming paper, we will examine in detail the associative memory aspects of these networks as capacity and error-correcting capability. In particular, we will show that the classification theory for the energy functions established in this paper also plays an important role in improving the performance of existing symmetric associative memories.

It is well known that the dynamics of symmetric Hopfield-type networks can be understood in terms of the minimization of some scalar quantity (in equilibrium to be identified with the free energy). The analysis conducted in the present paper shows that this understanding is still valid for some general asymmetric Hopfield-type networks (i.e., asymmetric Hopfield-type networks still have their energy function in some cases).

REFERENCES

Aiyer, S. V. B., Niranjan, M., & Fallside, F. (1990). A theoretical investigation into the performance of the Hopfield model. *IEEE Transactions on Neural Networks*, 1. 204–215.

Amit, D., Gutfreund, H., & Sompolinsky, H. (1985). Spin-glass models of neural networks, *Physical Review A*, **32**, 1007–1018.

Bondy, J. A., & Murty, U. S. R. (1976). *Graph theory with applications*. London: The Macmillan Press.

Bruck, J., (1990). On the convergence properties of the Hopfield model. *Proceedings of the IEEE*, **78**, 1579–1585.

Bruck, J., & Goodman, J. W. (1988). A generalized convergence theorem for neural networks. *IEEE Transactions on Information Theory*, **34**, 1089–1092.

Bruck, J., & Blaum, M. (1989). Neural networks, error correcting codes and polynomials over binary *n*-cubes. *IEEE Transactions on Information Theory*, **35**, 976–987.

Bruck, J., & Roychowdhury, V. P. (1989). On the number of spurious memories in the Hopfield Model. *IEEE Transactions on Information Theory*, **36**, 393–397.

Chiueh, T. D., & Goodman, R. M. (1991). Recurrent correlation associative memories. *IEEE Transactions on Neural Networks*, **2**, 275–284.

Coolen, T., & Sherrington, D (1993). Dynamics of attractor neural networks. In J. G. Taylor (Ed.), *Mathematical approaches to neural networks*. Amsterdam: Elsevier.

Cottrel, M. (1988). Stability and attractiveness in associative memory networks. *Biological Cybernetics*, **58**, 129–139.

Crisanti, A., & Sompolinsky, H. (1987). Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. *Physical Review A*, **36**, 4922–4936.

Dahl, E. D. (1988). Neural network algorithms for an NP-complete problem: Map and graph coloring. *Proceedings of International Conference on Neural Networks*, (III, pp. 113–120).

Dasgupta, S., Ghosh, A., & Cuykendall, R. (1989). Convergence in neural memories. *IEEE Transactions on Information Theory*, **35**, 1069–1072.

Derrida, B., Gardner, E., & Zippelius, A. (1987). An exactly soluble asymmetric neural network model. *Europhysics Letters*, **4**, 167–173.

Goles, E. (1986). Antisymmetrical neural networks. *Discrete Applied Mathematics*, **13**, 97–100.

Goles, E., Fogelman, F., & Pellegrin, D. (1985). Decreasing energy functions as a tool for studying threshold networks. *Discrete Applied Mathematics*, **12**, 261–277.

Guez, A., Protopopescu, V., & Barhen, J. (1988). On the stability, storage capacity and design of nonlinear continuous neural networks. *IEEE Transactions on System, Man, Cybernetics*, **18**, 80–87.

Hertz, J. A., Grinstein, G., & Solla, S. (1986). Memory networks with asymmetric bonds. In J. S. Denker (Ed.), *Neural networks for computing* (pp. 212–218). New York: American Institute of Physics.

Hertz, J. A., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Reading, MA: Addison-Wesley.

Hinton, G. E., & Sejnowski, T. J. (1983). Optimal perceptual inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington 1983 (pp. 448–453). New York: IEEE.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational ability. *Proceedings of the National Academy of Sciences, USA*, **79**, 2554–2558.

Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, **52**, 141–152.

Kanter, I., & Sompolinsky, H. (1987). Associative recall of memory without errors. *Physical Review A*, **35**, 308–392.

Kohonen, T. (1974). An adaptive associative memory principle. *IEEE Transaction on computers*, **C-23**, 444–445.

Kohonen, T. (1989). *Self-organization and associative memory* (3rd ed.). Berlin: Springer-Verlag.

Kohring, G. A. (1993). On the Q-state neuron problem in attractor neural networks. *Neural Networks*, **6**, 573–581.

Kosko, B. (1988). Bidirectional associative memories. *IEEE Transactions on System, Man, Cybernetics*, **18**, 49–60.

Little, W. A. (1974). The existence of persistent states in the brain. *Mathematical Biosciences*, **19**, 101–120.

Mazumadar, P., & Yih, J. S. (1989). Neural computing for built in self repair of embedded memory arrays. *Proceedings of the 19th International Symposium on Fault Tolerant Computing*, pp. 480–487.

McEliece, R. J., Posner, E. C., Rodemich, E. R., & Venkatesh, S. S. (1987). The capacity of the Hopfield associative memory. *IEEE Transactions on Information Theory*, **33**, 461–482.

Parisi, G. (1986). Asymetric neural networks and process of learning. Journal of Physics, **19**, L675–L680.

Personnaz, L., Guyon, I., & Dreyfus, G. (1985). Information storage and retrieval in spin-glass-like neural networks. *Journal de Physiue Letters (paries)*, **46**, 359–365.

Porat, S. (1989). Stability and looping in connectionist models with asymmetric weights. *Biological Cybernetics*, **60**, 335–344.

Pramanick, I. (1991). *Parallel dynamic interaction—An inherently parallel problem solving methodology*. Ph.D. Thesis, Department of ECE, University of Iowa (May).

Shrivastava, Y., Dasgupta, S., & Reddy, S. M. (1992). Guaranteed convergence in a class of Hopfield networks. *IEEE Transactions on Neural Networks*, **3**, 951–961.

Sompolinsky, H. (1987). The theory of neural networks: The Hebb rules and beyond. In J. L. van Hemmen & I. Morgenstern (Eds.), *Heidelberg Colloquium on Glassy Dynamics* (Heidelberg 1986), (pp. 485–527). Berlin: Springer-Verlag.

Sompolinsky, H., & Kanter, I. (1986). Temporal association in asymmetric neural networks. *Physical Review Letters*, **57**, 2861–2864.

van Hemmen, J. L., & Kühn, R. (1986). Nonlinear neural networks. *Physical Review Letters*, **57**, 913–916.

Wilson, V., & Pawley, G. S. (1988). On the stability of the TSP problem algorithm of Hopfield and Tank. *Biological Cybernetics*, **58**, 63–70.

Xu, Z. B., & Kwong, C. P. (1995a). Global convergence and asymptotic stability of asymmetric Hopfield neural networks. *Journal of Mathematical Analysis and Applications*, **191**, 405–427.

Xu, Z. B., & Kwong, C. P. (1995b). A decomposition principle for complexity reduction of artificial neural networks. *Neural Networks* (in press).

Xu, Z. B., Leung, Y., & He, X. W. (1994). Asymmetric bidirectional associative memories. *IEEE Transactions on System, Man, Cybernetics*, **24**, 1558–1564.

Xu, Z. B., Hu, G. Q., & Kwong, C. P. (1996). Some efficient strategies for improving the eigenstructure method in synthesis of feedback neural networks. *IEEE Transactions on Neural Networks*, **7**, 233–245.

APPENDIX A: THE PROOF OF THEOREM 1

Let

$$\varepsilon_i = \inf\left\{ t_i - \sum_{i=1}^{n} w_{ij} v_j : v_j \in \{-1, 1\}, \sum_{j=1}^{n} w_{ij} v_j < t_i \right\},$$

$$\tilde{T} = T - \frac{1}{2}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\mathrm{T}} = (\tilde{t}_1, \tilde{t}_2, \ldots, \tilde{t}_n)^{\mathrm{T}} \tag{A.1}$$

and construct the energy function

$$E(V) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} v_i v_j + \sum_{i=1}^{n} \tilde{t}_i v_i + C \tag{A.2}$$

where C is an arbitrary constant.

Writing $E(V)$ in a matrix form $E(V) = -\frac{1}{2} V^T W V + V^T \tilde{T} + C$ and then expanding it about $V(t)$ into a Taylor series:

$$E(V) = E(V(t)) + (\Delta V)^T DE + \frac{1}{2}(\Delta V)^T D^2 E(\Delta V) \tag{A.3}$$

where

$$\Delta V = V(t+1) - V(t)$$
$$DE = -\frac{1}{2}(W + W^T) V(t) + \tilde{T},$$

and

$$D^2 E = -\frac{1}{2}(W + W^T).$$

Thus

$$\begin{aligned}\Delta E(t) &= E(V(t+1)) - E(V(t)) \\ &= (\Delta V)^T \left[-\frac{1}{2}(W + W^T) V(t) + \tilde{T}\right] \\ &\quad -\frac{1}{4}(\Delta V)^T (W + W^T)(\Delta V).\end{aligned}$$

With any parameter matrix $R = \text{diag}\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, this can be further written as

$$\Delta E(t) = -\frac{1}{2}(\Delta V)^T (R + I)[WV(t) - \tilde{T}] - \sigma(t) \tag{A.4)}$$

where

$$\begin{aligned}\sigma(t) &= \frac{1}{2}(\Delta V)^T \{-R(WV(t) - \tilde{T}] + W^T V(t) - \tilde{T} \\ &\quad -\frac{1}{2}(W + W^T)(\Delta V(t))\} \\ &= \frac{1}{2}(\Delta V)^T [(W^T - RW) V(t) + (R - I)\tilde{T} \\ &\quad +\frac{1}{2}(W + W^T)\Delta V(t)].\end{aligned} \tag{A.5}$$

We first prove that $-(\Delta V)^T (R + I)[WV(t) - \tilde{T}] < 0$. Let

$$\tilde{H}_i(t) = \sum_{j=1}^{n} w_{ij} v_j(t) - \tilde{t}_i = H_i(t) + \frac{t_i}{2}$$

where $H_i(t)$ is defined as in (1.2). Then, from (A.1) it is clear that $\tilde{H}_i(t) \geqslant \varepsilon_i/2$ whenever $H_i(t) \geqslant 0$, and $\tilde{H}_i(t) < -\varepsilon_i/2$ whenever $H_i(t) < 0$. Hence, from (1.1) and (1.2), and

$$\Delta v_i(t) = v_i(t+1) - v_i(t) = \begin{cases} 0, & v_i(t) = v_i(t+1) \\ -2, & v_i(t) = 1, v_i(t+1) = -1 \\ 2, & v_i(t) = -1, v_i(t+1) = 1 \end{cases} \tag{A.6}$$

we conclude that if $V(t+1) \neq V(t)$ we always have

$$\begin{aligned}&-(\Delta V)^T (R + I)[WV(t) - \tilde{T}] \\ &= \sum_{i=1}^{n} \Delta v_i(t)(1 + \alpha_i)\tilde{H}_i(t) < 0,\end{aligned} \tag{A.7}$$

which is the first term (of (A.4). For the term $\sigma(t)$ in (A.4), we expand (A.5) as follows.

Let

$$I_1(t) = \{i \in \{1, 2, \ldots, n\} : \Delta v_i(t) = 0\} \tag{A.8}$$

and

$$I_2(t) = \{i \in \{1, 2, \ldots, n\} : \Delta v_i(t) \neq 0\}. \tag{A.9}$$

We write

$$\begin{aligned}\sigma(t) &= \sum_{i=1}^{n} \Delta v_i(t) \left\{ \sum_{j=1}^{n} (w_{ji} - \alpha_i w_{ij}) v_j(t) \right. \\ &\quad \left. + (\alpha_i - 1)\tilde{t}_i + \frac{1}{2} \sum_{j=1}^{n} (w_{ij} + w_{ji}) \Delta v_j(t) \right\} \\ &= \sum_{i \in I_2(t)} \Delta v_i(t) \left\{ \sum_{j \in I_1(t)} (w_{ji} - \alpha_i w_{ij}) v_j(t) \right. \\ &\quad + (\alpha_i - 1)\tilde{t}_i + \sum_{j \in I_2(t)} \left[(w_{ji} - \alpha_i w_{ij}) v_j(t) \right. \\ &\quad \left. \left. + \frac{1}{2}(w_{ij} + w_{ji}) \Delta v_j(t) \right] \right\}.\end{aligned} \tag{A.10}$$

However, from (A.6), $v_j(t) = -\frac{1}{2}\Delta v_j$, $\forall_j \in I_2(t)$. Then, in (A.10),

$$\begin{aligned}&\sum_{j \in I_2(t)} \left[(w_{ji} - \alpha_i w_{ij}) v_j(t) + \frac{1}{2}(w_{ij} + w_{ij}) \Delta v_j(t) \right] \\ &= \frac{1}{2}(1 + \alpha_i) \sum_{j \in I_2(t)} w_{ij} \Delta v_j.\end{aligned}$$

Therefore (A.10) can be rewritten as

$$\begin{aligned}2\sigma(t) &= \sum_{i \in I_2(t)} \Delta v_i \left\{ \sum_{j \in I_1(t)} (w_{ji} - \alpha_i w_{ij}) v_j(t) \right. \\ &\quad \left. + (\alpha_i - 1)\tilde{t}_i + \frac{1}{2}(1 + \alpha_i) \sum_{j \in I_2(t)} w_{ij} \Delta v_j \right\} \\ &= \frac{1}{2} \sum_{i \in I_2(t)} \sum_{j \in I_2(t)} (1 + \alpha_i) w_{ij} \Delta v_i \Delta v_j \\ &\quad + \sum_{i \in I_2(t)} \Delta v_i \left\{ \sum_{j \in I_1(t)} (w_{ji} - \alpha_i w_{ij}) v_j(t) + (\alpha_i - 1)\tilde{t}_i \right\}.\end{aligned} \tag{A.11}$$

Since $|v_j(t)| = 1$, $|\Delta v_i| = 2$ for any $i \in I_2(t)$ and $I_1(t) \cap I_2(t) = \varnothing$, we have

$$\sum_{i\in I_2(t)} \Delta v_i \left\{ \sum_{j\in I_1(t)} (w_{ji} - \alpha_i w_{ij}) v_j(t) + (\alpha_i - 1)\tilde{t}_i \right\}$$

$$\geqslant -\sum_{i\in I_2(t)} \left\{ \sum_{j\in I_1(t)} |w_{ji} - \alpha_i w_{ij}||\Delta v_i(t)||v_j(t)| + |\alpha_i - 1||\tilde{t}_i||\Delta v_i(t)| \right\}$$

$$= -\frac{1}{2} \sum_{i\in I_2(t)} \left\{ \sum_{j\in I_1(t)} |w_{ji} - \alpha_i w_{ij}| + |\alpha_i - 1||\tilde{t}_i| \right\} (\Delta v_i)^2$$

$$\geqslant -\frac{1}{2} \sum_{i\in I_2(t)} \left\{ \sum_{j=1, j\neq i}^{n} |w_{ji} - \alpha_i w_{ij}| + |\alpha_i - 1||\tilde{t}_i| \right\} (\Delta v_i)^2.$$

Combined with (A.11), this implies

$$2\sigma(t) \geqslant \frac{1}{2} \sum_{i\in I_2(t)} \sum_{j\in I_2(t)} (1+\alpha_i) w_{ij} \Delta v_i \Delta v_j - \frac{1}{2} \sum_{i\in I_2(t)} \left\{ \sum_{j=1, j\neq i}^{n} |w_{ji} - \alpha_i w_{ij}| + |\alpha_i - 1||\tilde{t}_i| \right\} (\Delta v_i)^2$$

$$\geqslant \frac{1}{2} \sum_{i\in I_2(t)} \Delta v_i(t) \left\{ \sum_{j\in I_2)t)} (1+\alpha_i) w_{ij} \Delta v_j(t) + \left(\sum_{j=1, j\neq i}^{n} |w_{ji} - \alpha_i w_{ij}| + |\alpha_i - 1||\tilde{t}_i| \right) \Delta v_i(t) \right\}$$

$$\geqslant \frac{1}{2} \sum_{i\in I_2(t)} \sum_{j\in I_2(t)} w_{ij}^* \Delta v_i(t) \Delta v_j(t). \qquad \text{(A.12)}$$

When N is operating in the serial mode, $I_2(t) = i$ and $I_1(t) = \{1, 2, \ldots, n\}/\{i\}$. In this case,

$$2\sigma(t) \geqslant \frac{1}{2} w_{ii}^* [\Delta v_i(t)]^2 = \frac{1}{2} \left\{ (1+\alpha_i) w_{ii} - \left(\sum_{j=1, j\neq i}^{n} |w_{ji} - \alpha_i w_{ij}| + |\alpha_i - 1||\tilde{t}_i| \right) \right\} \geqslant 0.$$

For the parallel mode operation, if W^* is positive semidefinite, its principal submatrices are positive semidefinite. This gives

$$\sum_{i\in I_2(t)} \sum_{j\in I_2(t)} w_{ij}^* \Delta v_i(t) \Delta v_j(t) \geqslant 0$$

and hence we obtain from (A.12) that $\sigma(t) \geqslant 0$. Together with (A.7), we have proved that, under the said assumptions, (A.4) is nonnegative, i.e., $E(V(t+1)) < E(V(t))$ when $V(t+1) \neq V(t)$. This completes the proof of Theorem 1.

APPENDIX B: THE PROOF OF THEOREM 2

(i) Let $V^* = (v_1^*, v_2^*, \ldots, v_n^*)^T$ be a local minimizer of E, i.e., $E(V^*) \leqslant E(V)$ for any $V \in B_H(V^*, 1)$. We need to show that $v_i^* = \text{sgn}\{H_i(V^*)\}$ for any $i = 1, 2, \ldots, n$, where

$$H_i(V^*) = \sum_{j=1}^{n} w_{ij} v_j^* - t_i.$$

For any fixed k, we let $V^{(k)}$ be the vector whose kth component differs from v_k^* and other components are the same with that of V^*. Then, clearly, $V^{(k)} \in B_H(V^*, 1)$ and we can write $V^{(k)} = V^* + \Delta V^*$ with $\Delta V^* = (0, \ldots, 0, -2v_k^*, \ldots, 0)^T$. Thus, we have

$$E(V^{(k)}) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} [v_i^* + (\Delta V^*)_i][v_j^* + (\Delta V^*)_j] + \sum_{i=1}^{n} t_i [v_i^* + (\Delta V^*)_i] + C$$

$$= E(V^*) - \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} v_i^* (\Delta V^*)_j + \sum_{j=1}^{n} \sum_{i=1}^{n} w_{ji} (\Delta V^*)_i v_j^* + \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (\Delta V^*)_i (\Delta V^*)_j \right) + \sum_{i=1}^{n} t_i (\Delta V^*)_i$$

$$= E(V^*) - \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij} + w_{ji}) (\Delta V^*)_i v_j^* - 2 \sum_{i=1}^{n} t_i (\Delta V^*)_i \right) - 2w_{kk}$$

$$= E(V^*) + v_k^* (H_k(V^*) + \left(\sum_{j=1}^{n} w_{jk} v_j^* - t_k \right) - 2w_{kk}.$$

This implies

$$E(V^{(k)}) - E(V^*) = (\alpha_k + 1) v_k^* H_k(V^*) + v_k^* \left(\sum_{j\neq k}^{n} (w_{jk} - \alpha_k w_{kj}) v_j^* + (\alpha_k - 1)(t_k) \right) - (1+\alpha_k) w_{kk}.$$

Since V^* is a local minimizer of E, it then follows that

$$(\alpha_k + 1) v_k^* H_k(V^*) = E(V^{(k)}) - E(V^*) + (1+\alpha_k) w_{kk} - v_k^* \left(\sum_{j\neq k}^{n} (w_{jk} - \alpha_k w_{kj}) v_j^* + (\alpha_k - 1)(t_k) \right).$$

$$\geqslant (1+\alpha_k) w_{kk} - \left(\sum_{j=1, j\neq i}^{n} |w_{jk} - \alpha_k w_{kj}| + (\alpha_k - 1 \parallel t_k|) \right). \qquad \text{(B.1)}$$

From (B.1) and by using the assumptions (3.3) and $\alpha_k > -1$, we now conclude that $v_k^* H_k(V^*) \geqslant 0$, i.e., $v_k^* = \text{sign}\{H_k(V^*)\}$. As k is arbitrarily fixed, this completes the proof of (i).

(ii) Assume that $V^* = (v_1^*, v_2^*, \ldots, v_n^*)^T$ is a stable state of N, i.e., $v_k^* H_k(V^*) \geqslant 0$ for any $i \in \{1, 2, \ldots, n\}$. We need to show $E(V^*) \leqslant E(V)$ for any $V \in B_H(V^*, 1)$.

For any fixed $V \in B_H(V^*, 1)$, we can write $V = V^* + \Delta V^*$ with $\Delta V^* = (0, \ldots, 0, -2v_k^*, \ldots, 0)^T$ for some k in $\{1, 2, \ldots, n\}$. Therefore, as in the proof of part (i), we can obtain

$$E(V) - E(V^*) = (1+\alpha_k) v_k^* H_k(V^*) + v_k^* \left(\sum_{j\neq k}^{n} (w_{jk} - \alpha_k w_{kj}) v_j^* + (\alpha_k - 1)(t_k) \right) - (1+\alpha_k) w_{kk}.$$

Since $v_k^* H_k(V^*) \geqslant 0$ for any $i \in \{1, 2, \ldots, n\}$, this implies

$$E(V) - E(V^*) \geqslant v_k^* \left(\sum_{j \neq k}^{n} (w_{jk} - \alpha_k w_{kj}) v_j^* + (\alpha_k - 1)(t_k) \right)$$

$$- (1 + \alpha_k) w_{kk}$$

$$\geqslant - \left(\sum_{j=1, j \neq i} |w_{jk} - \alpha_k w_{kj}| + |\alpha_k - 1 \parallel t_i| \right)$$

$$- (1 + \alpha_k) w_{kk}$$

$$= (1 + \alpha_k)(-\sigma(\alpha_k) - w_{kk}).$$

By (3.4), $E(V^*) \leqslant E(V)$ then follows. This finishes the proof of (ii).

(iii) Since there is an R such that $\sigma(\alpha_i) = 0$ for any i, the conditions (3.3) and (3.4) are simultaneously met. Hence (iii) follows directly from (i) and (ii).

APPENDIX C: THE PROOF OF THEOREM 3

Let $U^* = (u_1^*, u_2^*, \ldots, u_n^*)$ be any MIS of a graph G and $U = (u_1, u_2, \ldots, u_n)$ be any subset of the vertices of G. Denote, respectively, the number of non-zero components of U^* and U by r and $\tilde{r}$. Then we can write $\tilde{r} = r_1 + r_2$, where r_1 is the number of vertices in U that belong to an independent set. Clearly we have $r_1 \leqslant r$ and $E(U^*) = -r^2$.

We need to show that $E(U) \geqslant E(U^*)$ for any U. It is however calculated that

$$E(U) = -(\tilde{r})^2 + \frac{K}{2}(r_2)^2 = -(r_1)^2 - 2r_1 r_2 + \left(\frac{K}{2} - 1\right)(r_2)^2.$$

The inequality follows directly whenever $K \geqslant 2(2n + 1)$. This then completes the proof of Theorem 3.