Chaotic annealing for optimization

Chang-song Zhou[1] and Tian-lun Chen[2,1]
[1]*Department of Physics, Nankai University, Tianjin 300071, China*
[2]*CCAST (World Laboratory), Beijing, 100080, China*
and Institute of Theoretical Physics, Academia Sinica, P.O. Box 2735, Beijing, 100080, China
(Received 1 July 1996; revised manuscript received 18 October 1996)

We study the effect of chaotic transient for a global minima search in optimization. For a given energy or cost function, a chaotic evolution system in which chaos provides a scheme for searching the minima of the energy function in the state space can be constructed generally. By controlling a bifurcation parameter from the chaotic dynamics regime to the fixed-point regime gradually the system may eventually reach the global optimum state or its good approximation with very high probability. A double potential well and a traveling salesman problem are used to numerically illustrate the validity of chaotic transient search.
[S1063-651X(97)14603-7]

PACS number(s): 05.45.+b, 82.20.Mj, 43.72.+q

I. INTRODUCTION

In resent years, the roles or applications of complex dynamics, especially of chaos, have drawn great attention in many fields. One is chaos controlling [1] and synchronization [2]. Another field may be the biological implications and potential applications of chaos in neural networks. Many chaotic neural network models have been proposed [3–8]. It has been pointed out that chaos may play important roles in the information processing of neural networks. For example, Freeman suggested that chaos is necessary for a rabbit to memorize new odors [3]; Tsuda suggested that cortical chaos may serve for dynamically linking true memory and a memory search [5]; Nara *et al.* showed that complex dynamics of a recurrent neural network has promising efficiency in a complicated memory search [6].

Some researchers have considered the applications of chaos to global minima search in optimization. In Ref. [7], the authors presented a chaotic neurocomputer model in which each neuron is composed of two coupled logistic oscillators, and the neurons are globally coupled with synaptic connections as in the Hopfield model. The authors showed that during the evolution of the chaotic neurocomputer, some good approximate solutions of a 10 random city traveling salesman problem (TSP) can be visited. Recently, Hoyakawa *et al.* studied the performance of the ordinary Hopfield neural network driven by external chaotic noise, pointing out that short time correlation of chaos could work effectively for a global minima search [8]. In these models, chaos served as drive schemes, which enabled the system state to wander interminably in the phase space.

Compared with conventional methods for optimization, using chaos as a global minima search is just at its beginning stage. It is important at this stage to study some simple models. In Ref. [8], external chaotic noise is always present during the evolution of the network. It seems that the average amplitude of the chaotic noise may also play an important role in the performance of the network, which is not studied in Ref. [8]. In this paper, we consider such a question: what will happen if chaotic dynamics employed to search the state space vanishes gradually so that the system is reduced back to the ordinary stable system? Corresponding to the model in Ref. [8], the question is, what will happen if the amplitude of the external chaotic noise is gradually decreased? To examine this question, we present a method for constructing chaotic systems by introducing a simple nonlinear feedback into the gradient descent systems designed for optimization tasks. Chaotic dynamics vanish gradually when a bifurcation parameter of the nonlinear feedback is gradually decreased according to some scheme, and the chaotic system is reduced back to the original descent one; i.e., reaching a fixed state eventually. Numerical simulations show that with such a scheme, the system will converge to the global minimum or its good approximations with high probability.

In Sec. II, we show how a chaotic evolution system can be constructed generally for a given energy function. A simple double potential well is used to illustrate the global minimum search process in Sec. III. In Sec. IV, an energy function of a Hopfield neural network for solving a traveling salesman problem is used as a practical example. Section V is devoted to some discussions and outlooks.

II. DESCRIPTION OF THE MODEL

For simplicity, we discuss the method with only the one-dimensional energy function. Extending the method to multidimensional cases is straightforward, as will be shown in Sec. IV.

For a given energy function $E(x)$, we can construct an evolution system possessing gradient descent dynamics as

$$x(t+1)=f(x(t)). \tag{1}$$

A well-known choice of $f(x)$, for example, can be $f(x)=(1-\epsilon)x-\epsilon dE(x)/dx$ with sufficiently small ϵ. The system will finally reach a fixed point state x_F, which is one of the minima of $E(x)$, but is not capable of escaping from it. To reach the global minima, a mechanism that allows escaping from local minima is required. In [8], external chaotic noise is used to excite the state and kick it out of local minima. Our aim here is to introduce some simple new ingredient into the system, which allows the system to approach or visit the

local minima, and at the same time enables the system to escape from them. To destabilize the fixed points, a nonlinear feedback is introduced into the system, and the evolution becomes

$$x(t+1)=f(x(t))+g[x(t)-x(t-1)], \tag{2}$$

where the feedback is switched on at $t=1$ so that only $x(0)$ is sufficient to initialize the system. g is a nonlinear function satisfying the following demands: (1) It does not change the original fixed points of Eq. (1), namely, $g(0)=0$. (2) $y(t)=x(t)-x(t-1)$ can be regarded as the speed at which the system tends to a fixed point at time t. Large $y(t)$ implies that the system is far away from a fixed point (local minimum), and the system is expected to approach a fixed point almost in a gradient descent way in such a case, which demands that $g(y(t))$ has only a small perturbation on the system at large $y(t)$; i.e., $g(y(t))$ decreases rapidly to 0 at large $y(t)$. (3) At the intermediate value of $y(t)$, i.e., when the system comes into some close neighborhood of a local minimum, the system gets large drive from $g(y(t))$, which may enable it to climb over the energy hills and drop into other energy valleys.

In this paper, $g(y(t))$ is taken as

$$g(y(t))=py(t)\exp[-|y(t)|], \tag{3}$$

where p is a positive tunable parameter.

In fact, Eq. (2) can be rewritten as

$$x(t+1)=f(x(t))+g(y(t)), \tag{4}$$

$$y(t+1)=x(t+1)-x(t), \tag{5}$$

a two-dimensional evolution system. The stability of a fixed point $(x_F,0)$ of this system is governed by the Jacobian matrix at the fixed point, namely,

$$DF(x_F)=\begin{pmatrix} B & p \\ B-1 & p \end{pmatrix}, \tag{6}$$

where $B=[df(x)/dx]_{x=x_F}$. The two eigenvalues $\lambda_\pm$ of $DF(x_F)$ are determined by the characteristic equation

$$\lambda^2-(B+p)\lambda+p=0. \tag{7}$$

Denoting $D=(B+p)^2-4p$, we have

$$m_\pm=|\lambda_\pm|=\begin{cases} \left|\dfrac{(B+p)\pm\sqrt{D}}{2}\right|, & D\geqslant 0 \\ \sqrt{p}, & D<0. \end{cases} \tag{8}$$

A fixed point is stable if both $m_\pm<1$ and is unstable if both $m_\pm>1$. It is nonstable (a saddle point) if one of the $m_\pm$ is larger than 1 and the other is less than 1. For a stable fixed point x_F of Eq. (1), it must be true that $|B|<1$. Under this restriction, it is always true that $m_\pm<1$ for $0\leqslant p<1$, so that the fixed point will remain stable for $p<1$. Since $D<0$ at $p=1$ and the resulted $\lambda_\pm$ are a complex-conjugate pair, the bifurcation occurring at $p=1$ ($m_\pm=1$) is a Hopf bifurcation. The fixed point becomes unstable when p is larger than 1.0. It should be noted that the above analysis is true for any stable fixed point of any one-dimensional equilibrium system

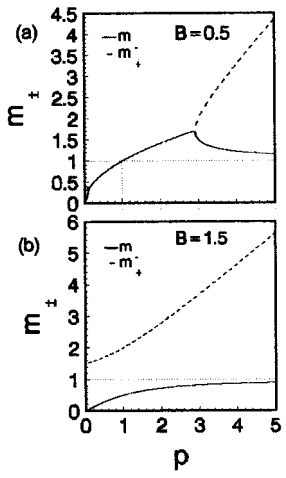

FIG. 1. Magnitude of the eigenvalue as a function of p.

$x(t+1)=f(x(t))$. As for a local maximum of $E(x)$ corresponding to an unstable fixed point of Eq. (1), it is true that $d^2E/dx^2\leqslant 0$ and $B>1$, which result in $m_+>1$ and $m_-<1$ for $p>0$. So an unstable fixed point of Eq. (1) becomes a saddle node of Eqs. (4) and (5). The added nonlinear feedback does not introduce new fixed points into the system. To illustrate the above analysis, $m_\pm$ are plotted as functions of p, with $B=0.5$ in Fig. 1(a) and $B=1.5$ in Fig. 1(b).

The dynamical behavior far from a fixed point, however, depends on the specific form of $f(x)$, and can be investigated with some numerical methods, such as calculating the bifurcation diagrams, the Lyapunov exponents, or the correlation dimension. Generally, when p is large enough, the system obtains the ability to wander in the state space, and the accessible region is larger with larger p, which will be demonstrated by examples in next sections.

Unlike in Ref. [8], where the short time correlation of *external* chaotic noise is employed to kick the system out of local minima when it is trapped, in our model, *internal* change of the system state is employed in such a way that it enables the system to access and escape from local minima. A physical image of the model is that the motion of a particle (or particles) is governed by a potential E as well as the previous momentum, for example, a particle in a potential field and a nonlinear adhesive medium, although the form of nonlinearity may not have physical realization.

However, the system is not allowed to wander in the state space interminably, but is reduced back to the stable dynamical system by decreasing p gradually according to some scheme, for example,

$$p(t+1)=p(1)/\ln(t), \quad t=2,3,4,\ldots . \tag{9}$$

In this paper, we will show that with this scheme the system can escape from most of the local minima of the energy and reach the global minimum or its good approximation with very high probability.

One can see that $g(y(t))$ plays a similar role of the temperature noise in simulated annealing (SA) [9,10], with p being the counterpart of the temperature T. Due to the similarity to SA, with chaotic search taking the place of stochastic search, our method can be referred to as chaotic annealing (CA), a term used in the following discussion.

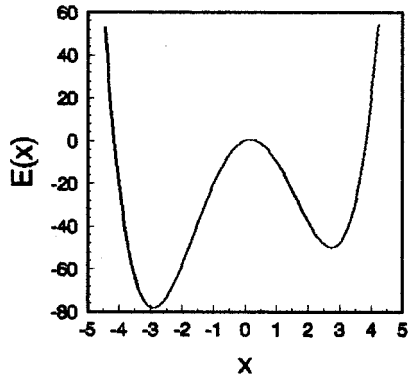

FIG. 2. A double well potential.

III. ONE-DIMENSIONAL EXAMPLE

To show how this method works, in this section, we choose a one-dimensional double well potential as the energy function [10],

$$E(x)=x^4-16x^2+5x, \tag{10}$$

as shown in Fig. 2.

The chaotic system is built as

$$x(t+1)=(1-\epsilon)x(t)-\epsilon[4x(t)^3-32x(t)+5] +g[x(t)-x(t-1)]. \tag{11}$$

We want to illustrate (1) the dynamical structure of the system and (2) the process of chaotic annealing and its performance of optimal optimization.

Let $\epsilon=0.01$ in all the simulations. To investigate the dynamical structure of the system, we calculate a bifurcation diagram and the largest Lyapunov exponent λ with respect to p, as shown in Figs. 3(a) and 3(b), respectively, with the same initial state $x(0)=1.0$ for each value of p. As expected, a Hopf bifurcation can be clearly detected at $p=1$. Chaos occurs in several regions where $\lambda>0$. Noting that when p is larger than some value $p^*(\approx 2.4)$, the state begins to wander

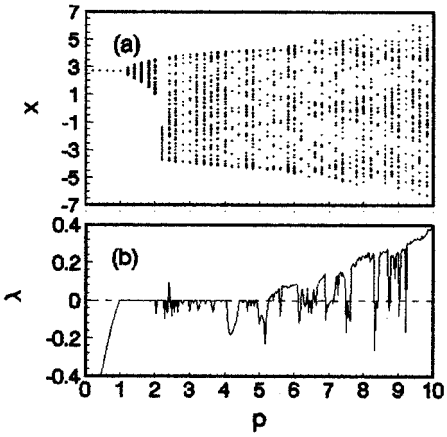

FIG. 3. (a) A bifurcation diagram of x against p. (b) The largest Lyapunov exponent λ.

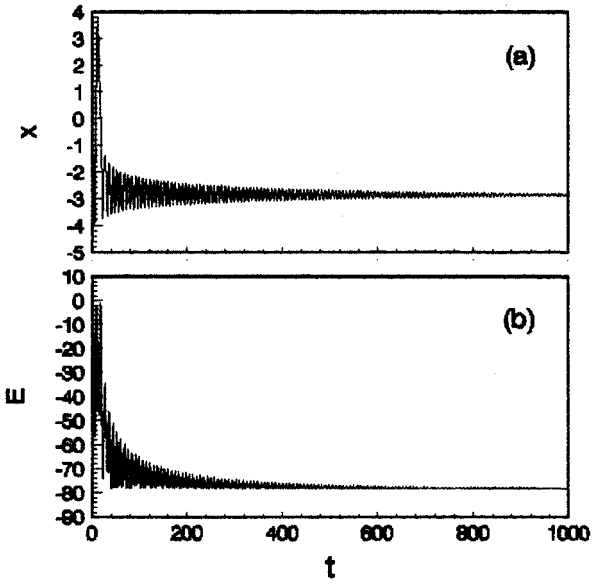

FIG. 4. Illustration of chaotic annealing process with $p(1)=7.5$. (a) x, (b) E.

(periodically, quisperiodically, or chaotically) between the two energy wells. Such region of p ($p>p^*$) is regarded as (chaotic) wandering region.

Now we choose a p in the wandering region and let it decrease according to Eq. (9). From an initial position in the shallower well the system can escape this local minimum and reach the deeper one, as illustrated by Fig. 4, an example of the chaotic annealing process with $p(1)=7.5$ and $x(0)=4$. Before being stable at the fixed point, the state of the system sweeps through the chaotic region of p. In this sense, the transient process preceding the stable behavior can be considered as a chaotic transient, a term used in later discussion.

Now we examine the general performance of chaotic annealing. We choose 1000 random initial states $x(0)$ uniformly distributed on $[-5,5]$. It is found that 52.2% of them approach the deeper minimum at $p(1)=0$ (gradient descent dynamics), with an average of 9 time steps to satisfy $\Delta E=|E(x(t))-E(x(t-1))|\leq 0.001$. Now chaotic annealing is carried out, employing the decreasing schedule Eq. (9) from different $p(1)$, and 1000 random initial states within $[-5,5]$ for each value of $p(1)$. The results are displayed in Fig. 5. In Fig. 5(a), P_R, the probability of converging to the global minimum, increases approximately with increasing $p(1)$, and reaches 1.0 when $p(1)$ is larger than about 7.5. Figure 5(b) is the corresponding plot of the largest Lyapunov exponent λ, which is computed by fixing $p=p(1)$, but not decreasing as Eq. (9). It is very interesting that the plot of P_R resembles that of λ. As seen in the plots, a $p(1)$ in the quasiperiodic regions [$\lambda=0$, e.g., $1.0<p(1)<2.4$, $3.5\leq p(1)\leq 4.0$] seems not a good choice for global optimization, because P_R is not significantly improved. Unlike the results in Ref. [8], which showed that an external drive from periodic windows of the logistic map is not a good candidate for kicking the system out of local minima, here a $p(1)$ in periodic windows [$\lambda<0$, e.g., $p(1)=2.5$, 2.75, 4.25, 4.75, 5.0, 6.5, 7.0, 7.5] seems to be conducive to a global minimum search. The region

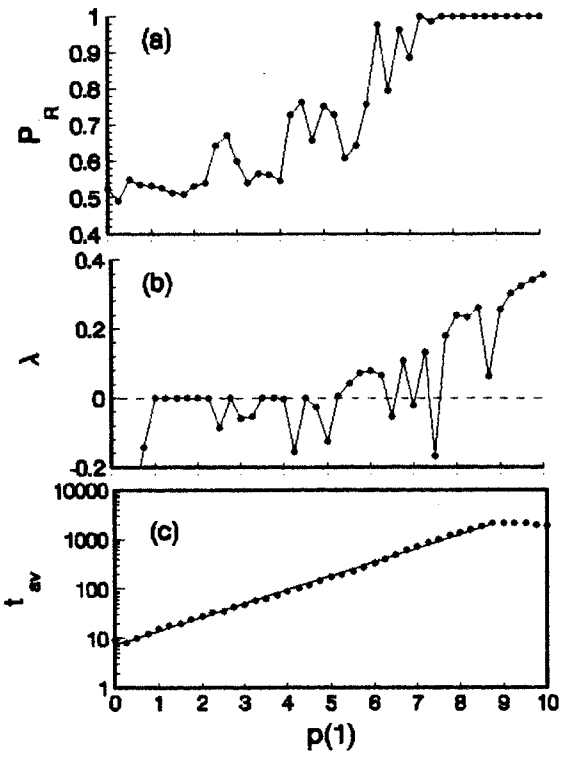

FIG. 5. (a) P_R, the probability of converging to the global minimum at different $p(1)$. (b) The largest Lyapunov exponents when p is fixed at $p(1)$. (c) The average time steps for the system to converge, which follows approximately an exponential law $t_{av}=7.3\exp[0.65p(1)]$.

$p(1)>6$ is the most interesting. It is clearly seen in the region $6<p(1)\leqslant 7.5$ that a $p(1)$ in the chaotic regime is better for global optimization than that in periodic windows. P_R is finally improved to 1 when $p(1)>7.5$ so that the system becomes rather chaotic.

The convergence rate of chaotic annealing is mainly governed by $p(1)$. As shown in Fig. 5(c), the average time steps t_{av} for the system to converge ($\Delta E\leqslant 0.001$) is approximately $t_{av}=7.3\exp[0.65p(1)]$, which is nearly the time steps needed for $p(t)$ to reach the fixed point regime ($p<1.0$).

This example of simple energy landscape has shown that the method has very high probability to approach the global minimum as long as $p(1)$ is sufficiently large, which is also similar to the requirement of sufficiently large initial temperature T in SA. How will this method perform if the energy landscapes are quite complex? This question is studied in the next section.

IV. EXAMPLE OF COMPLEX ENERGY LANDSCAPE

In this section we apply the chaotic annealing to a Hopfield neural network designed to solve a TSP. The energy landscape is much more complex and the dimension of the system is much higher.

A. The traveling salesman problem

The traveling salesman problem is that given N cities, a salesman is expected to find the shortest closed tour, visiting each of the N cities once and only once. Suppose that the N cities lie within a unit square. $L(i,j)$ is the link between the ith city and the jth city [$L(i,j)$ and $L(j,i)$ are considered the same], and D_{ij} is the distance between them. We use the following coding scheme [11]: each neuron V_{ij} $(i<j)$ corresponds to a link $L(i,j)$, and $L(i,j)$ is taken in the solution if $V_{ij}=1$, while it is not if $V_{ij}=0$. A possible energy function is

$$E=E_1+AE_2, \tag{12}$$

where

$$E_1=\sum_{i=1}^{N}\left(\sum_{j=i+1}^{N}V_{ij}+\sum_{j=1}^{i-1}V_{ji}-2\right)^2, \tag{13}$$

$$E_2=\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}V_{ij}D_{ij}. \tag{14}$$

$E_1=0$ assures that each city has two neighbors in the solution; E_2 is the cost (length) of the solution. However, $E_1=0$ cannot guarantee that a solution is a feasible tour (one cycle tour). Subtours consisting of several cycles also satisfy $E_1=0$. It is an essential difficulty of this coding scheme, and there is no simple and practical constraint that can be easily expressed in a neural network. However, some other algorithms can be introduced to merge several cycles obtained by the neural network into a single cycle tour [11], which is not done in this paper. An advantage of this coding scheme is that the number of neurons needed is $N(N-1)/2$, while it is N^2 in the Hopfield method [12].

We derive the synaptic connections and external inputs of the neural network by comparing the energy function Eq. (12) with the general energy expression of the Hopfield neural network [12],

$$E=-(1/2)\sum_{ij}\sum_{kl}W_{ijkl}V_{ij}V_{kl}-\sum_{ij}I_{ij}V_{ij}. \tag{15}$$

The evolution equations of the chaotic neural network are obtained as

$$\begin{aligned}U_{ij}(t+1)=&-2\left[\sum_{k>i}V_{ik}(t)+\sum_{k<i}V_{ki}(t)+\sum_{k>j}V_{jk}(t)\right.\\&\left.+\sum_{k<j}V_{kj}(t)\right]-AD_{ij}+8\\&+g[U_{ij}(t)-U_{ij}(t-1)],\end{aligned} \tag{16}$$

$$V_{ij}(t+1)=\tfrac{1}{2}\{1+\mathrm{sgn}[U_{ij}(t+1)]\}, \tag{17}$$

where U_{ij} is the local field of neuron V_{ij}.

In this paper, all the simulations are carried out with a 10-city problem used in [6]. Figure 6 shows the city distribution and its optimal path, whose length is $L_{op}=2.735$.

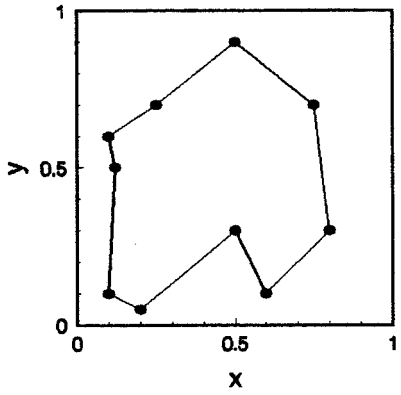

FIG. 6. The ten random cities and the shortest path $L_{\mathrm{op}}=2.735$.

Since our purpose in this paper is to show the effect of the chaotic transient in a global minima search, we fix $A=2$ in the following simulations, while we change $p(1)$ and compare the results with that of the Hopfield network.

B. Dynamical property of the chaotic neural network

The system now has a very high dimensionality of $N(N-1)=90$. So it is not easy to examine the dynamical property of the system analytically. We employ some numerical approaches to characterize the dynamical structure. Firstly, the bifurcation diagrams of the energy E and the local field of a neuron U_{12} against p are plotted in Fig. 7. For each value of p, the system starts from the same random initial condition and the first 1000 steps are discarded as a transient process. It is seen that when p is larger than $p^*\approx 0.95$, the system begins to wander in the state space. p^* is somewhat dependent on the initial conditions of the network. The appearance of the smeared bifurcation diagram itself is not proof of chaos. To characterize the dynamics further, at a certain p, let the system run for 50 000 steps, and record a time series of $S(i)=U_{12}(t)$ for the following $n=20\,000$ steps. A return map of such a time series is plotted in Fig. 8(a) for $p=4$. With this time series, the correlation dimension D_2 is calculated using the Grassberger-Procaccia method [13] in different embedded space R^M. The correlation function

$$C(r)=\frac{1}{n^2}\sum_{i,j} H[r-|\mathbf{S}(i)-\mathbf{S}(j)|] \tag{18}$$

where $H(x)=0$ if $x\leqslant 0$ while $H(x)=1$ if $x>0$, and $\mathbf{S}(i)=[S(i),S(i+1),\ldots,S(i+M-1)]$, is plotted with respect to r in Fig. 8(b). The slope of a plot is D_2 in the corresponding embedded space, which reaches saturation of about 1.613 for $M>1$. A fractal correlation dimension $D_2=1.613\pm 0.006$ demonstrates that the wandering orbit is chaotic but not periodic or quasiperiodic. We have also computed D_2 at other p values, for example, $D_2=1.268\pm 0.002$ for $p=1$, $D_2=1.491\pm 0.005$ for $p=2$, and $D_2=1.568\pm 0.004$ for $p=3$. Again the region $p>p^*$ is called the (chaotic) wandering region. During the wandering, the system explores the minima of the energy temporally and resides at them for one or several time steps. A similar temporal pattern process is possessed in many models [4,5,7], which may relate to chaotic itinerancy [5].

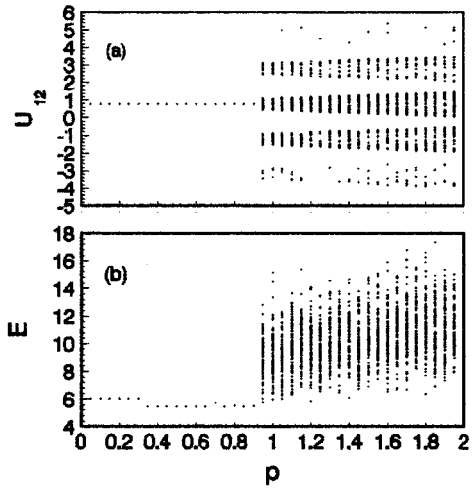

FIG. 7. A bifurcation diagram of a chaotic neural network with respect to p. (a) U_{12}, the local field of neuron V_{12}. (b) The energy E of the network.

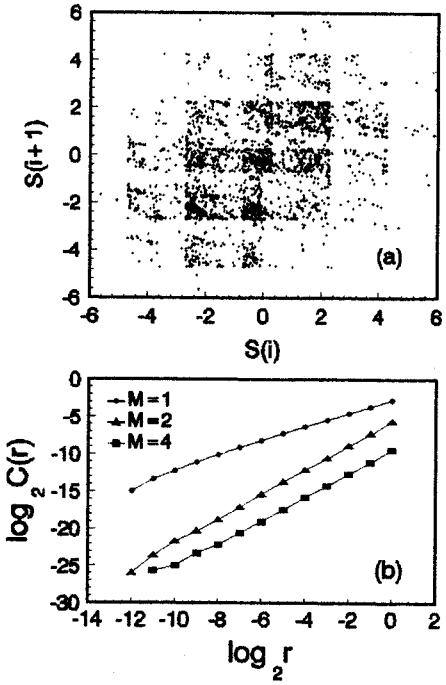

FIG. 8. (a) The return map $S(i)\sim S(i+1)$ for $p=4$; (b) the log-log plots of correlation function $C(r)$ with respect to r in different embedded space R^M.

C. Searching with chaotic transient dynamics

Now we begin to examine the performance of chaotic annealing on this complex energy landscape. Figure 9 is an example of the chaotic annealing process from $p(1)=5$. As seen in this figure, when p is gradually decreased according

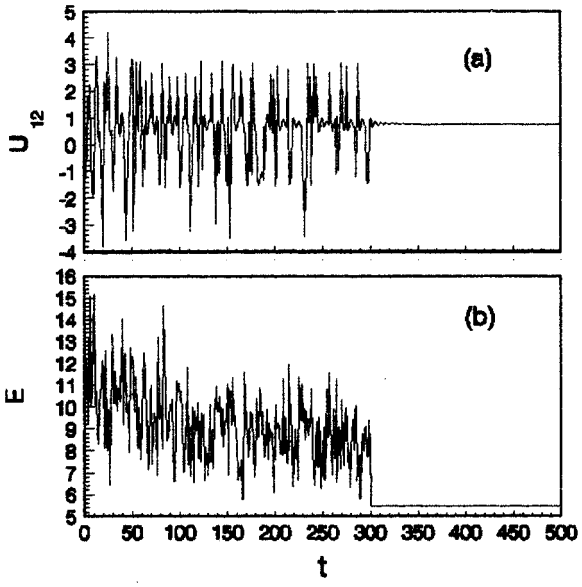

FIG. 9. Illustration of chaotic annealing process of the neural network with $p(1)=5$. (a) U_{12}, (b) E. The resulting stable state is the optimal minimum.

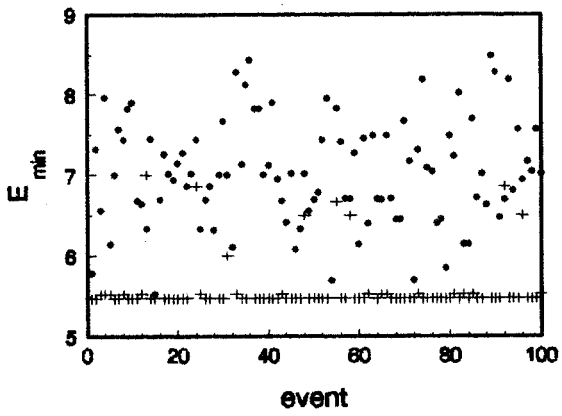

FIG. 10. Minimal energy $E_{\min}$ obtained with 100 random initial conditions. Dots for Hopfield network $[p(1)=0]$ and plusses for chaotic annealing with $p(1)=5$. Most of the minimal energy obtained by chaotic annealing are the global minimum.

to Eq. (9), the system wanders chaotically in the state space at first, visiting and escaping some local minima of the energy; at last it comes to a deep minimum (here it is the optimal one), from which it cannot escape and will stay there indefinitely. Before being stable at the optimal minimum, the system spends 45 time steps in local minima, visiting and escaping 39 local minima (in some local minima, the system stays successively for 2 time steps). Here a local minimum is referred to a state that has $E_1=0$, because when the system comes to such a state, it will stay there if p is set to 0 (i.e., it is a stable state of the Hopfield network). This process of searching for the minima of E is due to the special properties of $g(y(t))$, which allow the system to draw near a fixed point, but may drive it away when coming into some region around the fixed point. For $p(1)$ in the fixed point region $(p<p^*)$, this process is similar, so we do not examine this region separately, and also call the process as chaotic annealing even though p does not start from the (chaotic) wandering region.

In our next simulation, we start with 100 random initial conditions. The performances of the descent dynamics of the Hopfield network $[p(1)=0]$ and the chaotic annealing are compared by plotting the resulting minimal energy in Fig. 10. The improvement of the system performance by chaotic annealing is not trivial. For a Hopfield network $[p(1)=0]$, only 66 of the minima are feasible tours, and none of them is the optimal one. While for $p(1)=5$, all 100 minima are feasible tours, with 76 of them being the optimal one and 16 of them the second optimal one; among the other 8 minima only 2 of them have larger energy than those of $p(1)=0$. The result demonstrates that chaotic transient dynamics work much more efficiently than the descent dynamics of the Hopfield model to search for the optimal minimum or its good approximations.

In the following, we investigate the performance of the chaotic annealing with respect to $p(1)$. Several measures are used to characterize the performance. As has been pointed out, the minima of the energy E may not be feasible tours, our first measure is the probability of finding a feasible tour in a simulation. Another measure is the probability of finding the optimal tour in the obtained feasible tours. In SA, the convergence to global minima can be proved analytically in some cases [10], but it might not be guaranteed for CA. However, in practice, it may not be very fruitful to search the absolute optimum, and it may be better to find some good approximations of the absolute optimum in the period of time available, because these approximations and the absolute optimum may not be significantly different [9]. For example, in the present city distribution, the second optimal tour is $L=2.765$, only 1.09% worse than the optimal one, $L_{\rm op}=2.735$. Based on such considerations, the third measure used is the probability of finding a solution that is worse than the optimal one by a certain percentage describing the satisfaction of a solution in real optimization tasks. In our simulations, we estimate the probability of finding a tour with length $L\leqslant 1.05L_{\rm op}$ and $L\leqslant 1.1L_{\rm op}$ among the obtained feasible tours. In fact, only the optimal tour and the second optimal tour are included in the region $L\leqslant 1.05L_{\rm op}$ for the present city distribution.

These four measures are estimated with 1000 random initial conditions for each $p(1)$. For every initial condition, evolution of the network is terminated when it has stayed at a same state successively for 20 time steps. The results are plotted in Fig. 11(a). In the Hopfield network $[p(1)=0]$, only about 60% of the resulted minima are feasible tours, and among these feasible tours, more than 95% of them are longer than $1.1L_{\rm op}$ and only 2 of the 1000 random initial conditions lead to the optimal tour. Once the chaotic annealing is employed, the values of all these four measures are improved at once. Specifically, for $p(1)=0.2$, the system escapes from almost all those minima, which are not feasible tours; more than 70% of the obtained tours are within $1.1L_{\rm op}$ and more than 25% of the obtained tours are the optimal one or the second optimal one. When $p(1)$ is large enough (about $p(1)>5.5$) most of the initial conditions ($\approx$96%) will finally lead to the optimal tour or the second optimal tour. The

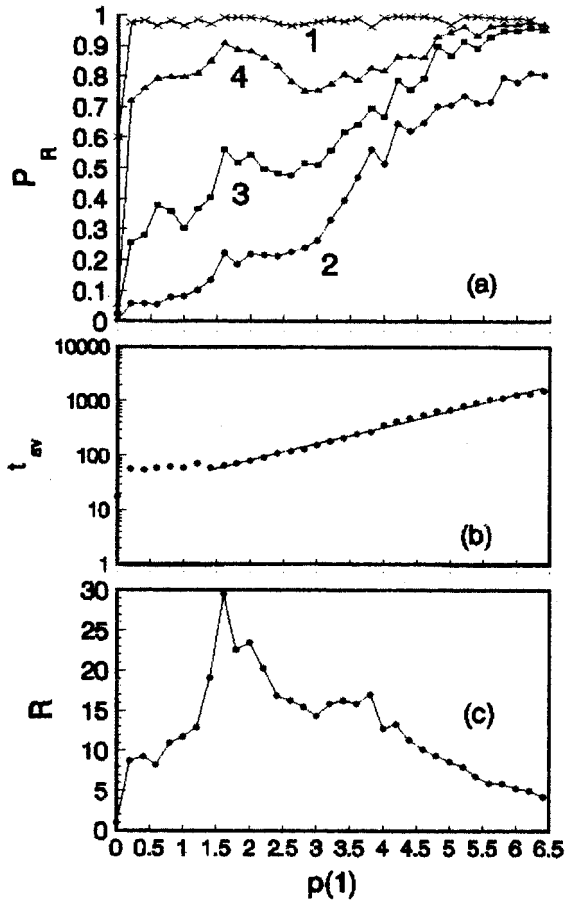

FIG. 11. Performance of chaotic annealing with respect to $p(1)$. (a) Plot 1 (cross) is the probability of finding a feasible tour; plot 2 (dot) is the probability of finding the optimal tour $L_{\rm op}$ among the obtained feasible tours; plot 3 (square) and plot 4 (triangle) are the probability of finding a tour among the obtained tours, with length $L \leqslant 1.05L_{\rm op}$ and $L \leqslant 1.1L_{\rm op}$, respectively. (b) Average time steps for the network to converge. Again it follows approximately an exponential law $t_{\rm av}=20.8\exp[0.70p(1)]$ for $p(1)$ in the wandering region [about $p(1)>1$]. (c) R.

probability of converging to the optimal tour in a simulation is improved up to about 0.8.

Similar to the one-dimensional case, the convergence rate of the system is mainly determined by $p(1)$. The average time needed for the system to converge is approximately $t_{\rm av}=20.8\exp[0.70p(1)]$ for $p(1)$ in the wandering region, as shown in Fig. 11(b). The time is longer than that of the Hopfield network. However, if considering the operation of random reinitializing the system once it reaches a stable state, the number for obtaining the optimal minimum during a sufficiently long period of time t_s is

$$N(t_s)=\frac{t_s}{t_{\rm av}}P_{\rm op}, \tag{19}$$

where $P_{\rm op}$ is the probability of obtaining the optimal tour in a simulation. When measured by $N(t_s)$, chaotic annealing is also much better than the Hopfield network, as illustrated by the ratio

$$R=\frac{N(t_s)\ \text{of chaotic annealing}}{N(t_s)\ \text{of Hopfield network}} \tag{20}$$

shown in Fig. 11(c).

We close this section with the conclusion that chaotic annealing works much more efficiently than the Hopfield model for a global minima search.

V. DISCUSSION

We have studied the role of chaotic transient in global optimization tasks. For a given energy function, we provide a general method for constructing a chaotic system based on the corresponding gradient descent system. The constructed system maintains some trend of quick descent to local minima, and at the same time has some chance of escaping from them. This property is utilized to search the local minima quickly. Chaos, which is generated temporally for searching for the minima in the state space, gradually vanishes when a bifurcation parameter is decreased gradually. It is shown that chaotic transient dynamics can serve as a more efficient global minima search than descent dynamics. The model is much simpler than that in Ref. [7] where it seems harder to analyze the role of chaos because an energylike function is not defined there [8].

Exploring the application of chaotic dynamics, including chaos control and synchronization has drawn much research attention recently. Unlike controlling chaos to a desired unstable periodic orbit by small modification of a system parameter, we control the system to a fixed state, which is the global minimum or its good approximations of the energy function of the system.

In a sense, we have developed a general chaotic annealing method for global optimization. The properties of the nonlinear self-feedback in this paper enable the method to be applied to a variety of energy minimum problems. When considering practical applications, our model has some advantages over the simple chaotic model in Ref. [8]. For a given energy function and the corresponding gradient descent system, many factors, such as the amplitude, the distribution, and the correlation of the external chaotic noise will affect the performance of network in Ref. [8], while the performance of our network is only governed by the initial value of parameter p, which is welcome for practical applications. Since our method is similar to simulated annealing in many ways, it should be meaningful in the future to compare it with simulated annealing as well as other conventional methods for optimization.

Adding some simple new ingredients such as the self-feedback in this paper to stable system seems to be a very direct way to construct systems with complex dynamics, which should prove useful when considering the applications of the complex dynamics in technology.

ACKNOWLEDGMENT

This project was supported by National Basic Research Project ''Nonlinear Science'' and the National Nature Science Foundation of China.

[1] E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990).
[2] L. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).
[3] W. J. Freeman, Biol. Cybern. **56**, 139 (1987).
[4] K. Aihara, T. Takabe, and M. Toyoda, Phys. Lett. A **144**, 333 (1990).
[5] I. Tsuda, Neural Networks **5**, 313 (1992).
[6] S. Nara, P. Davis, and H. Totsuji, Neural Networks **6**, 963 (1993).
[7] M. Inoue and A. Nagayoshi, Phys. Lett. A **158**, 373 (1991).
[8] Y. Hayakawa, A. Marumoto, and Y. Sawada, Phys. Rev. E **51**, R2693 (1995).
[9] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[10] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).
[11] X. Xu and W. T. Tsai, Neural Networks **4**, 193 (1991).
[12] J. J. Hopfield and W. D. Tank, Biol. Cybern. **52**, 141 (1985).
[13] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).