

**Б. В. КРЫЖАНОВСКИЙ**  
НИИ системных исследований РАН,  
Москва  
E-mail: kryzhanov@mail.ru

**О ТОПОЛОГИИ ПОТЕНЦИАЛЬНОЙ ПОВЕРХНОСТИ И  
ВЕРОЯТНОСТЯХ ОБНАРУЖЕНИЯ ЛОКАЛЬНЫХ МИНИМУМОВ  
В ЗАДАЧАХ БИНАРНОЙ ОПТИМИЗАЦИИ**

**Аннотация**

Обсуждается проблема бинарной оптимизации. На основе анализа обобщенной модели Хопфилда получены выражения, устанавливающие связь между глубиной локального минимума и шириной области притяжения, которую можно выразить в виде «глубже минимум – шире область притяжения – больше вероятность обнаружения в процессе случайного поиска». На основании этого, вероятность нахождения локального минимума при случайной инициализации нейронной сети удалось представить как функцию глубины этого минимума. В практических оптимизационных приложениях наличие таких выражений позволит по ряду уже найденных минимумов оценить вероятность нахождения более глубокого минимума и принять решение на остановку программы поиска или ее продолжение. Кроме того, описан итерационный алгоритм, позволяющий представить любую наперед заданную симметричную матрицу в виде взвешенного хеббовского разложения по внешним произведениям конфигурационных векторов. Возможность такого представления говорит о том, что все аналитические выводы, касающиеся ассоциативных нейросетей и оптимизационных алгоритмов, базирующихся на хеббовских матрицах, в равной степени применимы и для любого иного типа матриц. Развитая теория находится в хорошем согласии с результатами эксперимента.

**B. V. KRYZHANOVSKY**  
Institute of Systems Analysis, RAS,  
Moscow

**E-mail: kryzhanov@mail.ru**

**ABOUT POTENTIAL SURFACE TOPOLOGY AND RECOGNITION  
PROBABILITIES FOR LOCAL MINIMA IN BINARY OPTIMIZATION  
PROBLEMS**

**Abstract**

A binary optimization problem is discussed in the Lecture. Some expressions are derived basing on the Hopfield model to state an important relationship between a depth of local minimum and a width of attraction region. This relationship can be formulated as “deeper minimum – wider attraction region – greater detection probability during a random search process”. It has turned out to express the detection probability for local minimum as a function of the minimum depth for random initialization of a neural network. Such kind of expressions allow to evaluate the detection probability for deeper minimum in real-world applied optimization problems using a sequence of previous obtained minima in the searching process as well as to decide about shutdown or continuation of the search. An appropriate iteration algorithm is also suggested which makes it possible to represent any prescribed symmetric matrix as a weighted Hebbian expansion by external products of configuration vectors. An existence of this representation indicates that all the analytical conclusions related to associative neural networks and optimization algorithms based on the Hebbian matrices are equally applicable to any kind of matrices. Theoretical results obtained in the Lecture match well with experimental data.

**Введение**

Обычно нейронная система ассоциативной памяти рассматривается как система, решающая задачи распознавания или восстановления образов. Однако, ее можно рассматривать и как систему, решающую оптимизационную задачу – нейронная сеть в процессе релаксации находит конфигурацию, соответствующую минимуму энергии [1]. Это свойство нейронной сети можно использовать для решения различных *NP*-полных задач. Стандартный

подход сводится к нахождению такой архитектуры и параметров нейронной сети, при которых ценовая (целевая) функция совпадает с понятием нейросетевой энергии. Успешное применение нейронной сети к задаче коммивояжера [2], инициировало широкие исследования нейросетевых подходов к решению задач теории графов [3], нейросетевой оптимизации обработки изображений [4] и большому ряду других приложений. Эта область теории нейронных сетей активно развивается по настоящее время (см. работы [5–15] и ссылки в них).

Исследования в указанной области объединяет одно — сердцевиной подавляющего большинства нейросетевых оптимизационных алгоритмов является модель Хопфилда [1], а сам процесс оптимизации так или иначе сводится к нахождению в  $N$ -мерном конфигурационном пространстве глобального минимума некоего квадратичного функционала (энергии), построенного на заданной  $N \times N$ - матрице. Стандартный нейросетевой подход к решению указанной задачи сводится к процедуре случайного поиска оптимального решения: на первом этапе этой процедуры нейронная сеть случайным образом инициализируется, на втором этапе — нейронная сеть релаксирует в одно из возможных стабильных состояний, т. е. оптимизирует величину энергии. Поскольку искомым результат неизвестен и поиск проводится «вслепую», то нейросеть инициализируется раз за разом, с тем, чтобы найти как можно более глубокий минимум энергии. Несмотря на относительно успешное применение [2] этого подхода, всегда остается открытым вопрос — сколько таких случайных попыток следует сделать и когда проведенный поиск можно считать удачным (исходя из соотношения «затраченное время/полученный результат»). Сказанное справедливо как при спиновой [1–13], так и доменной [14,15] динамиках релаксации.

В работах [16–18] исследовались свойства энергетической поверхности и получены выражения, устанавливающие связь между глубиной локального минимума энергии и шириной области притяжения. На основании этого, вероятность нахождения локального минимума при случайной инициализации нейронной сети удалось представить как функцию глубины этого минимума. Полученные выражения можно сформулировать следующим образом: чем глубже локальный минимум, тем больше радиус его области притяжения и, следовательно, тем больше вероятность его обнаружения в процессе случайного поиска. Это означает, что наиболее глубокий минимум находится с наибольшей вероятностью. В практических приложениях наличие таких выражений позволит по ряду уже найденных минимумов оценить вероятность нахождения более глубокого минимума

и принять решение на остановку программы поиска или ее продолжение. Указанные выше результаты получены на основе анализа обобщенной модели Хопфилда — нейронной сети с хеббовской [26] (корреляционной) матрицей межсвязей. Для такого типа матриц получено прекрасное совпадение теории с экспериментом.

К сожалению, глубокий и последовательный анализ как свойств нейросети, так и возможности применения нейросети для оптимизационных целей, возможен только в случае, когда матрица связей нейронной сети построена по правилу Хебба или, хотя бы по предложенному в [16] обобщенному правилу Хебба. В связи с этим встает вопрос о применимости развитой теории нейросетей, в том числе и обсуждаемых здесь результатов работ [16–18], к случаю, когда матрица синаптических связей задана в произвольном (нехеббовском) виде. Для ответа на этот вопрос в настоящей работе описана нейросетевая итерационная процедура [19–21], позволяющая с заданной точностью представить любую наперед заданную симметричную  $N \times N$ -матрицу в виде взвешенного хеббовского разложения по конфигурационным векторам. Тем самым доказывается, что все аналитические выводы, касающиеся нейросетей с обобщенными хеббовскими матрицами и оптимизационных алгоритмов, в равной степени применимы и для любого иного типа матриц. Множественные эксперименты со случайными матрицами, подтверждают развитую в [16–18] теорию.

### Обобщенная модель Хопфилда

Рассмотрим модель Хопфилда. В теории нейронных сетей ее принято описывать как одномерную систему из  $N$  спинов-нейронов, которые могут ориентироваться вдоль, либо против заданной оси. Состояние такой нейронной сети характеризуется конфигурационным вектором  $\mathbf{S} = (s_1, s_2, \dots, s_N)$ , где  $s_i = \pm 1$ ,  $i \in (1, N)$ . Здесь мы будем рассматривать обобщенную модель, архитектура которой задается матрицей межсвязей

$$T_{ij} = \sum_{m=1}^M r_m s_{mi} s_{mj} \quad (1)$$

с нулевой диагональю ( $T_{ij} = 0$ ), организованной по правилу Хебба на  $M$  образах —  $N$ -мерных бинарных векторах  $\mathbf{S}_m = (s_{m1}, s_{m2}, \dots, s_{mN})$ . Обобщение состоит в том, что каждый образ  $\mathbf{S}_m$  добавляется в матрицу

$T_{ij}$  со своим статвесом  $r_m$ . Для упрощения выражений принята нормировка статвесов  $\sum r_m^2 = 1$  не нарушающая общности рассмотрения. Это незначительное видоизменение модели оказывается весьма существенным, поскольку, в отличие от стандартной модели, позволяет описывать нейронную сеть с невырожденным спектром минимумов.

Энергия нейросети описывается выражением

$$E = - \sum_{i,j=1}^N s_i T_{ij} s_j, \quad (2)$$

а ее динамика заключается в следующем. Задается начальное состояние сети  $\mathbf{S}$  (начальные направления спинов устанавливаются в соответствии со знаками компонент вектора  $\mathbf{S}$ ). Затем, вычисляется локальное поле  $h_i = -\partial E / \partial s_i$ , действующее на произвольно выбранный  $i$ -й спин со стороны всех остальных спинов системы в момент времени  $t$  и определяется энергия спина в этом поле  $\varepsilon_i = -s_i h_i$ . Если направление спина совпадает с направлением локального поля ( $\varepsilon_i < 0$ ), то его положение энергетически устойчиво и в последующий момент времени состояние спина остается неизменным. В противном случае ( $\varepsilon_i > 0$ ) положение спина неустойчиво и он разворачивается вдоль направления этого поля, переходя в состояние  $s_i(t+1) = -s_i(t)$  с энергией  $\varepsilon_i(t+1) < 0$ . Такая процедура последовательно применяется ко всем спином нейронной сети. При каждом перевороте спина энергия сети понижается. Это означает, что сеть за конечное число шагов перейдет в стабильное состояние, соответствующее локальному минимуму энергии.

### Размер области притяжения

Исследуем при каких условиях присутствующий в матрице (1) образ  $\mathbf{S}_m$  будет являться неподвижной точкой [1], в которой энергия сети  $E$  достигает своего (локального) минимума  $E_m$ . Для получения корректных оценок рассмотрение проведем в асимптотическом пределе  $N \rightarrow \infty$ . Определим область притяжения образа  $\mathbf{S}_m$  как совокупность точек  $N$ -мерного пространства, из которых нейронная сеть релаксирует в конфигурацию  $\mathbf{S}_m$  и попробуем оценить размер этой области. Пусть начальное состояние сети  $\mathbf{S}$  находится в некоторой окрестности образа  $\mathbf{S}_m$ . Тогда вероятность того,

что сеть сойдется к точке  $S_m$ , опишется выражением:

$$\text{Pr} = 2^{-N}(1 + \text{erf}z)^N. \quad (3)$$

Здесь  $\text{erf}z$  — функция ошибок переменной  $z$ :

$$z = r_m \frac{\sqrt{N}}{\sqrt{2(1 - r_m^2)}} \left(1 - \frac{2n}{N}\right), \quad (4)$$

где  $n$  — хеммингово расстояние между  $S_m$  и  $S$ . Выражение (3) можно получить методами теоретико-вероятностного подхода, повторив хорошо известные для случая  $r_m = 1/\sqrt{M}$ ,  $\forall m \in (1, M)$  выкладки [22].

Из (3) следует, что область притяжения определяется как совокупность близких к  $S_m$  точек конфигурационного пространства, для которых справедливо соотношение  $n \leq n_m$ :

$$n_m = \frac{N}{2} \left(1 - \frac{r_0}{r_m} \sqrt{1 - r_m^2}\right), \quad (5)$$

где

$$r_0 = \sqrt{\frac{2 \ln N}{N}}. \quad (6)$$

Действительно, при  $n \leq n_m$  вероятность схождения в точку  $S_m$  с ростом  $N$  асимптотически стремится к единице; в противном случае ( $n > n_m$ ) имеем  $\text{Pr} \rightarrow 0$ . Это означает, что величину можно рассматривать как радиус области притяжения локального минимума  $E_m$ .

Из (5) следует, что при  $r_m \rightarrow r_0$  радиус области притяжения стремится к нулю (рис. 1). Это означает, что образы, прописанные в матрицу межсвязи (1) с статвесом, меньшим  $r_0$ , попросту не образуют локальных минимумов. Локальные минимумы имеются только в точках  $S_m$ , образы которых прописаны в матрице межсвязей с относительно большими статвесами  $r_m > r_0$ . Более того, если статвес одного из образов, например образа  $S_1$ , сделать значительно больше чем у остальных ( $r_1 \rightarrow 1$ ), то область его притяжения ( $n_1 \sim N/2$ ) охватит половину всего  $N$ -мерного пространства (остальное полупространство занимает область притяжения его негатива  $S_1^* = -S_1$ ). При этом, сеть будет иметь всего лишь два локальных минимума в точках  $S_1$  и  $S_1^*$ , остальные локальные минимумы исчезнут. Этот предельный случай не представляет практического интереса для оптимизационных задач и ниже мы его рассматривать не будем.

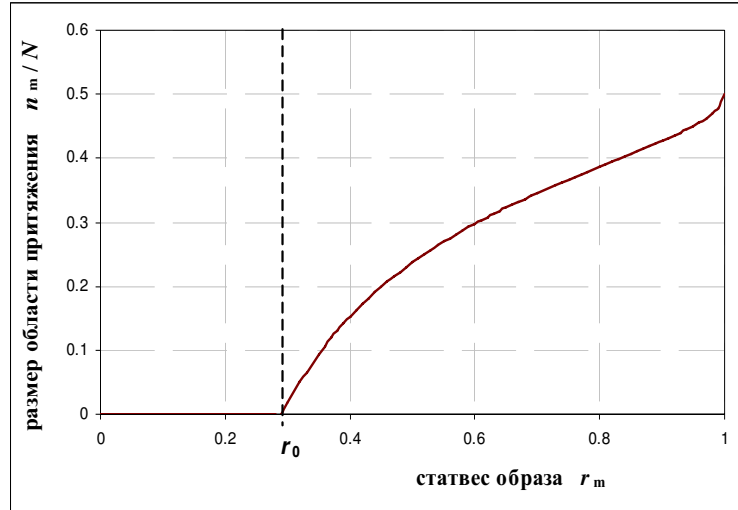


Рис. 1. Характерный вид зависимости ширины области притяжения от статвеса образа. Локальный минимум существует только для образов, статвес которых в матрице межсвязей больше  $r_0$ . При  $r_m \rightarrow r_0$  размер области притяжения стремится к нулю, т. е. образы со статвесом  $r_m < r_0$  не создают локальных минимумов.

### Объем памяти обобщенной модели

Из проведенного выше анализа следует, что не все из прописанных в матрице межсвязей образов образуют локальные минимумы и могут распознаваться нейронной сетью. В связи с этим следует дать четкое определение понятия объема нейросетевой памяти  $\bar{M}$ . Определим его (объем памяти) как число тех образов, которые образуют локальные минимумы и, следовательно, могут быть восстановлены нейронной сетью без ошибок (абсолютная память согласно определению, введенному в [23,25]).

Очевидно, что объем памяти зависит от того, как сформирована матрица межсвязей и, в частности, от распределения статвесов. Рассмотрим, например, случай, когда статвеса образов случайны и равномерно распределены на отрезке  $r_m^2 \in [0, 2/M]$  со средним  $\bar{r}_m^2 = 1/M$ . В этом случае, вероятность того, что статвес образа удовлетворяет условию  $r_m > r_0$ , равна

величине  $1 - Mr_0^2/2$ . Соответственно, объем памяти модели определится выражением:

$$\bar{M} = M \left( 1 - \frac{Mr_0^2}{2} \right). \quad (7a)$$

Зависимость объема памяти от числа прописанных в матрице межсвязей образов  $M$ , представлена на рис. 2. Как видим, максимальный объем памяти  $\bar{M} = N/4 \ln N$  достигается при числе образов  $M = N/2 \ln N$ . При  $M \geq N/\ln N$  объем памяти сети обращается в ноль — нет ни одного образа, образующего локальный минимум.

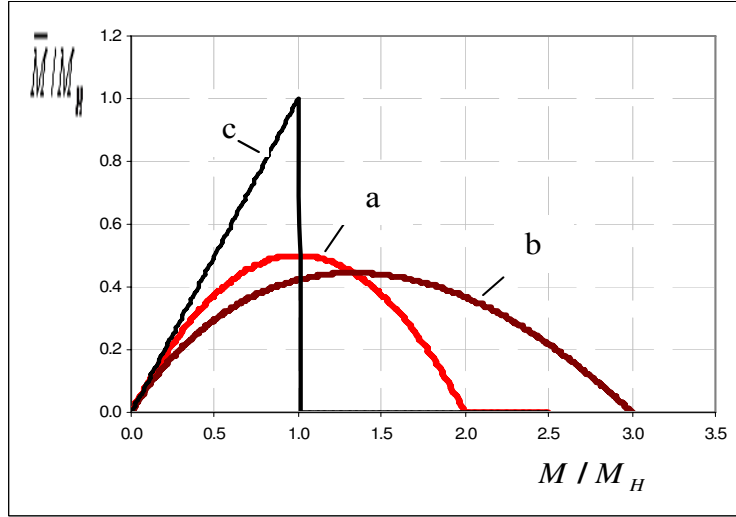


Рис. 2. Зависимость числа устойчивых точек от числа образов  $M$  при разных распределениях статвесов: **a** — распределение статвесов (7a); **b** — распределение (7b); **c** — стандартная модель Хопфилда. Все величины нормированы на величину объема памяти  $M_H = N/2 \ln N$  стандартной модели Хопфилда.

В случае, когда статвеса образов случайны и равномерно распределены на отрезке  $r_m \in [0, \sqrt{3/M}]$  со средним  $\bar{r}_m^2 = 1/M$ , зависимость объема памяти от числа образов  $M$  определяется следующим выражением:

$$\bar{M} = M \left( 1 - r_0 \sqrt{\frac{M}{3}} \right). \quad (7b)$$



Максимальный объем памяти  $\overline{M} = 2N/9 \ln N$  в этом случае достигается при числе образов  $M = 2N/3 \ln N$ , а при  $M \geq 3N/2 \ln N$  объем памяти сети обращается в ноль.

В другом предельном случае, когда все статвесы равны друг другу ( $r_m = 1/\sqrt{M}$ ), условие существования локальных минимумов  $r_m > r_0$  преобразуется в известное [5] ограничение  $M < N/2 \ln N$  на емкость памяти стандартной модели Хопфилда. В этом случае понятие «емкость памяти» совпадает с числом образов  $M$ , поскольку либо распознаются все прописанные в матрице межсвязей образы, либо не распознается ни один из них. Соответственно, максимальный объем памяти определяется выражением  $\overline{M} = N/2 \ln N$ .

В заключение этого раздела отметим, что удаление из матрицы межсвязей образов, прописанных в матрице межсвязей с статвесами, меньшими критического значения ( $r_m < r_0$ ), не влияет на размер объема памяти, но улучшает ее распознающие характеристики — размер областей притяжения оставшихся образов увеличится и, следовательно, сеть сможет распознавать более сильно искаженные образы.

### Энергия локального минимума

Из анализа выражения (2) следует, что энергию локального минимума  $E_m$  с точностью до незначительной флуктуации порядка

$$\sigma_m = N \sqrt{1 - r_m^2} \quad (8)$$

можно представить в виде

$$E_m = -r_m N^2. \quad (9)$$

Тогда, с учетом соотношения (5) из (9) нетрудно получить выражение

$$E_m = E_0 \frac{1}{\sqrt{\left(1 - \frac{2n_m}{N}\right)^2 + \frac{E_0^2}{N^4}}}, \quad (10)$$

где

$$E_0 = -N \sqrt{2N \ln N}, \quad (11)$$

устанавливающее связь между глубиной локального минимума и шириной его области притяжения. Как видим, чем шире область притяжения, тем

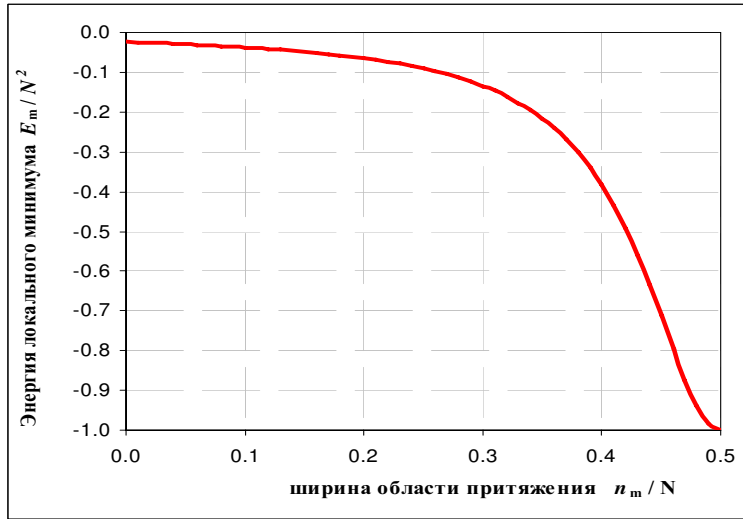


Рис. 3. Зависимость энергии локального минимума от ширины области притяжения.

глубже минимум, и наоборот — чем глубже минимум, тем шире область его притяжения (см. рис. 3).

Введенная здесь величина  $E_0$  характеризует сразу два параметра нейронной сети. Во-первых, ею определяется полуширина лоренцевского контура распределения (10). Во-вторых, из (10) вытекает неравенство  $E_m \leq E_0$ , т. е.  $E_0$  является верхней границей спектра локальных минимумов. Эти результаты хорошо согласуются с результатами компьютерных экспериментов, в ходе которых проверялось, существует ли в точке  $S_m$  локальный минимум или его нет. Результаты одного из экспериментов ( $N = 500$ ,  $M = 25$ ) приведены на рис. 4. Как видим, хорошо прослеживается линейная зависимость энергии локального минимума от величины статвеса образа. Причем, экспериментальные точки, соответствующие локальным минимумам, расположены только в правом нижнем квадранте, где  $r_m > r_0$  и  $E_m < E_0$ . Кроме того, на рис. 4 видно, что в соответствии с (8) флуктуации энергий минимумов с ростом статвеса уменьшаются.

В заключение этого раздела приведем вытекающие из предыдущего анализа полезные соотношения: глубина локального минимума ограниче-

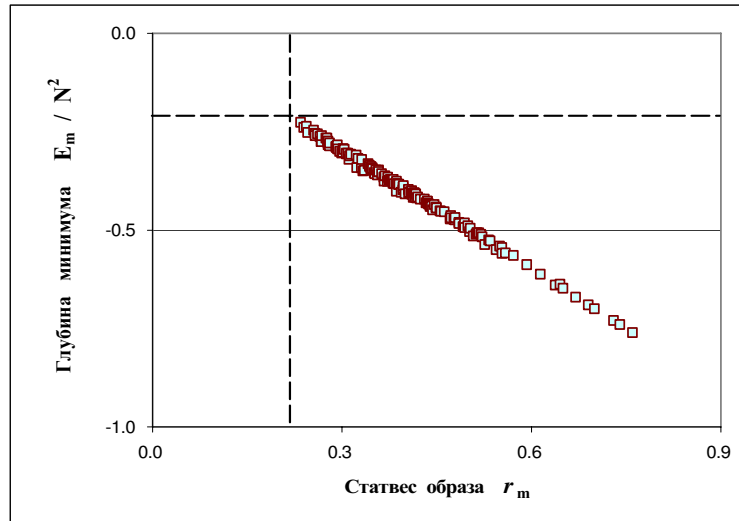


Рис. 4. Зависимость энергии  $E_m$  локального минимума от статвеса образа  $r_m$ .

на соотношением  $-N^2 \leq E_m \leq E_0$ ; сумма квадратов энергий локальных минимумов удовлетворяет соотношению  $\sum E_m^2 \leq N^4$ , в котором знак равенства стоит в том случае, когда всем образам соответствуют локальные минимумы.

### Вероятность обнаружения минимума

Определим теперь вероятность  $W$  нахождения локального минимума  $E_m$  при случайном поиске. Искомая вероятность по определению совпадает с вероятностью того, что задавая начальную конфигурацию мы попадаем в область притяжения образа  $S_m$ . Следовательно величина  $W = W(n_m)$  есть число точек в сфере радиуса  $n_m$ , приведенное к общему числу точек в  $N$ -мерном пространстве:

$$W = 2^{-N} \sum_{n=1}^{n_m} C_N^n. \quad (12)$$

Выражениями (8) и (12) в неявном виде задается связь между глубиной локального минимума и вероятностью его нахождения. Применяя к биномиальным коэффициентам асимптотическое разложение Стирлинга и заменяя в (12) суммирование интегрированием, можно представить искомую связь как

$$W = W_0 e^{-Nh}, \quad (13)$$

где  $h$  — функция Шеннона

$$h = \frac{n_m}{N} \ln \frac{n_m}{N} + \left(1 - \frac{n_m}{N}\right) \ln \left(1 - \frac{n_m}{N}\right) + \ln 2. \quad (14)$$

Здесь  $W_0$  — не существенная для дальнейшего анализа медленная функция от  $E_m$ , которую можно получить асимптотической оценкой (12) при условии  $n_m \gg 1$ , а зависимость  $W = W(n_m)$  всецело определяется быстрой экспонентой.

Как следует из (12), вероятность обнаружения локального минимума малой глубины ( $E_m \sim E_0$ ) убывающе мала как  $W \sim 2^{-N}$ . Заметно отличной от нуля вероятность  $W$  становится только в случае достаточно глубоких минимумов  $|E_m| \gg |E_0|$ , размеры областей притяжения которых сопоставимы с величиной  $N/2$ . Выражение (13) с учетом (10) в этом случае преобразуется к зависимости  $W = W(E_m)$ , задаваемой выражением

$$W = W_0 \exp \left[ -N E_0^2 \left( \frac{1}{E_m^2} - \frac{1}{E_\infty^2} \right) \right], \quad (15)$$

где

$$E_\infty^2 = \sum_m E_m^2. \quad (16)$$

Здесь мы ввели новую константу  $E_\infty$ , хотя в принятой нами нормировке из (9) вытекает  $E_\infty = -N^2$  и введение нового обозначения не имеет особого смысла. Однако, для других типов нормировок вполне возможна иная зависимость  $E_\infty$  от  $N$ , что может привести к недоразумениям.

Как видим, вероятность обнаружения минимума растет с ростом его глубины. Эта зависимость подтверждается результатами экспериментов, проведенных для матриц (1) с малым загрузочным параметром  $M/N \leq 0/05$ , при котором проведенный выше анализ справедлив. Сплошная кривая на рис. 5 построена по формуле (13), точки — эксперимент. Как видим, хорошее согласие достигается прежде всего для наиболее глубоких локальных минимумов, которые соответствуют записанным в матрице межсвязей

образам  $S_m$  (на рисунке — это область энергий  $E_m < -0.48N^2$ ). Экспериментально обнаруженные минимумы малой глубины (точки в области  $E_m > -0.44N^2$ ) — это так называемые «химеры» [1]. В стандартной модели Хопфилда ( $r_m \equiv 1/\sqrt{M}$ ) они появляются при относительно больших нагрузках. В рассматриваемом здесь более общем случае они могут появиться и раньше. Причины их появления хорошо исследованы методами статистической физики в работах [24–25], где показано, что химеры образуются как следствие интерференции минимумов  $S_m$ . При малой нагрузке химеры отделены от минимумов  $S_m$  энергетической щелью, отчетливо видной на рис. 5. С ростом параметра нагрузки  $M/N$  соответствующие химерам минимумы опускаются и, при нагрузке  $M/N \sim 0.1$ , сравниваются по глубине с локальными минимумами  $S_m$ . Более того, при  $M/N \geq 0.14$  появляются глубоко расположенные так называемые ложные (spurious) минимумы, образованные из больших фрагментов векторов  $S_m$ , а большинство локальных минимумов  $S_m$  исчезает. Эта экспериментально наблюдавшаяся картина имеет простое объяснение в рамках проведенного выше анализа: с ростом  $M$  уменьшается число образов, удовлетворяющих условию  $r_m > r_0$  существования локального минимума.

Несмотря на такую сложную картину энергетической поверхности описываемая выражениями (13)–(15) зависимость «глубже минимум — больше область притяжения — больше вероятность попадания в этот минимум» как правило сохраняется при любых параметрах нагрузки. Отметим, что зависимость (13) выполняется тем лучше, чем глубже описываемый этим выражением минимум. А поскольку для задачи оптимизации важно описать поведение именно наиболее глубоких минимумов, то это позволяет сформулировать эвристический подход для отыскания глобального минимума функционала (2) с произвольной наперед заданной матрицей (не обязательно хеббовской). Состоит он в том, чтобы использовать выражение (13) с неизвестными параметрами  $W_0$ ,  $E_0$  и  $E_\infty$ . Для этого запускается процедура случайного поиска и находится какое-то число минимумов. По полученным данным определяются характерные для данной матрицы значения  $E_0$  и  $E_\infty$  и величина подгоночного параметра  $W_0$ . Подстановка этих величин в (13) позволит оценить вероятность нахождения неизвестного более глубокого минимума  $E_m$  (если он существует) и принять решение на остановку программы поиска (если оценка пессимистична) или ее продолжение.

Такой подход опробован на хеббовских матрицах при достаточно больших нагрузочных параметрах ( $M/N = 0.2 \div 2$ ), когда нейросетевая система

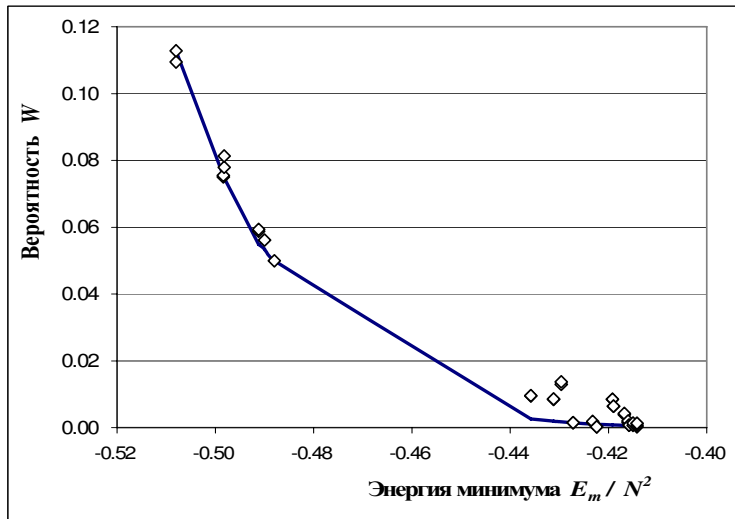


Рис. 5. Зависимость вероятности  $W$  обнаружения локального минимума от его глубины  $E_m$ .

переходит в состояние спинового стекла [23] и применение полученных выше результатов становится не совсем корректным. Результат одного из экспериментов приведен на рис. 6. В ходе эксперимента по обнаруженным минимумам (точки  $A$ ) были рассчитаны параметры  $W_0$ ,  $E_0$  и  $E_\infty$  и построена зависимость  $W = W(E_m)$  (сплошная кривая). Затем, в результате многократно повторяемой процедуры случайного поиска ( $\sim 10^5$  стартов нейронной сети) были установлены другие минимумы (точки  $B$ ) и точные значения вероятностей попадания в них. Как видим, несмотря на разброс, предсказанные значения по порядку величины хорошо совпадают с реальными значениями вероятностей.

### Форма локального минимума

Установленная выше зависимость «глубже минимум – больше область притяжения – больше вероятность попадания в этот минимум» на первый взгляд может показаться достаточно удивительной. Кажется, что она про-

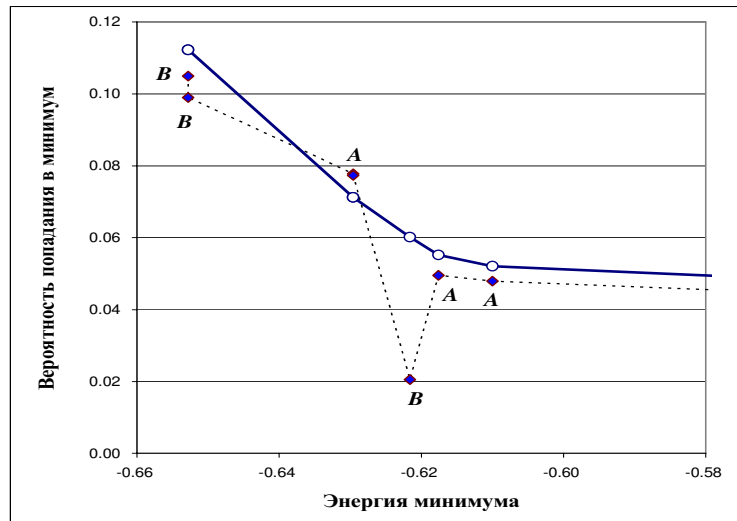


Рис. 6. Сравнение предсказанных вероятностей (сплошная линия) и экспериментально найденных значений (точки, связанные пунктиром).

тиворечит здравому смыслу, который подсказывает, что могут существовать очень глубокий, но очень узкий минимум, вероятность обнаружения которого очень мала, или наоборот — относительно неглубокий, но очень широкий минимум, вероятность обнаружения которого очень велика. Ниже мы покажем, что такого противоречия нет, поскольку все минимумы имеют строго заданную форму, в которой глубина и ширина (радиус области притяжения) взаимосвязаны так, что ширина минимума растет с ростом его глубины.

Чтобы доказать это утверждение, рассмотрим поведение энергии вблизи некоторой точки  $S_m$ , в которой энергия достигает своего локального минимума. Подставляя (1) в (2) для энергии этого локального минимума с точностью до несущественной константы получим:

$$E_m = -r_m N^2 - \sum_{\mu \neq m} r_\mu (S_m S_\mu)^2. \quad (17)$$

Введем понятие  $n$ -окрестности точки  $S_m$  — это совокупность точек  $N$ -

мерного пространства, отстоящих от  $S_m$  на расстоянии  $n$ , т. е. каждый вектор  $S_m^{(n)}$  отличается от  $S_m$  на  $n$  бит ( $n$  перевернутых спинов). Например, 1-окрестность содержит  $N$  точек, отличающихся от  $S_m$  в одном бите, 2-окрестность содержит  $N(N-1)/2$  точек,  $n$ -окрестность —  $C_N^n$  точек. Рассмотрим одну из точек  $n$ -окрестности  $S_m^{(n)}$ . Для простоты обозначений будем полагать, что векторы  $S_m^{(n)}$  и  $S_m$  отличаются в первых  $n$  битах (первые  $n$  спинов в этой точке  $S_m^{(n)}$  перевернуты относительно  $S_m$ ). Тогда энергию сети в этой точке из  $n$ -окрестности можно записать в виде:

$$E_m^n = -r_m N^2 \left(1 - \frac{2n}{N}\right)^2 - \sum_{\mu \neq m} r_\mu (S_m S_\mu)^2 + 4 \sum_{\mu \neq m} \sum_{i=1}^n \sum_{j=n+1}^N r_\mu s_{mi} s_{\mu i} s_{\mu j} s_{mj} \quad (18)$$

или в виде

$$E_m^{(n)} = E_m + \delta E_m(n) + R, \quad (19)$$

где

$$R = 4 \sum_{\mu \neq m} \sum_{i=1}^n \sum_{j=n+1}^N r_\mu s_{mi} s_{\mu i} s_{\mu j} s_{mj}, \quad (20)$$

$$\delta E_m(n) = 4r_m N^2 \frac{n}{N} \left(1 - \frac{n}{N}\right). \quad (21)$$

Величина  $R$  незначительно меняется от точки к точке в  $n$ -окрестности. Нетрудно заметить, что ее можно описать как нормально распределенную случайную величину с нулевым средним ( $\langle R \rangle = 0$ ) и среднеквадратичным отклонением

$$\sigma_m(n) = 2N \sqrt{\frac{r_m n}{N} \left(1 - \frac{n}{N}\right)}. \quad (22)$$

Соответственно, величина  $\delta E_m(n) = \langle E_m^n - E_m \rangle$  есть не что иное, как среднее значение разницы в энергии между точками в  $n$ -окрестности  $\{S_m^{(n)}\}$  и в минимуме  $S_m$ . Именно зависимостью величины  $\delta E_m(n)$  от  $n$  и задается форма локального минимума. Выражение (21) находится в прекрасном согласии с результатами численного эксперимента. В ходе эксперимента генерировалась случайная матрица, случайно находился какой либо локальный минимум и затем исследовалось среднее значение энергии вокруг этого минимума. Результаты одного из таких экспериментов приведены на рис. 7.



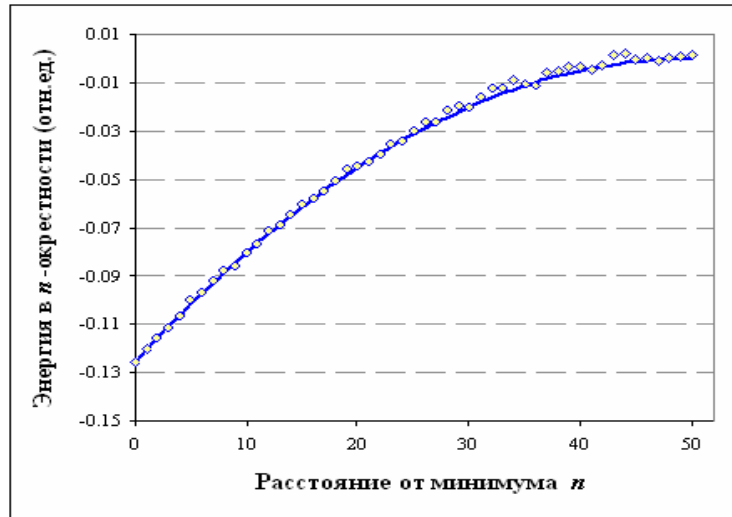


Рис. 7. Форма локального минимума: среднее значение энергии в  $n$ -окрестности локального минимума. Кривая соответствует формуле (21), маркер — экспериментальные точки.

### Итерационный алгоритм

Опишем теперь итерационную процедуру, позволяющую с заданной точностью представить любую заданную симметричную матрицу  $\hat{\mathbf{T}}$  в виде разложения (1) по неподвижным точкам модели Хопфилда.

Введем обозначение матрицы  $\hat{\mathbf{T}}_k$ :

$$\hat{\mathbf{T}}_k = \hat{\mathbf{T}} - \sum_{m=1}^k r_m S_m^+ S_m, \quad (23)$$

матричные элементы которой имеют вид:

$$T_{k,ij} = T_{ij} - \sum_{m=1}^k r_m s_{mi} s_{mj},$$

и зададим невязку на  $k$ -м шаге итерации в виде:

$$\varepsilon_k = \sum_{i=1}^N \sum_{j=1}^N T_{k,ij}^2, \quad (24)$$

Определим процедуру минимизации невязки таким образом, чтобы с ростом  $k$  имело место  $\varepsilon_k \rightarrow 0$ . Если при некотором числе итераций  $k \geq M$  величина невязки станет меньше заданного значения  $\varepsilon_M$ , то тем самым мы решим поставленную выше задачу: равенство (1) будет выполняться с заданной точностью.

Предлагаемая здесь итерационная процедура сводится к следующему. На  $k$ -м шаге итерации генерируем случайный конфигурационный вектор  $\mathbf{S}_k$ . С учетом (23) выражение для невязки (24) представим в виде:

$$\varepsilon_k = \sum_{i=1}^N \sum_{j=1}^N (T_{k-1,ij} - r_k s_{ki} s_{kj})^2, \quad (25)$$

или в виде:

$$\varepsilon_k = \varepsilon_{k-1} + 2r_k E_k + r_k^2 N^2, \quad (26)$$

где

$$E_k = - \sum_{i=1}^N \sum_{j=1}^N T_{k-1,ij} s_{ki} s_{kj}. \quad (27)$$

Из (26) следует, что минимум невязки  $\varepsilon_k$  будет достигнут, если величину статвеса задать в виде:

$$r_k = - \frac{E_k}{N^2}. \quad (28)$$

Тогда минимальное значение невязки примет вид:

$$\varepsilon_k = \varepsilon_{k-1} - r_k^2 N^2. \quad (29)$$

Как видно из (29), величину  $\varepsilon_k$  можно сделать еще меньше, если увеличить статвес  $\varepsilon_k$ . Для этого надо запустить стандартную спиновую динамику [2] оптимизации квадратичного функционала  $E_k$ , т. е. оптимизировать энергию модели Хопфилда (27), построенной на матрице  $\hat{\mathbf{T}}_{k-1}$ . При этом вектор  $\mathbf{S}_k$  заменяется на вектор  $\mathbf{S}_k^{onm}$ , соответствующей экстремуму функционала  $E_k$ . В точке  $\mathbf{S}_k^{onm}$ , в соответствии с (28), достигается максимальное (по модулю) значение статвеса  $r_k$  и минимальное значение невязки  $\varepsilon_k$ .

Далее, на основе вектора  $\mathbf{S}_k = \mathbf{S}_k^{onm}$  и отвечающего ему значения статвеса  $r_k$  в соответствии с (23) строим матрицу  $\hat{\mathbf{T}}_k = \hat{\mathbf{T}}_{k-1} - r_k \mathbf{S}_k^+ \mathbf{S}_k$  и переходим к следующему шагу итерации.

Покажем, что описанная выше итерационная процедура сходится за конечное число итераций. Для этого, с учетом полученного выше соотношения  $E_k = -r_k N^2$  между энергией экстремума и статвесом паттерна, преобразуем (29) к виду:

$$\varepsilon_{k+1} = \varepsilon_k (1 - r_k^2), \quad (30)$$

из которого несложно получить следующее соотношение

$$\varepsilon_{k+1} = \varepsilon_0 \prod_{m=1}^k (1 - r_m^2), \quad (31)$$

где  $\varepsilon_0$  — начальное значение невязки, равное сумме квадратов матричных элементов исходной матрицы  $\hat{\mathbf{T}}$ . Поскольку на статвес паттерна, соответствующего экстремуму функционала энергии, выше получено ограничение

$$r_k^2 \geq \frac{2 \ln N}{N}, \quad (32)$$

то для величины невязки из (30)–(31) вытекает соотношение

$$\varepsilon_k \sim \varepsilon_0 \exp\left(-\frac{2k \ln N}{N}\right), \quad (33)$$

из которого нетрудно оценить число итераций  $M$ , обеспечивающих заданную точность разложения (1). Нетрудно заметить, что число итераций конечно. Численные эксперименты подтверждают сказанное (см. рис. 8).

### Обсуждение результатов

Проведенный анализ показал, что свойства обобщенной модели описываются двумя критическими параметрами  $r_0$  и  $E_0$ . Первый из них определяет минимальное значение статвеса, при котором образ образует локальный минимум и может быть восстановлен нейронной сетью. Второй — это верхняя граница локальных минимумов энергетической поверхности. Существенно, что величины обоих параметров никак не зависят от числа образов  $M$ , на которых построена матрица межсвязей. Существенно

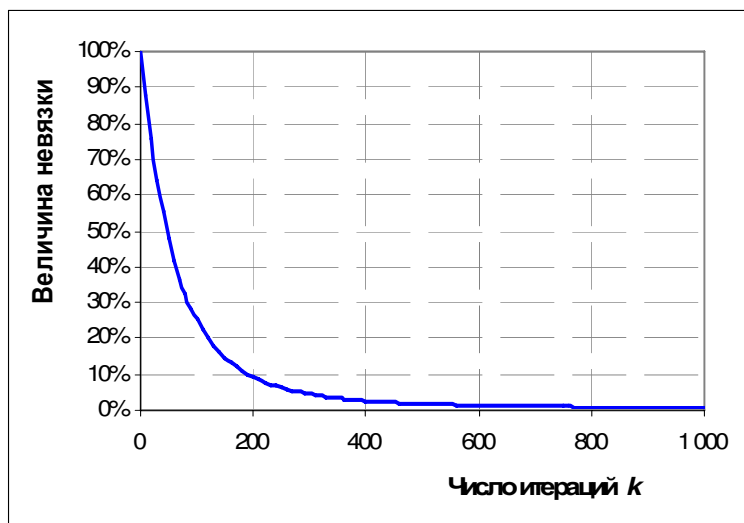


Рис. 8. На рисунке показано как невязка  $\varepsilon_k$  уменьшается с ростом числа итераций  $k$ . Кривая взята из эксперимента, проведенного при условиях: матричные элементы исходной матрицы  $\mathbf{T}$  генерировались случайным образом с равномерным распределением на отрезке  $[-1, +1]$  и нулевым средним; размерность конфигурационного пространства —  $N = 1000$ .

и то, что удаление из матрицы межсвязей образов, прописанных в матрице межсвязей с статвесами, меньшими критического значения ( $r_m < r_0$ ), хотя и не влияет на размер объема памяти, однако улучшает ее распознающие характеристики - размер областей притяжения оставшихся образов увеличится и, следовательно, сеть может распознавать более сильно искаженные образы. Вместе с этим, обратим внимание на тот факт, что с ростом размерности сети радиус области притяжения локального минимума уменьшается. Как следует из (5), такое уменьшение имеет место даже в том случае, когда загрузка сети с ростом  $N$  логарифмически уменьшается как  $M/N = \text{const}/2 \ln N$ . Это означает, что с ростом  $N$  уменьшается уровень допустимых искажений, при которых сеть способна восстанавливать образы (сеть меньших размеров способна распознавать более сильно искаженные образы).

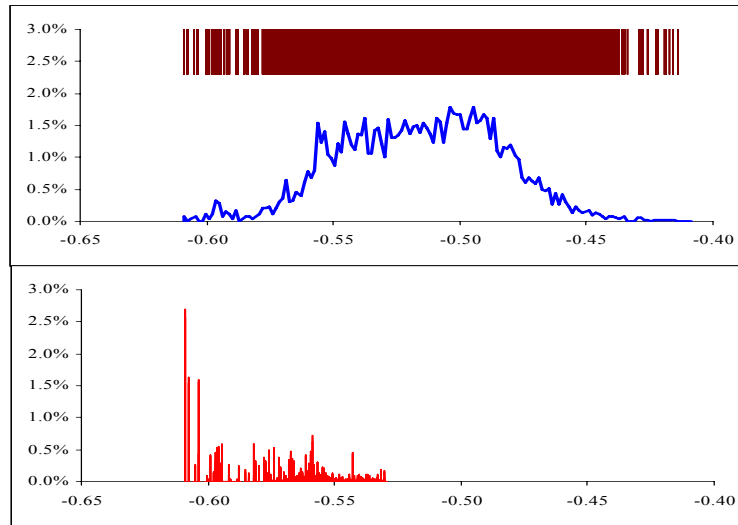


Рис. 9. Случай хеббовской матрицы с большой загрузкой ( $M/N = 1$ ).  
**(a)** Кривая показывает спектральную плотность минимумов (число минимумов на единичном интервале). В верхней части рисунка представлен сам спектр минимумов — каждая вертикальная черта соответствует минимуму. По оси ординат отложена спектральная плотность в %, по оси абсцисс — энергия в относительных единицах  $E/N^2$ .  
**(b)** Зависимость вероятности нахождения минимума от его глубины. По оси ординат отложена вероятность в %, по оси абсцисс — энергия минимума.

Во избежание недоразумений следует указать на то, что связь между радиусом области притяжения  $n_m$  и объемом этой области  $W = W(n_m)$  существенно нелинейная (12). Это означает, что даже при очень большом значении радиуса (например  $n_m \sim 0.49N$ ) область притяжения может занимать очень малую часть  $N$ -мерного пространства, а доля пространства вне области притяжения (т. е. все точки, отстоящие от минимума на расстоянии от  $n_m$  до  $0.5N$ ) с ростом  $N$  нарастает экспоненциально и при достаточно больших значениях  $N$  стремится к  $1/2$ .

В заключение сделаем несколько практических замечаний, относящихся к сравнению теории с экспериментом. Во-первых, напомним, что карти-

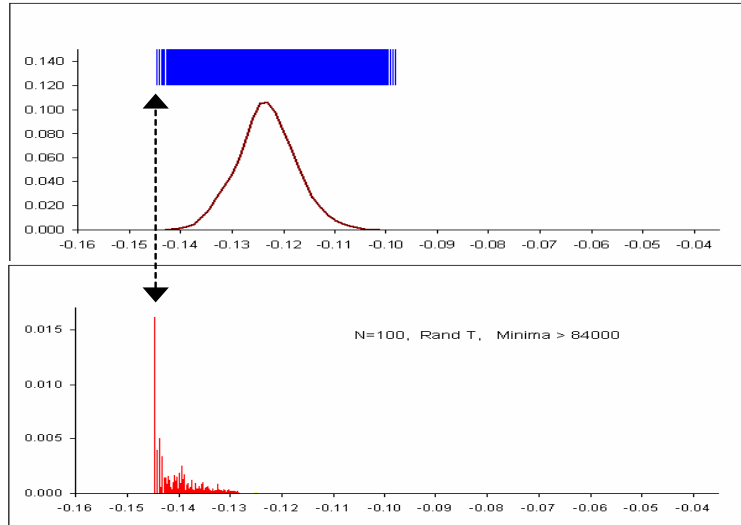


Рис. 10. Случайная матрица с равномерным распределением матричных элементов. **(а)** Кривая показывает спектральную плотность минимумов (число минимумов на единичном интервале). В верхней части рисунка представлен сам спектр минимумов — каждая вертикальная черта соответствует минимуму. По оси ординат отложена спектральная плотность в %, по оси абсцисс — энергия в относительных единицах. **(б)** Зависимость вероятности нахождения минимума от его глубины. По оси ординат отложена вероятность в %, по оси абсцисс — энергия минимума  $E/N^2$ .

на энергетического ландшафта симметрична относительно изменения знака вектора состояния нейронной сети и, следовательно, каждому локальному минимуму в точке  $S$  соответствует точно такой же минимум в точке  $S^* = -S$ . Поэтому, экспериментально измеренную вероятность нахождения минимума с энергией  $E$  следует сравнивать с величиной  $W = W(E)$ , если только в ходе эксперимента фиксировалась координата минимума, в противном случае — с величиной  $2W(E)$ .

Итак, получены выражения, устанавливающие связь между глубиной локального минимума и шириной области притяжения, которую можно выразить в виде «глубже минимум — шире область притяжения — больше»

вероятность обнаружения в процессе случайного поиска». Тем самым показано, что в процессе случайного поиска с наибольшей вероятностью находится наиболее глубокий минимум. Эти выражения получены на основе анализа обобщенной модели Хопфилда. Однако, показано, что существует итерационный алгоритм, позволяющий за конечное число шагов представить любую наперед заданную симметричную матрицу в виде взвешенного хеббовского разложения (1) по внешним произведениям конфигурационных векторов. Возможность такого представления говорит о том, что все аналитические выводы, касающиеся ассоциативных нейросетей и оптимизационных алгоритмов, базирующихся на хеббовских матрицах, в равной степени применимы и для любого иного типа матриц. Развитая теория находится в хорошем согласии с результатами эксперимента. Действительно, многочисленные минимизационные эксперименты, проведенные на хеббовских матрицах с произвольной загрузкой (как очень малой, так и очень большой), на случайных матрицах с гауссовским или равномерным распределением матричных элементов, показывают, что с практически во всех случаях наибольшей вероятностью находится наиболее глубокий минимум (см. рис. 9 и рис. 10).

Работа выполнена при поддержке РФФИ (проект 06-01-00109).

### Литература

1. *Hopfield J. J.* Neural Networks and physical systems with emergent collective computational abilities // *Proc. Nat. Acad. Sci. USA.* – v. 79, pp. 2554–2558 (1982).
2. *Hopfield J. J., Tank D. W.* Neural computation of decisions in optimization problems // *Biological Cybernetics.* – v. 52, pp. 141–152 (1985); *Hopfield J. J., Tank D. W.* Computing with neural circuits: A Model // *Science.* – v. 233, pp. 625–633 (1986).
3. *Fu Y., Anderson P. W.* Application of statistical mechanics to NP-complete problems in combinatorial optimization // *Journal of Physics A.* – v. 19, pp. 1605–1620 (1986).
4. *Poggio T., Girosi F.* Regularization algorithms for learning that are equivalent to multilayer networks // *Science.* – 247, pp. 978–982 (1990).
5. *Smith K. A.* Neural networks for combinatorial optimization: A review of more than a decade of research // *INFORMS Journal on Computing.* – v. 11 (1), pp. 15–34 (1999).
6. *Joya G., Atencia M., Sandoval F.* Hopfield neural networks for optimization: Study of the different dynamics // *Neurocomputing.* – v. 43, pp. 219–237 (2002).
7. *Hartmann A. K., Rieger H.* *New Optimization Algorithms in Physics.* – Wiley-VCH, Berlin, 2004.

8. Tang Huajin, Tan K. C., Yi Zhang. A columnar competitive model for solving combinatorial optimization problems // *IEEE Trans. Neural Networks.* – v. 15, pp. 1568–1574 (2004).
9. Kwok T., Smith K. A. A noisy self-organizing neural network with bifurcation dynamics for combinatorial optimization // *IEEE Trans. Neural Networks.* – v. 15, pp. 84–98 (2004).
10. Salcedo-Sanz S., Santiago-Mozos R., Bousoño-Calzon C. A hybrid Hopfield network-simulated annealing approach for frequency assignment in satellite communications systems // *IEEE Trans. Systems, Man and Cybernetics.* – v. 34, pp. 1108–1116 (2004).
11. Wang L. P., Li S., Tian F. Y., Fu X. J. A noisy chaotic neural network for solving combinatorial optimization problems: Stochastic chaotic simulated annealing // *IEEE Trans. System, Man, Cybern, Part B–Cybernetics.* – v. 34, pp. 2119–2125 (2004).
12. Wang L. P., Li S. A gradual noisy chaotic neural network for solving the broadcast scheduling problem in packet radio networks // *IEEE Trans. Neural Networks.* – vol. 17, pp. 989–1000 (2006).
13. Kryzhanovsky B., Magomedov B. Application of domain neural network to optimization tasks // *Proc. of ICANN'2005.* Warsaw. LNCS 3697, Part II, pp. 397–403 (2005).
14. Kryzhanovsky B., Magomedov B. Domain dynamics in optimization tasks. – *Lecture Notes in Computer Science.* – Springer, Berlin–Heidelberg. ISSN: 0302-9743, Volume 4029 / 2006, pp. 37–45 (2006).
15. Litinskii L. B., Magomedov B. M. Global Minimization of a Quadratic Functional: Neural Networks Approach // *Pattern Recognition and Image Analysis.* – v. 15(1), pp. 80–82 (2005).
16. Крыжановский Б. В., Магомедов Б. М., Микаэлян А. Л. Взаимосвязь глубины локального минимума и вероятности его обнаружения в обобщенной модели Хопфилда // *Доклады АН.* – т. 405, № 3, с. 320–324, 2005.
17. Kryzhanovsky B. V., Magomedov B. M., Fonarev A. B. On the Probability of Finding Local Minima in Optimization Problems // *Proc. of International Joint Conference on Neural Networks IJCNN-2006,* pp. 5888–5892. Vancouver, Canada.
18. Kryzhanovsky B. V., Kryzhanovsky V. M., Mikaelian A. L. Binary optimization: A relation between the depth of a local minimum and the probability of its detection // *4th International Conference on Informatics in Control, Automation and Robotics.* – pp. 5–10. ICINCO 2007, Anger, France.



19. Крыжановский Б. В. Представление наперед заданной матрицы в виде хэббовского разложения по конфигурационным весторам // IX Всероссийская научно-техническая конференция «Нейроинформатика-2007». – М.: МИФИ, т. 1, с. 178–182 (2007).
20. Kryzhanovsky B. V. Expansion of a matrix in terms of external products of configuration vectors // *Optical Memory & Neural Networks*. – vol. 16, No. 4, pp. 187–199, 2007.
21. Крыжановский Б. В. Представление заданной матрицы в виде взвешенного разложения по конфигурационным векторам. // *Многопроцессорные вычислительные и управляющие системы–2007*. Материалы Международной научно-технической конференции. Т. 2. - Таганрог: Изд-во ТТИ ЮФУ, 2007. – с. 9-13.
22. Perez-Vincente C. J. Finite capacity of sparse-coding model // *Europhys. Lett.* – v. 10, pp. 627–631 (1989).
23. McEllice R., Posner E., Rodemich E. R. Capacity of the Hopfield associative memory // *IEEE Trans. on Inf. Theory*. – v. 33, pp. 461–482 (1987).
24. Amit D. J., Gutfreund H., Sompolinsky H. // *Annals of Physics*. – v. 173, pp. 30–67 (1987).
25. Amit D. J., Gutfreund H., Sompolinsky H. Spin-glass models of neural networks // *Physical Review A*. – v. 32, pp. 1007–1018 (1985.)
26. Hebb D. O. *The Organization of Behavior*. – New York: Wiley, 1949.

**Борис Владимирович КРЫЖАНОВСКИЙ**, директор Центра оптико-нейронных технологий Научно-исследовательского института системных исследований РАН (Москва), член-корреспондент РАН, доктор физико-математических наук, Президент Российской ассоциации нейроинформатики. Область научных интересов: нейронные сети, квантовая механика, нелинейная оптика. Имеет более 80 научных публикаций.