

О минимизации квадратичного бинарного функционала¹

Л.Б. Литинский

Центр оптико-нейронных технологий
Научно-исследовательского института системных исследований,
Российская академия наук, Москва
Поступила в редколлегию 24.09.2008

Аннотация—Рассматривается задача минимизации квадратичного функционала, зависящего от большого числа N бинарных переменных. Для матриц связей спин-стекольного типа средствами компьютерного моделирования изучены 3 процедуры минимизации. Показано, что при прочих равных условиях заметным превосходством обладает максимальная динамика (greedy algorithm). Изучена зависимость результатов минимизации от расстояния между стартовыми точками и глобальным минимумом. Установлено, что характер распределения локальных минимумов критически зависит от этого расстояния.

1. ВВЕДЕНИЕ

Обсуждается задача минимизации квадратичной формы от большого числа N бинарных переменных $s_i = \{\pm 1\}$:

$$E(\mathbf{s}) = -\frac{(\mathbf{J}\mathbf{s}, \mathbf{s})}{2} = -\frac{1}{2} \sum_{i,j=1}^N J_{ij}s_i s_j \xrightarrow{\mathbf{s}} \min, \quad \mathbf{J} = (J_{ij})_1^N, \quad \mathbf{s} = (s_1, s_2, \dots, s_N). \quad (1)$$

Данная задача возникает во многих отраслях научного знания [1]. К минимизации функционала (1) сводится отыскание максимального разреза в графе и распределение вычислительных процессов по процессорам [2], [3], обработка результатов физических экспериментов [4], [5] и многие другие проблемы. В физике магнитных материалов [6] данная задача возникает в связи с отысканием *основного состояния* – бинарного вектора $\mathbf{s}$, который доставляет функционалу (1) глобальный минимум.

Без ограничения общности матрицу связей $\mathbf{J} = (J_{ij})_1^N$ можно считать симметричной, с нулевыми элементами на главной диагонали: $J_{ij} = J_{ji}$, $J_{ii} = 0$. Используя физическую терминологию бинарные переменные $s_i = \{\pm 1\}$ будем называть *спинами*, N -мерные векторы $\mathbf{s} = (s_1, s_2, \dots, s_N)$, которые задают состояние системы как целого - *конфигурационными векторами* или *конфигурациями*, функционал $E(\mathbf{s})$ – *энергией* состояния $\mathbf{s}$, а конфигурацию, доставляющую глобальный минимум функционалу (1) – *основным состоянием*.

Каждая содержательная проблема порождает свою, специфическую матрицу связей $\mathbf{J}$. Общим для всех матриц является отсутствие эффективной вычислительной процедуры, с помощью которой можно за приемлемое время находить основное состояние. Обычно число локальных минимумов экспоненциально велико по N , и пробиться к самому глубокому минимуму очень трудно. Полный перебор всех 2^N конфигураций возможен только для небольших размерностей $N \sim 20$. Для размерностей $N \sim 10^2 - 10^3$, как правило, ограничиваются отысканием

¹ Работа выполнялась в рамках программы "Интеллектуальные компьютерные системы"(проект 2.45) при финансовой поддержке грантом РФФИ 06-01-00109.

как можно более глубокого локального минимума, без гарантии, что найденный минимум является глобальным. Разработке эффективных методов минимизации функционала (1) посвящена огромная литература. Представление о последних достижениях в этой области можно составить по недавней коллективной монографии [7]. Обзоры более ранних работ см. в [8]-[10].

Пожалуй, наибольшие успехи достигнуты для модели Эдвардса-Андерсона спинового стекла (ЕА-модель) [11]. В этой модели спины расположены в узлах регулярной решетки (обычно квадратной или кубической); кроме того предполагается, что взаимодействуют только ближайшие спины. Подавляющее большинство элементов матрицы связей J_{ij} равны нулю, а ненулевые элементы случайным образом выбираются из стандартного нормального распределения. Для таких матриц разработан эффективный алгоритм направленного перебора [12], который позволяет даже для очень больших размерностей ($N > 2000$) находить основное состояние за приемлемое время. Существует интернет-сайт, куда каждый может прислать свою матрицу, а в ответ получит основное состояние функционала (1) (см. [13]). Нам приходилось пользоваться услугами этого сайта.

Другой интенсивно исследуемый физиками объект - модель Шеррингтона-Киркпатрика спинового стекла (SK-модель) [14]. Здесь предполагается, что все спины взаимодействуют друг с другом, а матричные элементы суть случайные числа, выбранные из стандартного нормального распределения. Матрица в этом случае плотно заполнена. С физической точки зрения эта модель нереалистична, однако ее свойства изучены теоретически. Энергетическая поверхность представляет собой систему иерархически вложенных друг в друга "долин", а число локальных минимумов экспоненциально растет с увеличением N . Про основное состояние SK-модели известно только, что оно должно быть невырожденным. Неизвестна ни соответствующая конфигурация, ни величина глобального минимума по энергии. Не существует алгоритма, позволяющего для матриц большой размерности находить основное состояние. В тех случаях, когда по постановке задачи необходимо знать глобальный минимум, проводят большое число случайных испытаний, и в качестве основного состояния принимают конфигурацию, отвечающую самому глубокому минимуму ([15], [16]). Мы в своих экспериментах поступали точно так же (см. раздел 4).

Заканчивая краткий перечень современных методов минимизации, отметим очень интересный GRA-алгоритм, который развивается авторами [17], [18]. Несмотря на то, что алгоритм является эвристическим, он основан на глубоких физических идеях, и, судя по публикациям, демонстрирует хорошие результаты. Важно, что GRA-алгоритм является универсальным и не опирается на специфику матрицы связей.

Любой алгоритм минимизации основан на том или ином варианте процедуры локальной минимизации. Систему стартуют с некоторого начального состояния $\mathbf{s}(0)$, переворачивают по очереди спины, и каждый раз проверяют – уменьшается значение функционала (1) или нет? Если уменьшается – данному спину присваивают новое (перевернутое) значение; система как целое переходит в новое состояние. Если же значение функционала не уменьшается – восстанавливают прежнее значение спиновой переменной. Процедуру локальной минимизации можно реализовать по-разному; соответственно, результат минимизации будет зависеть от способа реализации [17]-[20]. Основная цель нашей работы - экспериментальное исследование трех вариантов процедуры локальной минимизации. В следующем разделе дается формальная постановка задачи. В следующих затем разделах описываются компьютерные эксперименты и обсуждаются полученные результаты.

В работах [21], [22] теоретически изучалась энергетическая поверхность функционала (1) для модели Хопфилда нейронной сети [1]. Основу аппарата составил теоретико-вероятностный подход, который используется для оценки емкости памяти модели Хопфилда [1]. Авторам удалось получить целый ряд интересных результатов о минимумах функционала (1). В частности,

оказалось возможным связать глубину локального минимума с вероятностью его отыскания в случайном поиске. Полученное аналитическое выражение свидетельствует о том, что чем глубже локальный минимум, тем шире его область притяжения, тем больше вероятность его отыскания. Несмотря на кажущуюся очевидность этого утверждения, оно весьма нетривиально. В частности, из него следует, что при большом числе случайных испытаний вероятность попадания системы в основное состояние (или близко к нему) больше вероятности попадания в менее глубокие минимумы. До конца статьи будем называть это *правилом глубоких минимумов*.

Для Хеббовских матриц связей, фигурирующих в модели Хопфилда, авторы [21], [22] получили хорошее согласие своей теории с компьютерным экспериментом. В настоящее время исследуются другие типы матриц - см. [23], [24]. В данной работе мы проверяли - при каких условиях правило глубоких минимумов выполняется для спин-стекольных матриц? В этом состояла еще одна цель нашей работы.

2. РАЗЛИЧНЫЕ ДИНАМИКИ

Впервые процедура локальной минимизации была использована для минимизации функционала (1) в [25]. На русском языке аналогичный подход впервые был предложен в [26]. В теории нейронных сетей данная процедура известна под названием последовательной динамики: систему стартуют с начальной конфигурации $\mathbf{s}(0)$, в случайном порядке перебирают спины, и каждому спину присваивают значение, равное знаку действующего на него локального поля

$$h_i(t) = \sum_{j=1}^N J_{ij}s_j(t).$$

Формально динамическое правило имеет вид:

$$s_i(t+1) = \begin{cases} s_i(t), & \text{если } s_i(t)h_i(t) \geq 0; \\ -s_i(t), & \text{если } s_i(t)h_i(t) < 0. \end{cases}$$

Если спин не совпадает по знаку с действующим на него локальным полем – если спин неудовлетворен: $s_i(t)h_i(t) < 0$, – его переворачивают: $s_i(t+1) = -s_i(t)$. В противном случае значение спина не меняется, и переходят к следующему спину. Данный алгоритм – один из исследуемых в настоящей статье. Будем называть его *случайной динамикой*.

Легко видеть, что переворот неудовлетворенного спина понижает энергию состояния:

$$E(\mathbf{s}(t+1)) = E(\mathbf{s}(t)) - 2|h_i(t)|. \quad (2)$$

Рано или поздно система окажется в состоянии, в котором все спины будут удовлетворены, и эволюция системы прекратится. Такое состояние является локальным минимумом по энергии.

Максимальная динамика, которая также изучается в настоящей работе, отличается от случайной динамики тем, что на каждом шаге переворачивают тот неудовлетворенный спин, переворот которого приводит к наибольшему понижению энергии. Иными словами, если в состоянии $\mathbf{s}(t)$ имеется несколько неудовлетворенных спинов, переворачивают тот из них, которому отвечает максимальное по модулю локальное поле $h_i(t)$ – см. выражение (2). Мотивировка для введения максимальной динамики очевидна: если уж переворот неудовлетворенного спина приводит к понижению энергии $E(\mathbf{s})$, а наша цель – найти как можно более глубокий минимум, то успешной может оказаться стратегия максимального понижения энергии *на каждом шаге*. Максимальная динамика исследовалась другими авторами, и известна под названием greedy-алгоритм [19], [20].

Наконец мы изучали еще *минимальную динамику*, когда переворачивают тот неудовлетворенный спин, переворот которого сопровождается наименьшим понижением энергии. В работе [20], где такая динамика названа *reluctant-алгоритмом*, установлено, что для SK-модели спинового стекла данная динамика дает лучшие результаты, чем остальные динамики. Этот результат вызвал удивление у самих авторов [20], поэтому важно было его проверить. Забегая вперед, скажем, что для SK-модели минимальная динамика действительно дает наилучшие результаты, но только когда динамическую систему запускают из точек *общего положения* - см. раздел 4.

На протяжении всей статьи будем пользоваться тремя введенными выше терминами: случайная, максимальная и минимальная динамики. Мы сравнивали их, минимизируя функционал (1) для матриц, отвечающих двумерной ЕА-модели спинового стекла с периодическими граничными условиями, и для матриц, отвечающих SK-модели спинового стекла (см. Введение). Несмотря на то, что оба типа матриц моделируют спиновое стекло, название *спин-стекольные матрицы* мы сохраним только за матрицами ЕА-модели. Матрицы, отвечающие SK-модели, для удобства различения будем называть *гауссовыми матрицами*.

Мы стремились выяснить, как результат минимизации зависит от типа матрицы и от ее размерности N ? Какая из динамик и при каких условиях дает лучшие результаты? Для каждой матрицы на предварительной стадии мы находили основное состояние. Это позволило поставить следующий вопрос: как результат минимизации зависит от расстояния стартовых точек до основного состояния? Расстояние между двумя конфигурациями $\mathbf{s}$ и $\mathbf{s}'$ всегда (кроме особо оговоренных случаев) будем измерять относительным расстоянием Хемминга:

$$d = \frac{N - |(\mathbf{s}, \mathbf{s}')|}{2N}, \text{ где } (\mathbf{s}, \mathbf{s}') = \sum_{i=1}^N s_i s'_i. \quad (3)$$

Иначе говоря, d есть относительное число координат, которые у двух конфигураций не совпадают. Легко видеть, что $d \in [0, 0.5]$; конфигурации, отстоящие друг от друга на расстояние 0.5, являются ортогональными.

Все три определенные выше динамики принадлежат к классу одношаговых процедур, когда на каждом такте переворачивают только один спин. Иногда рассматривают k -оптимальные процедуры, когда разом преворачивается группа из k спинов, или даже процедуры с переменной глубиной поиска, когда размер переворачиваемой группы спинов не фиксирован, а выбирается динамически. Некоторое представление о существующих здесь возможностях можно составить по [17], [18]. В следующих разделах описываются компьютерные эксперименты и полученные результаты.

3. РЕЗУЛЬТАТЫ ДЛЯ СПИН-СТЕКОЛЬНЫХ МАТРИЦ

В это разделе речь идет о случайных матрицах, отвечающих двумерной ЕА-модели спинового стекла с периодическими граничными условиями и взаимодействием между ближайшими спинами. Мы исследовали 9 матриц - по 3 матрицы размерностями $N = 100, 400$ и 900 . Благодаря возможностям, предоставленным сайтом [13], для каждой матрицы мы знали основное состояние функционала (1).

Для каждой матрицы генерировали 10000 случайных стартовых конфигураций $\mathbf{s}(0)$, удаленных на расстояние d от основного состояния, и запускали с них все три динамики: случайную, максимальную и минимальную. Для каждой динамики получали 10000 локальных минимумов, для которых фиксировали: энергию E ; расстояние D до основного состояния; число шагов S , за которое динамическая система спустилась в локальный минимум (о том, чем интересна S -характеристика, см. ниже). Все это проделывалось для следующих значений

стартового расстояния: $d = 0.05, 0.1, 0.2, 0.3, 0.4$ и 0.5 ; под *стартовым расстоянием* всегда будем понимать расстояние от стартовой конфигурации до основного состояния.

1) Подробные результаты для одной из матриц размерности $N=100$ приведены на рис.1-6; поясним, что именно показано на различных панелях рисунков. Левый столбец каждого рисунка отвечает максимальной динамике, средний столбец – случайной динамике, правый столбец – минимальной динамике. По оси абсцисс всюду отложена энергия локальных минимумов E . Энергия основного состояния для данной матрицы равна $E_{GS} = -1.232$.

На верхних панелях рисунков по оси ординат отложена вероятность попадания в локальные минимумы. Здесь же приведены число разных локальных минимумов K_{lm} и энергия самого глубокого из них E_{min} . На нижних панелях по оси ординат отложено расстояние D от локальных минимумов до основного состояния. Таким образом, на нижних панелях рисунка можно наблюдать распределение локальных минимумов в координатной плоскости $\{E, D\}$.

1а) На рис.1 приведены результаты стартов с расстояния $d=0.05$. Как видим, основное состояние удастся найти во всех трех динамиках ($E_{min} = E_{GS}$ на всех панелях рисунка). Однако, вероятность попадания в основное состояние в максимальной динамике (~ 0.9) в 3 раза больше аналогичной вероятности для случайной динамики (~ 0.3), которая в 2 раза больше вероятности попадания в основное состояние в минимальной динамике.

В максимальной динамике получено всего 43 локальных минимума. Это в 20 раз меньше числа минимумов, найденных в случайной динамике, которое в 2 раза меньше числа минимумов, найденных в минимальной динамике. Указанные 43 минимума являются самыми глубокими: они локализируются на левом краю широкого энергетического интервала, заполненного энергиями локальных минимумов, найденных в случайной и/или минимальной динамике. Можно сказать, что для стартового расстояния $d=0.05$ максимальная динамика позволяет находить в среднем более глубокие минимумы, чем случайная и/или минимальная динамики.

Проанализируем теперь нижние панели рис.1. Видно, что большая часть 43-х минимумов, найденных в максимальной динамике, находятся ближе к основному состоянию, чем стартовые точки (для которых $d=0.05$). Напротив, подавляющее большинство минимумов, найденных в случайной и минимальной динамиках, удалены от основного состояния на большее расстояние, чем стартовые точки. Для этих локальных минимумов, как правило, выполняется: $D > 0.05$. Можно сказать, что максимальная динамика сдвигает стартовую точку по направлению к основному состоянию, в то время как в случайной (или минимальной) динамике система, как правило, уходит от основного состояния.

Все сказанное можно повторить для стартов с расстояний $d=0.1$ и $d=0.2$ – см. рис.2 и рис.3. По мере того, как стартовые точки отдаляются от основного состояния, превосходство максимальной динамики постепенно уменьшается. Однако можно сказать, что для стартовых расстояний $d \leq 0.2$ максимальная динамика существенно превосходит и случайную, и минимальную динамики.

Для стартового расстояния $d=0.3$ максимальная динамика все еще заметно превосходит другие динамики – см. рис.4. Действительно, основное состояние теперь удастся найти только в максимальной динамике. Кроме того, распределение локальных минимумов в координатной плоскости $\{E, D\}$ для максимальной динамики более компактно, чем для остальных динамик (ср. между собой нижние панели рис.4). Отметим, однако, что теперь и для максимальной динамики число локальных минимумов K_{lm} становится почти равным числу стартов K : $K_{lm} \approx K = 10000$.

Для еще больших стартовых расстояний – $d=0.4$ и $d=0.5$ – разница между тремя динамиками перестает ощущаться – см. рис.5 и рис.6. Вообще говоря, самый глубокий локальный минимум и здесь отыскивается именно в максимальной динамике. Однако и общее число ло-

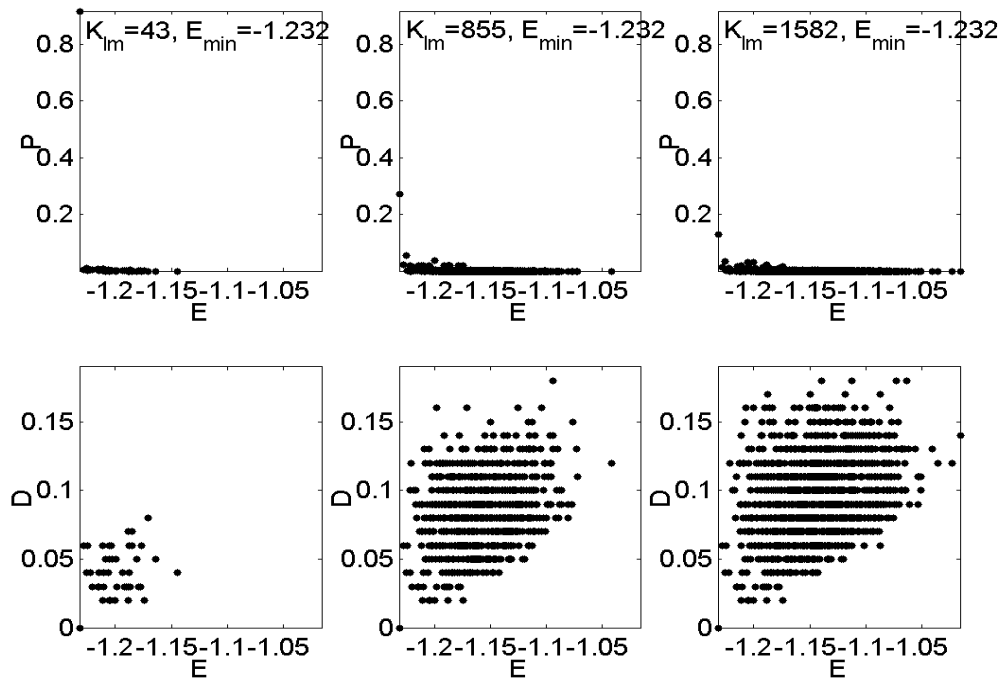
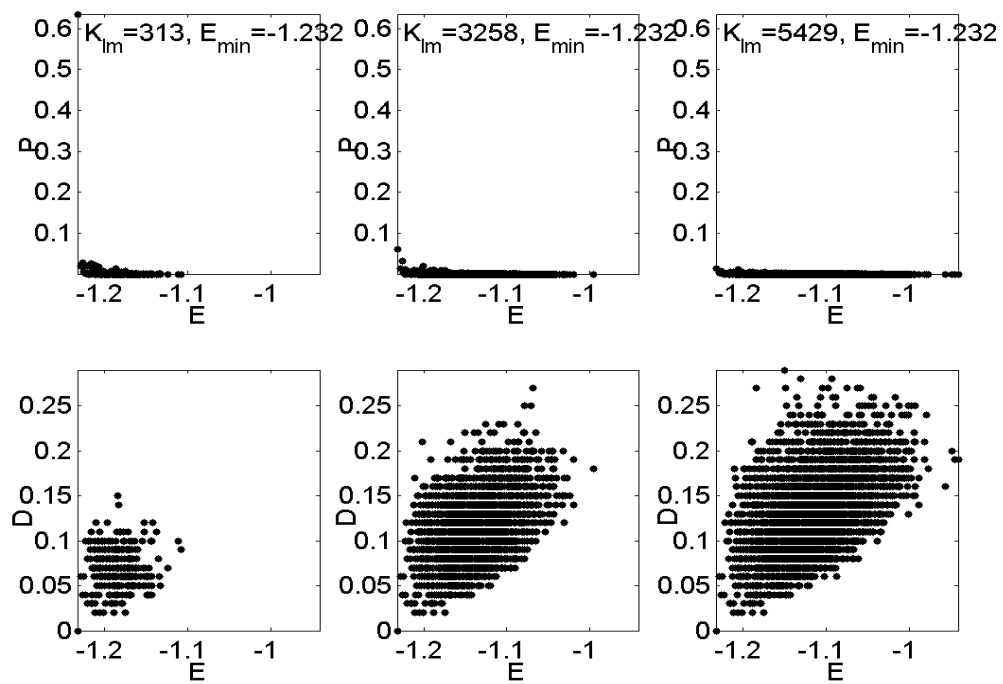
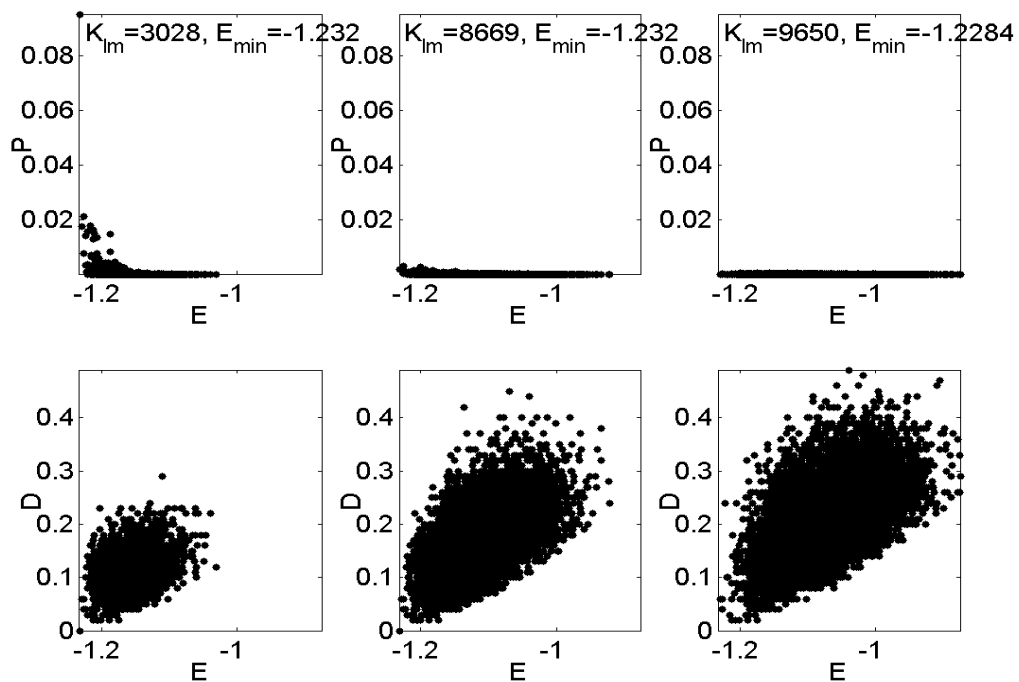


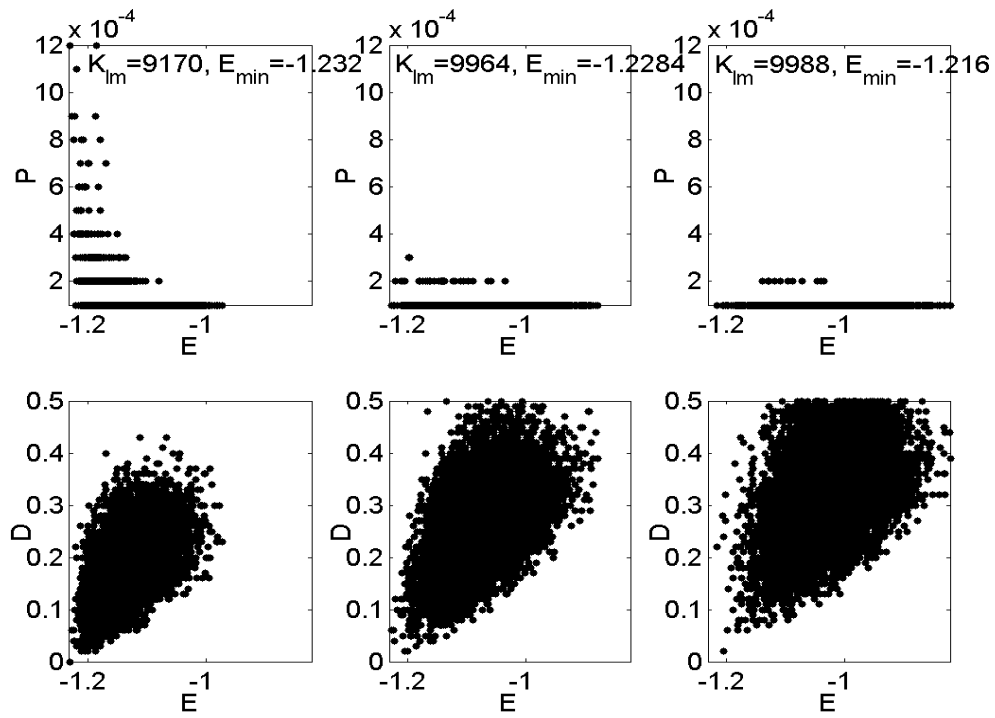
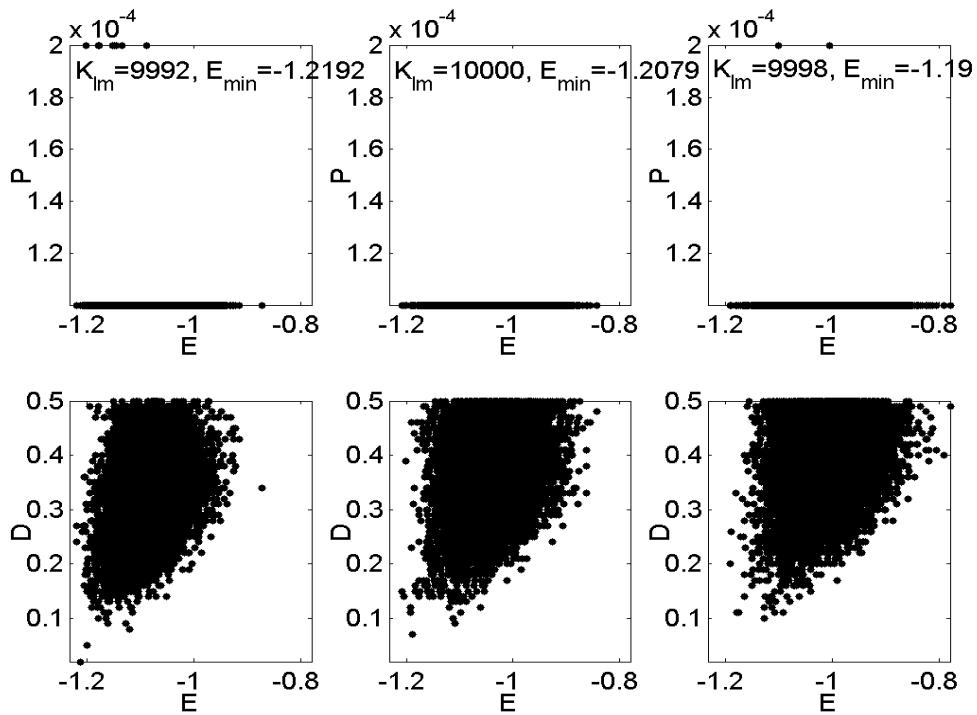
Рис.1. Результаты 10000 стартов с расстояния $d=0.05$ от основного состояния для спин-стекольной матрицы $N=100$. Левый столбец – максимальная динамика, средний столбец – случайная динамика, правый столбец – минимальная динамика. По оси абсцисс – энергии локальных минимумов. По оси ординат: на верхних панелях – вероятность P попадания в локальный минимум; на нижних панелях – расстояние D от локального минимума до основного состояния. K_{lm} – число различных локальных минимумов, E_{min} – минимальная найденная энергия.

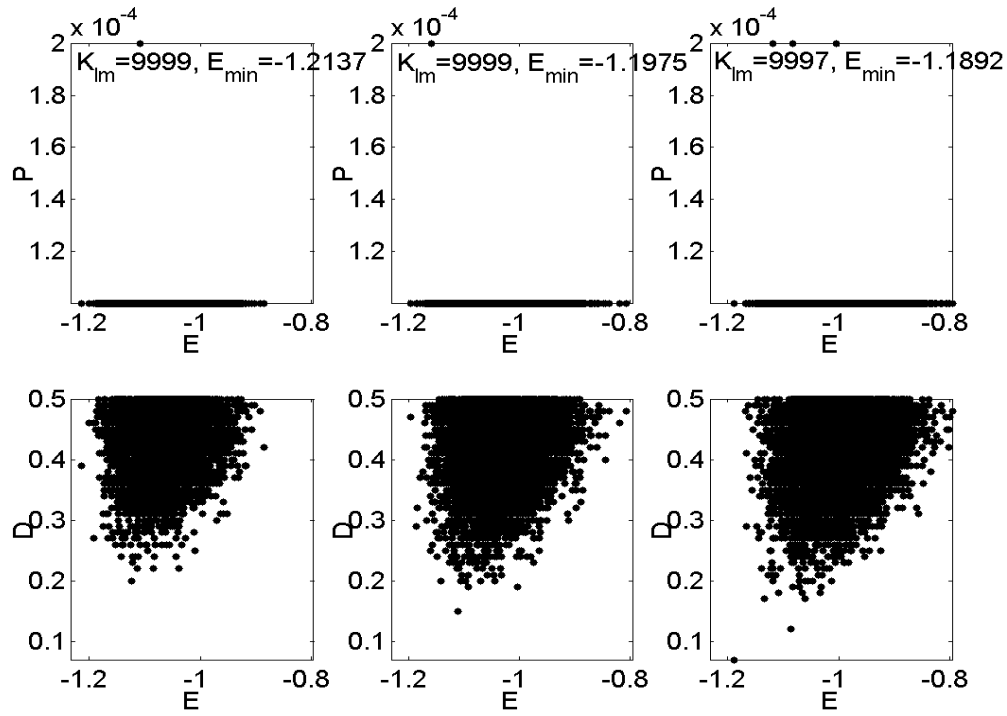
кальных минимумов, и характер их распределения в плоскости $\{E, D\}$ практически одинаковы для всех трех динамик.

16) Обратим внимание на то, что для стартовых расстояний $d \leq 0.3$ вероятность попадания в самые глубокие минимумы больше вероятности попадания в менее глубокие минимумы – об этом свидетельствует характерный пик в начале графиков на верхних панелях рис.1-4. Для этих стартовых расстояний выполняется правило глубоких минимумов – см. Введение. Напротив, для $d > 0.3$ правило глубоких минимумов не выполняется – об этом свидетельствуют графики на верхних панелях рис.5 и рис.6. Число локальных минимумов здесь почти не отличается от числа стартов $K=10000$: все минимумы становятся равновероятными, с вероятностью попадания в них $\sim 1/K$.

Может быть, невыполнение правила глубоких минимумов связано с недостаточной статистикой, с тем, что 10000 стартов мало? Чтобы устранить эти сомнения был проделан дополнительный эксперимент, а именно: с максимального стартового расстояния $d=0.5$ случайную динамику запускали 1 миллион раз. Результаты эксперимента приведены на рис.7. На обеих панелях по оси абсцисс отложена энергия локальных минимумов. На нижней панели показана гистограмма плотности распределения локальных минимумов по энергии; здесь же приведено их общее число K_{lm} . На верхней панели по оси ординат отложена вероятность попадания в локальные минимумы; здесь же приведена энергия самого глубокого локального минимума (из числа найденных). Все вычисления велись с использованием 8 значащих десятичных цифр мантиссы.

Рис.2. То же, что на рис.1 для стартового расстояния $d=0.1$.Рис.3. То же, что на рис.1 для стартового расстояния $d=0.2$.

Рис.4. То же, что на рис.1 для начального расстояния $d=0.3$.Рис.5. То же, что на рис.1 для начального расстояния $d=0.4$.

Рис.6. То же, что на рис.1 для начального расстояния $d=0.5$.

Мы видим, что за миллион стартов было получено огромное число локальных минимумов: $K_{lm} \approx 900000$, но основное состояние не было найдено ни разу: $E_{\min} > E_{GS}$. График на верхней панели рисунка свидетельствует о том, что вероятности попадания в самые глубокие минимумы меньше вероятностей попадания в минимумы, находящиеся в середине энергетического интервала. Иными словами, правило глубоких минимумов здесь не выполняется.

Резюмируем сказанное. Для начальных расстояний d , меньших некоторого критического значения d_c , выполняется правило глубоких минимумов. В этом случае максимальная динамика заметно превосходит другие динамики. Напротив, для начальных расстояний больших критического значения, $d > d_c$, правило глубоких минимумов не выполняется, а вероятность попадания в самые глубокие минимумы очень мала. Превосходство максимальной динамики становится здесь едва заметным. Признаком того, что мы находимся в области $d > d_c$, является совпадение числа локальных минимумов с числом стартов: $K_{lm} \approx K$.

Описанная картина воспроизводится и для двух других матриц размерности $N=100$: пока начальное расстояние не превосходит критического значения $d_c \approx 0.3$, максимальная динамика заметно превосходит и случайную, и минимальную динамики. С увеличением начального расстояния превосходство максимальной динамики постепенно уменьшается, и в области $d > d_c$ разница между динамиками практически стирается.

2) В дальнейшем будем пользоваться более компактным способом представления результатов, помещая на одном графике результаты для всех начальных расстояний $0.05 \leq d \leq 0.5$. Мы сосредоточимся на анализе нескольких основных характеристик. На рис.8 для всех трех динамик показана зависимость этих характеристик от начального расстояния d ; чтобы не дублировать изложение, на рисунке приведены результаты, полученные для второй спин-стекольной матрицы размерности $N=100$.

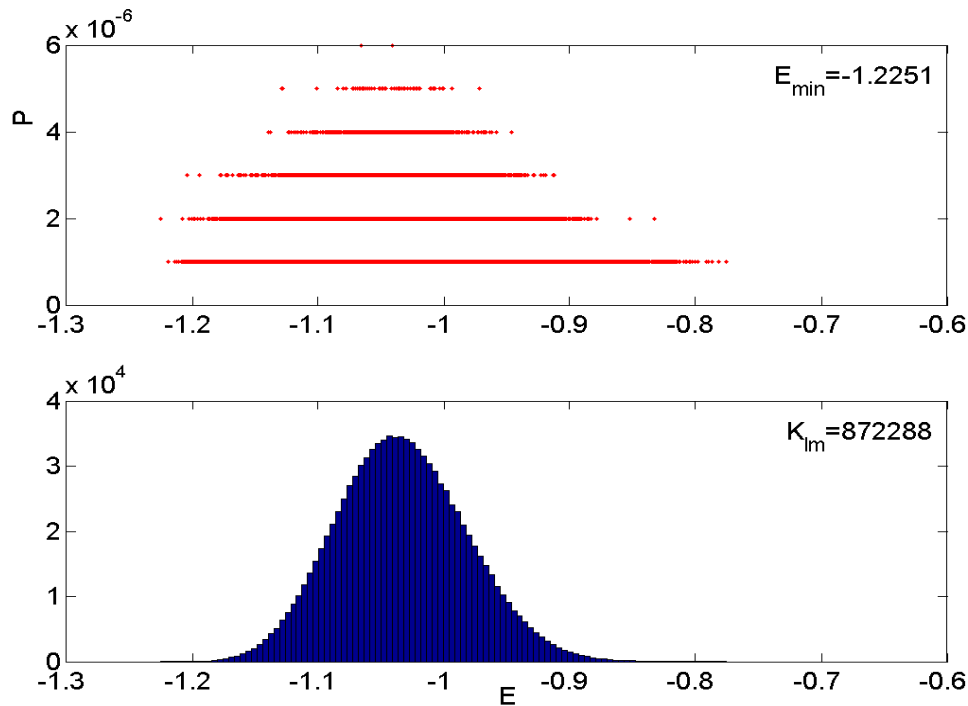


Рис.7. Результаты 10^6 стартов для спин-стекольной матрицы $N=100$. Внизу – гистограмма плотности распределения локальных минимумов по энергии;верху – вероятности попадания в локальные минимумы.

Сплошная линия отвечает максимальной динамике, штриховая линия – случайной динамике, штрих-пунктир – минимальной динамике. На двух верхних и нижней левой панелях рисунка по оси абсцисс отложено стартовое расстояние d в терминах относительного расстояния Хемминга (3). На нижней правой панели рисунка используется обычное расстояние Хемминга $d \cdot N$ (число несовпадающих координат у двух конфигураций); причины, по которым для этой панели сделано исключение, станут ясны ниже.

На левой верхней панели по оси ординат отложена P – вероятность отыскания основного состояния. Видно, что сначала максимальная динамика доминирует над другими динамиками, а затем, когда стартовое расстояние переваливает за критическое значение d_c , все три динамики выравниваются.

На нижней левой панели по оси ординат отложено число различных минимумов, найденных в каждой динамике. Видно, что и случайная динамика (штриховая линия), и минимальная динамика (штрих-пунктир) гораздо раньше максимальной динамики выходят на режим, когда число локальных минимумов перестает отличаться от числа стартов: $K_{lm} \approx K = 10000$. Однако для $d > 0.3$ это равенство выполняется уже и для максимальной динамики.

На правой верхней панели по оси ординат отложена разность между энергией основного состояния $E_{GS} = -1.2788$ и минимальной энергией, найденной за 10000 стартов: $E_{GS} - E_{min}$. Для малых значений стартового расстояния d (когда основное состояние еще удастся отыскать за 10000 стартов) эта разность равна нулю для всех динамик – см. начальную часть графиков. С ростом стартового расстояния, когда 10000 стартов уже не хватает для отыскания основного состояния, разность энергий становится отрицательной. Однако кривая, отвечающая максимальной динамике, мажорирует здесь остальные кривые. Иначе говоря, максимальная дина-

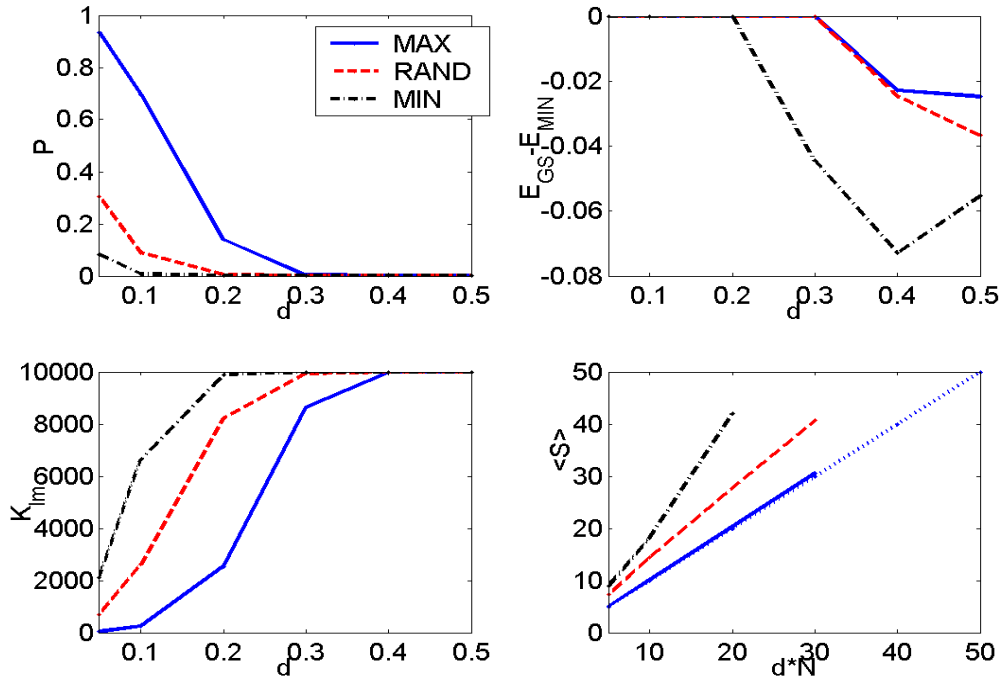


Рис.8. Зависимость результатов минимизации от начального расстояния для спин-стекольной матрицы $N=100$. По оси абсцисс: на правой нижней панели – расстояние Хемминга, на остальных панелях – относительное расстояние Хемминга d (3). Сплошная линия – максимальная динамика, штриховая линия – случайная динамика, штрих-пунктир – минимальная динамика (см. пояснения в тексте).

мика и здесь демонстрирует лучшие результаты (с ее помощью отыскиваются более глубокие минимумы).

Наконец, на правой нижней панели по оси ординат отложено среднее число шагов $\langle S \rangle$, за которое динамическая система скатывается в основное состояние. Фактически, речь идет о длине траектории, ведущей из стартовой точки в основное состояние, причем пройденный путь измеряется числом перевернутых спинов (числом шагов). Естественно, кривые на этом графике определены только до тех значений начального расстояния, для которых основное состояние еще удастся найти. Для минимальной динамики – до $d \cdot N = 20$, для случайной и максимальной динамик – до $d \cdot N = 30$ (ср. с графиками на правой верхней панели). Поясним, чем интересна S -характеристика.

Поскольку мы знаем стартовое расстояние $d \cdot N$, интересно понять – по какой траектории происходит спуск в основное состояние? Одно дело, если динамическая система движется прямо в основное состояние, сокращая расстояние до него на каждом шаге эволюции. Длина такой траектории будет приблизительно равна стартовому расстоянию: $S \approx d \cdot N$. Другое дело, если система движется по более запутанной траектории, то приближаясь к основному состоянию, то удаляясь от него. Движение может носить даже хаотический характер, лишь постепенно приближая динамическую систему к основному состоянию. В этих случаях число спинов, перевернутых за время спуска в основное состояние, будет больше стартового расстояния: $S > d \cdot N$. На правой нижней панели рис.8 как раз и показана средняя длина траектории, по которой динамическая система скатывается в основное состояние; усреднение происходило

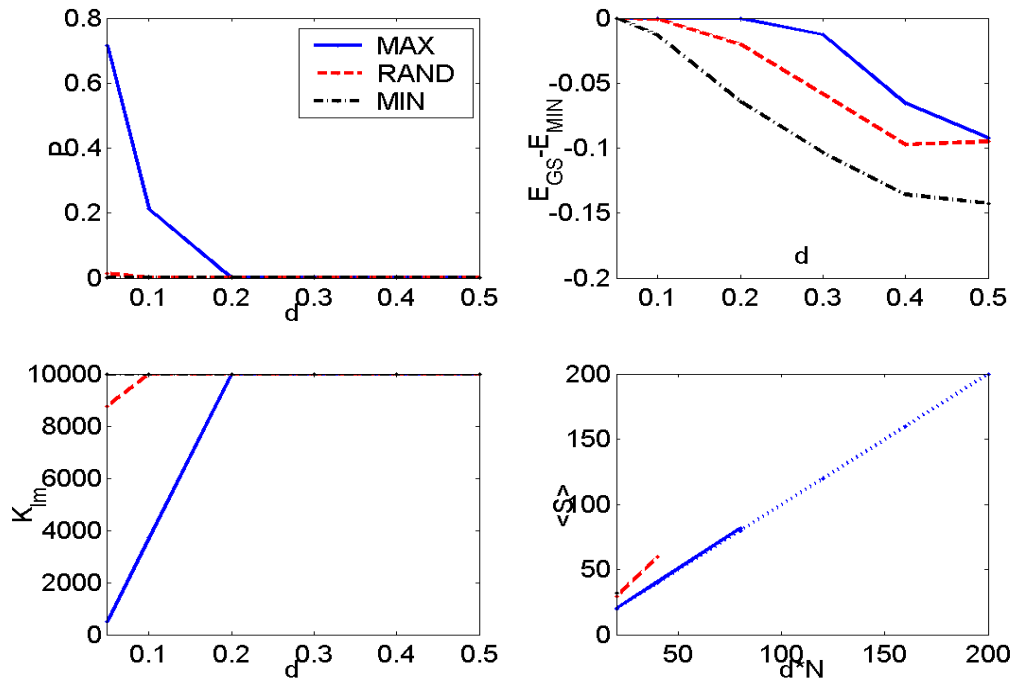


Рис.9. То же, что на рис.8 для спин-стекольной матрицы размерности $N=400$.

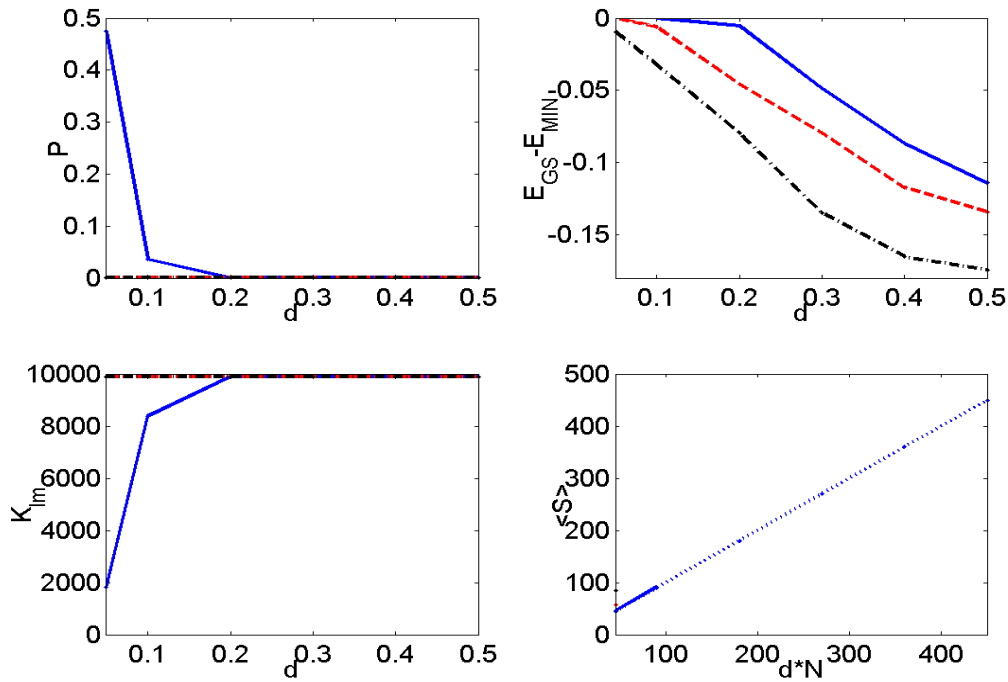
по всем попаданиям в основное состояние. Точечная прямая, идущая в правый верхний угол рисунка – биссектриса угла: $\langle S \rangle = d \cdot N$; с ней удобно сравнивать остальные кривые.

Мы видим, что $\langle S \rangle$ -кривая, отвечающая максимальной динамике, практически не отличается от прямой линии $\langle S \rangle = d \cdot N$. Иначе говоря, под действием максимальной динамики система по кратчайшему пути движется прямо в основное состояние. Для случайной динамики $\langle S \rangle$ -характеристика больше стартового расстояния $d \cdot N$. А для минимальной динамики – еще больше. В целом, однако, значения $\langle S \rangle$ -характеристики для трех динамик отличаются друг от друга не очень сильно; совсем иначе обстоит дело с гауссовыми матрицами – см. следующий раздел.

В точности та же картина имеет место и для спин-стекольных матриц больших размерностей. На рис.9 показаны аналогичные графики для одной из матриц размерности $N=400$, а на рис.10 – для одной из матриц размерности $N=900$. На графиках можно наблюдать все характерные особенности, отмеченные выше для матриц размерности $N=100$. Меняется только критическое значение стартового расстояния d_c , начиная с которого число локальных минимумов K_{lm} становится равным числу стартов K . По результатам обчета всех 3-х матриц для $N=400$ критическое расстояние можно оценить как $d_c \approx 0.2$, а для $N=900$ – как $d_c \approx 0.15$. Эти значения могут быть уточнены в более подробных компьютерных экспериментах.

4. РЕЗУЛЬТАТЫ ДЛЯ ГАУССОВЫХ МАТРИЦ

Так мы называем симметричные матрицы, элементы которых независимым и случайным образом выбираются из стандартного нормального распределения. Были исследованы 6 таких матриц – по 2 матрицы размерностями $N = 50, 100$ и 400 . Проблема, с которой пришлось столкнуться в первую очередь – как найти основное состояние функционала (1)?

Рис.10. То же, что на рис.8 для спин-стекольной матрицы размерности $N=900$.

1) Не существует алгоритма, который для гауссовых матриц большой размерности позволяет находить основное состояние. Вслед за другими исследователями ([15], [16], [18], [20]) мы использовали следующий рецепт: делается большое число случайных стартов, $K \sim 10^6$, и в качестве основного состояния выбирают конфигурацию, отвечающую самому глубокому минимуму. Желательно, чтобы число попаданий в самый глубокий минимум существенно превосходило $1/K$ – это явится хоть какой-то гарантией того, что более глубокого минимума не существует.

На рис.11-13 представлены результаты большого числа случайных стартов для матриц размерностями $N=50$, $N=100$ и $N=400$ соответственно (по одной матрице каждой размерности). Для размерностей $N=50$ и $N=100$ делалось 500000 стартов, для размерности $N=400$ – миллион стартов. Представленные на рисунках графики аналогичны тем, что даны на рис. 7. По оси абсцисс отложена энергия локальных минимумов. На нижней панели приведена гистограмма плотности распределения локальных минимумов по энергии. На верхней панели по оси ординат отложена вероятность попадания в локальные минимумы. Кроме того, приводится число локальных минимумов K_{lm} . Все расчеты велись в одинарной точности (с 8 значащими десятичными цифрами мантииссы). Использовалась случайная динамика, а стартовые конфигурации можно считать "точками общего положения", поскольку среднее расстояние от них до основного состояния равно 0.5: $\langle d \rangle = 0.5$.

Мы видим, что для матриц размерностями $N=50$ и $N=100$ вероятности попадания в самые глубокие минимумы существенно больше вероятностей попадания в менее глубокие минимумы – см. распределения вероятностей на верхних панелях рис.11 и рис.12. Иными словами, правило глубоких минимумов здесь очевидным образом выполняется. Конечно, большая часть локальных минимумов приходится на область средних энергий – об этом свидетельствует горб на гистограмме (см. нижние панели рисунков). Однако вероятность попадания в отдельный

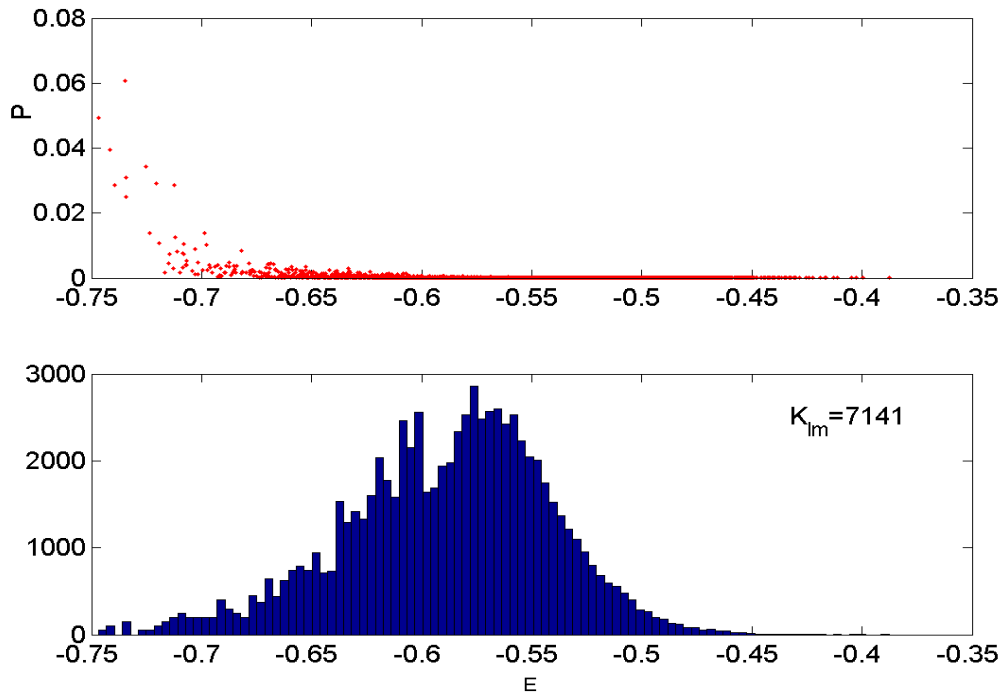


Рис.11. Для гауссовой матрицы размерности $N=50$ результаты 500000 случайных стартов. Внизу – гистограмма плотности распределения локальных минимумов по энергии;верху – вероятности попадания в локальные минимумы.

минимум больше всего для самых глубоких минимумов. Для $N=50$ имеем: $P(E_{\min}) \approx 0.05$, а для $N=100$: $P(E_{\min}) \approx 0.017$. Есть все основания полагать, что в этом случае конфигурация, отвечающая самому глубокому минимуму, является основным состоянием.

Иначе обстоит дело с матрицами размерности $N=400$. Правило глубоких минимумов здесь явно не выполняется – см. распределение вероятностей на верхней панели рис.13. Из миллиона случайных стартов в самый глубокий минимум динамическая система попадает только один раз. Можно ли доверять столь малой вероятности? Возможно, имеется более глубокий минимум? Чтобы ответить на эти вопросы было проделано следующее.

Согласно [20] для гауссовых матриц наилучшие результаты дает минимальная динамика; значит именно минимальную динамику необходимо использовать для отыскания основного состояния. Для обеих матриц размерности $N=400$ дополнительно было проделано по 500000 случайных стартов в минимальной динамике. Для первой матрицы и случайная, и минимальная динамики в качестве самого глубокого минимума дали одну и ту же конфигурацию; но в случайной динамике эта конфигурация была найдена всего 1 раз, а в минимальной динамике – 37 раз. Для второй матрицы в минимальной динамике был найден более глубокий минимум, чем в случайной динамике; в этот более глубокий минимум динамическая система скатывалась 50 раз. В результате, для матриц размерности $N=400$ в качестве "основного состояния" выбирали конфигурацию, найденную с помощью большого числа случайных стартов в минимальной динамике.

Прежде, чем переходить к завершающей части описания результатов для гауссовых матриц, отметим наличие здесь критической размерности N_c : если размерность задачи меньше N_c , правило глубоких минимумов выполняется; наоборот, если размерность матрицы больше N_c ,

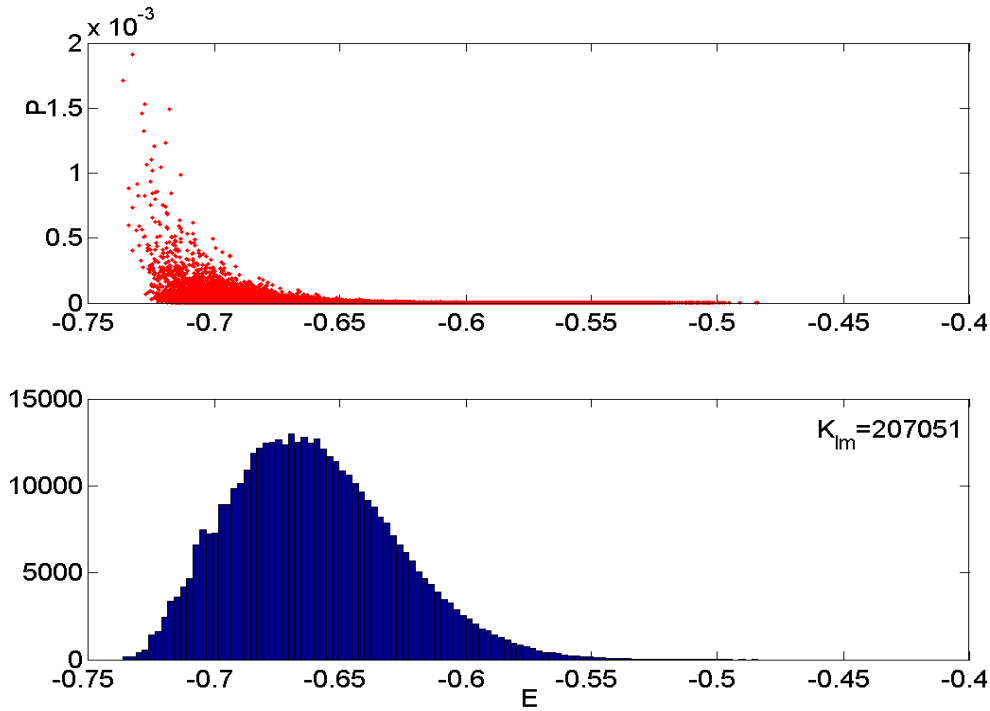


Рис.12. То же, что на рис.11 для гауссовой матрицы размерности $N=100$.

правило глубоких минимумов не выполняется. Мы оцениваем здесь значение критической размерности величинной $N_c \approx 320 - 350$.

2) Остальная часть эксперимента проходила точно так же, как для спин-стекольных матриц. Для каждой матрицы генерировали по 10000 стартовых конфигураций, отстоящих на расстояние d от основного состояния, и запускали с них все три динамики. Стартовым расстояниям d придавались те же, что и раньше значения; для локальных минимумов фиксировали те же, что и раньше характеристики E , D и S – см. начало предыдущего раздела. Результаты экспериментов для матриц размерностями $N = 50, 100$ и 400 приведены на рис.14-16 соответственно. Форма представления результатов та же, что на рис.8-10: сплошная линия отвечает максимальной динамике, штриховая – случайной, штрих-пунктирная линия – минимальной динамике. По оси абсцисс всюду отложено расстояние между стартовыми конфигурациями и основным состоянием. Как и раньше, только на правой нижней панели рисунков используется расстояние Хемминга $d \cdot N$; на остальных панелях по оси абсцисс отложено относительное расстояние Хемминга d (3).

На левой верхней панели по оси ординат отложена вероятность отыскания основного состояния P . На панели прямо под нею по оси ординат отложено число локальных минимумов. Для стартовых расстояний $d < 0.4$ мы видим знакомую картину: наилучшие результаты дает максимальная динамика, наихудшие результаты – минимальная динамика, а случайная динамика занимает промежуточное между ними положение. Ни о каком превосходстве минимальной динамики здесь не может быть и речи.

Неожиданным образом для стартовых расстояний $d \geq 0.4$ минимальная и максимальная динамики меняются местами. Для размерностей $N=50$ и $N=100$ это отчетливо видно по графикам на левых панелях рис.14, 15. Действительно, теперь и вероятность отыскания основного состояния больше всего для минимальной динамики, и число локальных минимумов меньше

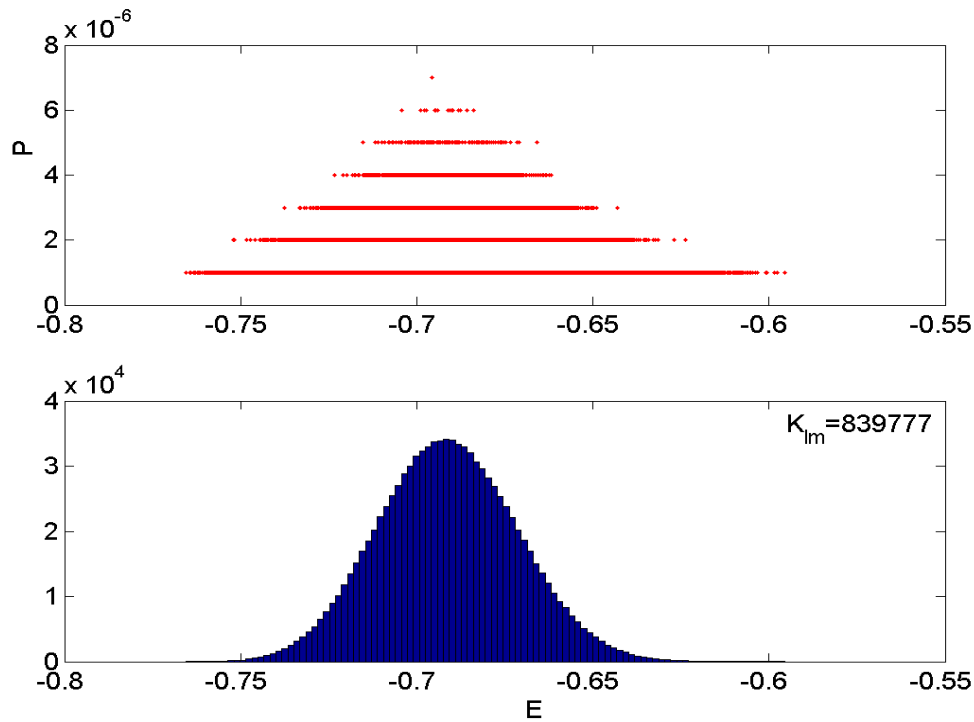


Рис.13. То же, что на рис.11 для гауссовой матрицы размерности $N=400$ (10^6 случайных стартов).

всего именно для нее. Для размерности $N=400$ о превосходстве минимальной динамики свидетельствуют графики на правой верхней панели рис.16. По оси ординат здесь отложена разность $E_{GS} - E_{\min}$ между энергией "основного состояния" и минимальной энергией, найденной за 10000 стартов; мы видим, что для стартовых расстояний $d \geq 0.4$ минимальная динамика позволяет находить "основное состояние", а случайная и максимальная динамики – нет.

Превосходство минимальной динамики для больших значений стартового расстояния ($d \geq 0.4$) наблюдалось для всех исследованных матриц. Для стартовых расстояний $d \approx 0.5$ превосходство минимальной динамики много раньше наблюдалось авторами [20]. Подчеркнем, однако, что для $d < 0.4$ минимальная динамика дает наихудшие результаты. Инвертирование минимальной и максимальной динамик при больших значениях стартового расстояния d составляет одно из характерных отличий гауссовых матриц от спин-стекольных.

Второе характерное отличие связано с поведением $\langle S \rangle$ -характеристики – см. правые нижние панели рис.14-16. Как и в случае спин-стекольных матриц, $\langle S \rangle$ -график для максимальной динамики (сплошная линия) почти не отклоняется от прямой $\langle S \rangle = d \cdot N$ (точечная прямая). Для случайной динамики средняя длина траектории, приводящей систему в основное состояние, несколько больше: штриховая линия на всех рисунках проходит чуть выше сплошной линии. А для минимальной динамики (штрих-пунктир) средняя длина траектории больше обеих указанных величин, по меньшей мере, на порядок – см. подавляющее превосходство штрихпунктирной линии на правых нижних панелях рис. 14-16. В этом случае динамическая система движется в основное состояние каким-то запутанным, кружным путем. Ничего более определенного о характере движения по энергетической поверхности мы сказать не можем.

Напомним, что для спин-стекольных матриц каждой размерности наблюдалось критическое значение стартового расстояния d_c , слева от которого правило глубоких минимумов вы-

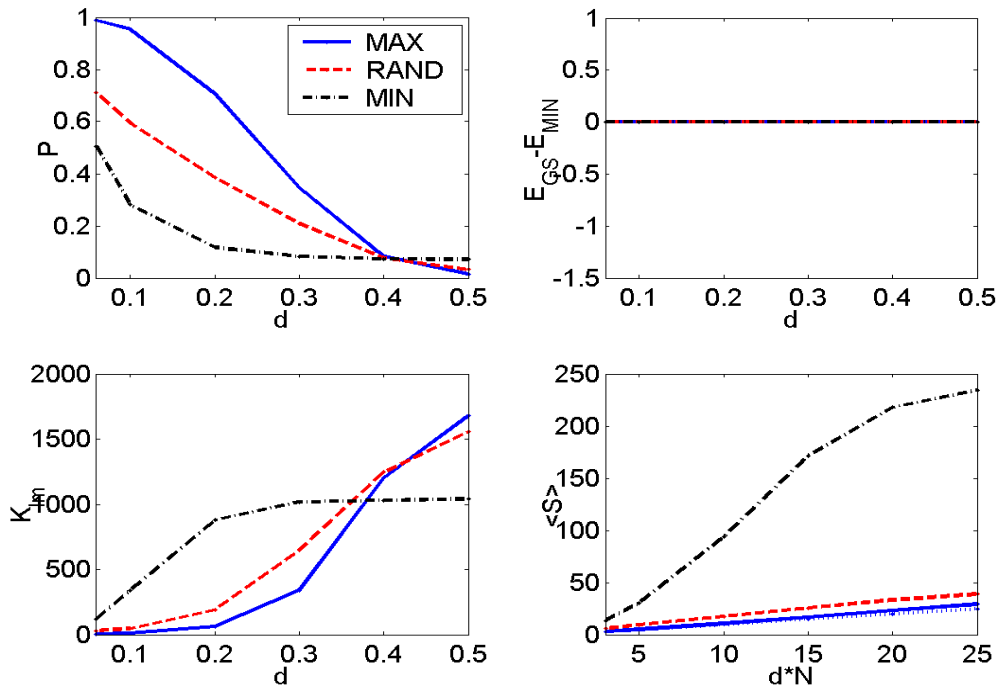


Рис.14. Зависимость результатов минимизации от стартового расстояния для гауссовой матрицы $N=50$. По оси абсцисс: на правой нижней панели – расстояние Хемминга, на остальных панелях – относительное расстояние Хемминга d (3). Сплошная линия – максимальная динамика, штриховая линия – случайная динамика, штрих-пунктир – минимальная динамика (см. пояснения в тексте).

полнялось, а справа – нет. Признаком того, что стартовое расстояние d больше d_c , служило равенство числа локальных минимумов числу стартов: $K_{lm} \approx K$. Для гауссовых матриц не очень больших размерностей $N = 50, 100$ правило глубоких минимумов выполняется для любых стартовых расстояний (см. верхние панели рис.11 и рис.12). Для гауссовых матриц размерности $N=400$ критическое значение стартового расстояния меньше 0.4: $d_c \approx 0.35$ – см. нижнюю левую панель рис.16.

Резюмируем сказанное. Пока стартовые точки отстоят от основного состояния на расстояние $d < 0.35$, максимальная динамика дает наилучшие результаты, минимальная динамика – наихудшие, а случайная динамика занимает промежуточное положение. Для $d > 0.35$ максимальная и минимальная динамики меняются местами. Для всех значений стартового расстояния d минимальная динамика работает по крайней мере на порядок медленнее остальных динамик.

5. ОБСУЖДЕНИЕ И ВЫВОДЫ

Средствами компьютерного моделирования исследованы 3 варианта процедуры одношаговой минимизации: максимальная, случайная и минимальная динамики. Исследование проводилось для матриц связей двух типов: спин-стекольных матриц (так у нас называются матрицы, отвечающие модели Эдвардса-Андерсона спинового стекла) и гауссовых матриц (отвечающих модели Шеррингтона-Киркпатрика спинового стекла). Размерности матриц принимали значения из интервала $N \in [50, 1000]$.

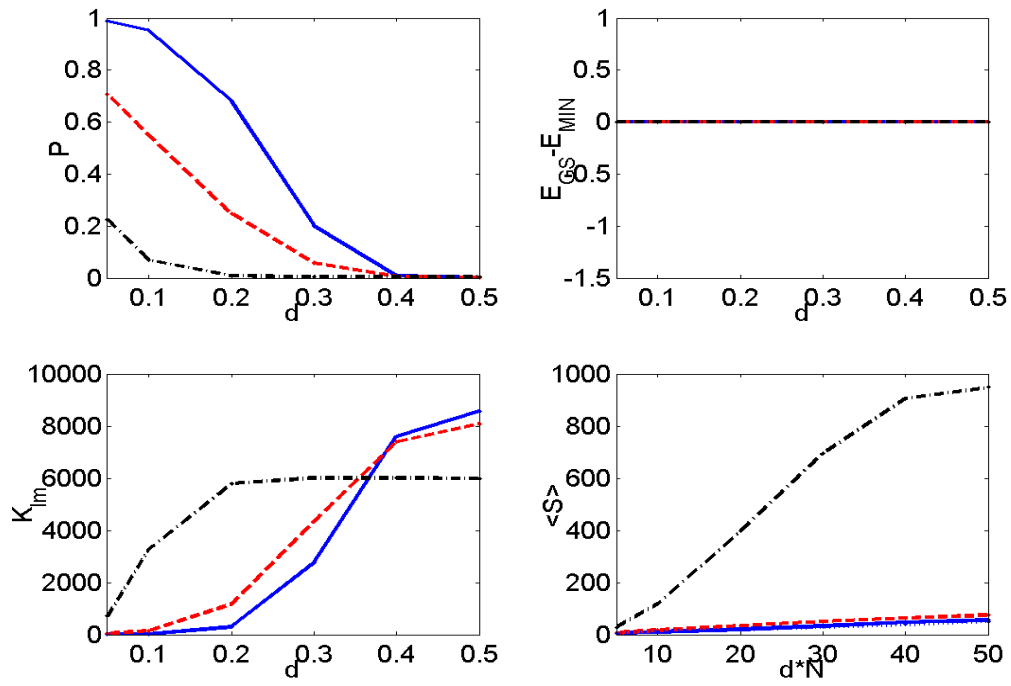


Рис.15. То же, что на рис.14 для гауссовой матрицы размерности $N=100$.

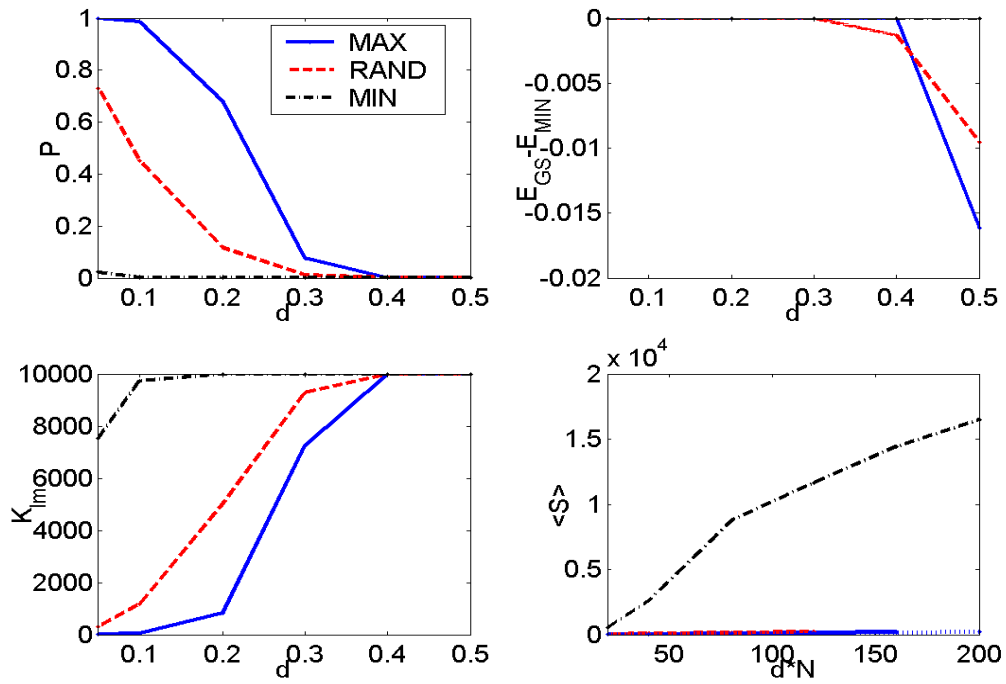


Рис.16. То же, что на рис.14 для гауссовой матрицы размерности $N=400$.

Основной вывод состоит в том, что максимальная динамика, как правило, демонстрирует заметно лучшие результаты, чем другие динамики. Для спин-стекольных матриц превосходство максимальной динамики имеет место для всех значений стартового расстояния. Для гауссовых матриц превосходство максимальной динамики имеет место только для стартовых расстояний $d < 0.35$; для больших значений стартового расстояния, $d \in [0.4, 0.5]$ более глубокие минимумы позволяет находить минимальная динамика (см. также [20]). Отметим, что для гауссовых матриц минимальная динамика работает на 1-2 порядка медленнее, чем максимальная или случайная динамики.

Для гауссовых матриц и больших значений стартового расстояния именно минимальная динамика позволяет получить наилучшие результаты; на фоне почти тотального превосходства максимальной динамики этот "всплеск" эффективности минимальной динамики выглядит достаточно неожиданным.

Равно неожиданным является и то, что в случае гауссовых матриц минимальная динамика приводит к очень длинным, запутанным траекториям, по которым система движется в основное состояние; длина этих траекторий больше чем на порядок превосходит расстояние между стартовой точкой и основным состоянием. Можно предположить, что причиной указанных отличий являются какие-то различия в общем устройстве энергетических поверхностей для двух типов матриц. Однако в чем состоят эти различия – мы не знаем. Общее впечатление таково, что для гауссовых матриц энергетическая поверхность является более "гладкой"; ее склоны менее "изрезаны" локальными минимумами, чем в случае спин-стекольных матриц. Именно поэтому для гауссовых матриц основное состояние удастся находить почти всегда, а для спин-стекольных матриц – только когда стартовые точки отстоят от основного состояния не слишком далеко.

Для обоих типов матриц при определенных условиях выполняется правило глубоких минимумов [21], [22]: динамическая система с большей вероятностью попадает в более глубокие минимумы, чем в менее глубокие. В то же время, это правило выполняется далеко не всегда. Например, для спин-стекольных матриц всех исследованных размерностей оно перестает выполняться, когда стартовые расстояния превосходят некоторое критическое значение $d_c(N)$; в такой ситуации вероятность попадания в основное состояние становится пренебрежимо малой. Для гауссовых матриц размерности $N \leq 300$ правило глубоких минимумов выполняется для всех стартовых расстояний. Для матриц размерности $N = 400$ это правило перестает выполняться, когда стартовые расстояния превосходят критическое значение $d_c \approx 0.35$.

Превосходство максимальной динамики тем сильнее, чем ближе к основному состоянию расположены стартовые точки. Именно в такой ситуации максимальная динамика наиболее эффективна. В связи с этим возникает вопрос: можно ли научиться выбирать стартовые точки поближе к основному состоянию, когда само основное состояние нам неизвестно? В работах [27], [28] введены и изучены так называемые EIGEN-конфигурации: конфигурационные векторы, ближайшие к собственным векторам матрицы связей. Некоторые из EIGEN-конфигураций обладают полезными свойствами, в особенности – наибольшие EIGEN-конфигурации, те, что близки к собственным векторам, отвечающим наибольшим собственным значениям. Например, 10-20 наибольших EIGEN-конфигураций лежат по энергии очень глубоко, на уровне энергий локальных минимумов (это справедливо для матриц любого типа и любой размерности). Установлено (см. [29], [30]), что старты с этих конфигураций приводят динамическую систему в самые глубокие локальные минимумы и т.д. В настоящей работе экспериментально установлено, что среди наибольших EIGEN-конфигураций непременно найдутся 2-3 конфигурации, отстоящие от основного состояния на расстояние $d \approx 0.25 - 0.3$; как правило, это EIGEN-конфигурации, ближайшие к 2-3 наибольшим собственным векторам. (Остальные EIGEN-конфигурации удалены от основного состояния гораздо дальше, на рас-

стояние порядка 0.5.) Этот результат справедлив для обоих типов исследованных матриц всех размерностей N . Мы надеемся использовать его для эффективного порождения целого набора случайных конфигураций, удаленных от основного состояния на расстояние $d \approx 0.25 - 0.3$. Эти конфигурации могут оказаться хорошими стартовыми точками для отыскания основного состояния.

Автор признателен Д.В. Вылегжанину за его участие на начальном этапе настоящих исследований, а также участникам семинара Центра оптико-нейронных технологий НИИСИ РАН, где данная работа докладывалась.

СПИСОК ЛИТЕРАТУРЫ

1. Hertz J., Krogh A., Palmer R. *Introduction to the Theory of Neural Computation*. NY: Addison-Wesley, 1991.
2. Костенко В.А. Задача построения расписания при совместном проектировании аппаратных и программных средств. *Программирование*, 2002 №3, стр. 64–80.
3. Крыжановский М.В., Мальсагов М.Ю. Задача распределения нагрузки в двухпроцессорной системе. *Вестник компьютерных и информационных технологий*, 2008, в печати.
4. Ососков Г.А., Полянский А., Пузынин И.В. Современные методы обработки экспериментальных данных в физике высоких энергий. *Физика элементарных частиц и атомного ядра*, 2002, т. 33(3), стр. 676–740.
5. Ососков Г.А., Стадник А.В. Эффективные нейросетевые алгоритмы для обработки экспериментальных данных. *Информационные технологии и вычислительные системы*, 2004, №1, стр. 103–125.
6. Бэкстер Р. *Точно решаемые модели статистической физики*. М.: Мир, 1985.
7. Hartmann A.K. and Rieger H. (Eds.) *New optimization algorithms in physics*. Berlin: Wiley-VCH, 2004.
8. Hartmann A.K., Rieger H. *Optimization Algorithms in Physics*. Berlin: Wiley-VCH, 2001.
9. Rieger H. Combinatorial Optimization. In: *Advances in Computer Simulation*, vol. 501 of Lecture Notes in Physics. Eds. J.Kertesz and I.Kondor. Heidelberg: Springer, 1998.
10. Smith K.A. Neural Networks for Combinatorial Optimization: A Review of More Than a Decade of Research. *INFORMS Journal on Computing*, 1999, v.11(1), pp.15–34.
11. Edwards S.F. and Anderson P.V. *J. de Physique*, 1975, v.5, pp. 965–978.
12. Liers F., Junger M., Reinelt G., and Rinaldi G. Computing Exact Ground States of Hard Ising Spin Glass Problems by Branch-and-cut. In [7], pp. 47–70.
13. Junger M. The Spin Glass Ground State Server. http://www.informatik.uni-koeln.de/lis_juenger/research/sgs/sgs.html.
14. Sherrington D., Kirkpatrick S. *Physical Review Letters*, 1975, v.36, pp.1972–1975.
15. Pal K.F. Hysteretic optimization for the Sherrington-Kirkpatrick spin glass. http://xxx.lanl.gov/PS_cache/cond-mat/pdf/0601/0601027v1.pdf.
16. Boettcher S. Extremal Optimization for Sherrington-Kirkpatrick Spin Glasses. *Euro. Phys. Journal B*, v.46, 2005, pp. 501–505. http://xxx.lanl.gov/PS_cache/cond-mat/pdf/0407/0407130v2.pdf.
17. Houdayer J., Martin O.C. Hierarchical approach for computing spin glass ground states. *Physical Review E*, v.64(5), 2001 056704-1 – 056704-7.
18. Martin O.C. Probing spin glasses with heuristic optimization algorithms. In: [7], pp. 23–45.
19. Horner H. Dynamics of spin glasses and related models of neural networks. In: *Computational Systems – Natural and Artificial*. Proceedings of the International Symposium on Synergetics. Ed. H. Haken. Berlin: Springer, 1988.

20. Parisi G. On the statistical properties of the large time zero temperature dynamics of the SK-model. http://xxx.lanl.gov/PS_cache/cond-mat/pdf/9501/9501045v1.pdf
21. Крыжановский Б.В., Магомедов Б.М., Микаэлян А.Л. Взаимосвязь глубины локального минимума и вероятности его обнаружения в обобщенной модели Хопфилда. *Доклады АН*, 2005, т. 405(3), стр.320–324.
22. Kryzhanovsky B.V., Magomedov B.M., Fonarev A.B. On the probability of finding local minima in optimization problems. *Proceedings of International Joint Conference on Neural Networks*, 2006, pp. 5888–5892, Vancouver: IEEE Press, 2006.
23. 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 ICINCO 2007*, pp.5–10, Anger: IEEE Press, 2007.
24. Kryzhanovsky B.V. Expansion of a matrix in terms of external products of configuration vectors. *Optical Memory & Neural Networks (Information Optics)*, 2007, v.16(4), pp.187–199.
25. Lin S. Computer solutions of the traveling salesman problem. *Bell System Technical Journal*, 1965, v.44, p.2245.
26. Браверман Э.М. Методы экстремальной группировки параметров и задача выделения существенных факторов. *Автоматика и телемеханика*, 1970, No.1, стр. 123–132.
27. Вылегжанин Д.В., Литинский Л.Б., Мурашкин А.А. О выборе стартовых состояний при минимизации квадратичного дискретного функционала. *Искусственный интеллект*, 2005, т.3, стр. 334–342.
28. Litinskii L.B. Eigenvalue problem approach to discrete minimization. In: *Proceedings of International Conference on Artificial Neural Networks ICANN-2005*, Eds. W.Duch et al. 2005, Berlin: Springer-Verlag, pp. 405–410.
29. Вылегжанин Д.В. *Исследование энергетической поверхности квадратичного дискретного функционала для нескольких типов случайных матриц*. Магистерская диссертация, М.: МФТИ (ГУ), 2006.
30. Мурашкин А.А. *О выборе стартовых состояний и динамик в задачах минимизации квадратичного дискретного функционала*. Магистерская диссертация, М.: МФТИ (ГУ), 2006.