

# Содержание

|  |           |
|--|-----------|
| <b>Введение . . . . .</b>                        | <b>3</b>  |
| <b>Глава 1. Теоретическая часть . . . . .</b>    | <b>5</b>  |
| 1.1 Квантовые состояния нейронной сети . . . . . | 5         |
| 1.2 Вариационный метод Монте-Карло . . . . .     | 8         |
| 1.3 Используемые алгоритмы оптимизации . . . . . | 10        |
| <b>Глава 2. Практическая часть . . . . .</b>     | <b>17</b> |
| 2.1 Исследуемые гамильтонианы . . . . .          | 17        |
| 2.2 Результаты обучения нейронной сети . . . . . | 18        |
| <b>Заключение . . . . .</b>                      | <b>23</b> |
| <b>Список литературы . . . . .</b>               | <b>25</b> |

# Введение

С увеличением объема данных и вычислительных мощностей методы машинного обучения все больше используются как средство для анализа. Терабайты разнообразной информации хранят незримые для человека закономерности, которые желательно выявлять для более глубокого понимания процессов, происходящих в окружающем нас мире. Методы машинного обучения являются, по-сути, единственным инструментом познания окружающего мира посредством анализа больших данных. Огромное количество данных непосредственно производится в научной сфере, вследствие чего применение методов машинного обучения для решения физических задач заинтересовало множество научных групп, связанных с информационными технологиями.

Так, недавно были проведены исследования по обнаружению нетривиальных состояний материи посредством машинного обучения [1]. Эти и подобные исследования позволяют предположить наличие потенциала в применении машинного обучения к физическим исследованиям и его дальнейшее использование в будущем [2].

Особый интерес представляет применение искусственных нейронных сетей для приближенного решения стационарного уравнения Шредингера многочастичных систем. В данной работе будет рассматриваться решение последнего для основного состояния системы спинов  $S = 1/2$ :

$$\hat{\mathcal{H}}|\psi_0\rangle = E_0|\psi_0\rangle$$

Для системы спинов размерность гильбертова пространства растет как  $2^N$ , где  $N$  — число спинов. Так как для вычисления средних значений наблюдаемых необходимо считать матричные элементы на состояниях всего гильбертова пространства, то вычисление оных может быть нереализуемо в силу конечной вычислительной мощности и вместимости памяти. Так, для  $N = 40$  спинов необходимо порядка 10 терабайтов памяти для полного задания состояния спиновой системы. Для моделирования состояния подобного ансамбля спинов необходимо характеризовать данную систему меньшим, но

достаточным числом параметров.

Наиболее известны два метода решения таких задач: стохастический метод (например, квантовый метод Монте-Карло [3]) и метод эффективного сжатия гильбертова пространства системы спинов (например, ренормгруппа матрицы плотности [4]). В недавней статье [5] был рассмотрен новый метод решения данной задачи, являющийся эффективным объединением двух предыдущих методов. Суть данного метода состоит в аппроксимации вектора основного состояния системы спинов *искусственной нейронной сетью*, которая представляет собой многомерную функцию, являющуюся суперпозицией двух одномерных функций с некоторым количеством неопределенных параметров. Математические основы достижения данным методом удовлетворительной аппроксимации опираются на известную математическую теорему Колмогорова-Арнольда о представлении произвольной многомерной функции суперпозицией одномерных функций [6]. Следуя этой теореме, можно подобрать такую непрерывную функцию, что с ее помощью можно будет с произвольной точностью выразить вектор основного состояния. Поэтому, подобрав достаточно хорошую функцию, коей мы выбираем согласно [5] *ограниченную машину Больцмана*, можно будет с ее помощью решить данную задачу. Этот метод и будет рассмотрен в данной работе.

В качестве критерия точности аппроксимации ограниченной машиной Больцмана вектора основного состояния будет выступать энергия основного состояния системы спинов:

$$E[\psi] = \frac{\langle \psi | \hat{\mathcal{H}} | \psi \rangle}{\langle \psi | \psi \rangle}$$

Если известно точное значение энергии основного состояния, то его можно сравнить с полученным по ходу *обучения* искусственной нейронной сети. Обучение искусственной нейронной сети представляет собой процесс подбора таких параметров, которые бы позволяли ей решать поставленную задачу. Если же точное значение энергии не известно, то точность аппроксимации можно проверить, проверяя устойчивость вычисляемого значения энергии при использовании различных методов обучения ограниченной машины Больцмана.

# Глава 1. Теоретическая часть

В данной главе рассматриваются основные сведения об ограниченной машине Больцмана, а также методах ее обучения и используемых приближениях при решении квантовой задачи многих тел. Также вводятся алгоритмы оптимизации, используемые в практической части для обучения ограниченной машины Больцмана.

## 1.1 Квантовые состояния нейронной сети

Как было сказано во введении, в данной работе будет рассмотрен подход, предложенный в статье [5]. Представим вектор основного состояния спиновой системы следующим образом:

$$|\psi_0\rangle = \sum_{\mathbf{v}} \psi(\mathbf{v}, \mathbf{p}) |\mathbf{v}\rangle \quad (1.1)$$

Состояние будем характеризовать удвоенной проекцией спина на ось  $z$  или же вектором-столбцом *состояний стохастических нейронов видимого слоя*:

$$\mathbf{v} = \begin{pmatrix} \sigma_1^z \\ \sigma_2^z \\ \vdots \\ \sigma_N^z \end{pmatrix}$$

В качестве коэффициентов мы возьмем ограниченную машину Больцмана с тем лишь отличием, что вместо квадратов модулей коэффициентов разложения (1.1) будем рассматривать сами коэффициенты разложения. Тогда данная искусственная нейронная сеть будет представлять соответствующие квантовые состояния спиновой системы, а сам метод такого представления квантовых состояний называют *квантовыми состояниями нейронной сети* (Neural-network Quantum States или NQS). Теперь можно явно записать вид коэффициентов вектора основного состояния в нашем методе:

$$\psi(\mathbf{v}, \mathbf{p}) = \sum_{\mathbf{h}} e^{-\mathcal{E}(\mathbf{v}, \mathbf{h}, \mathbf{p})} \quad (1.2)$$

где суммирование происходит по всем возможным *состояниям стохастических нейронов скрытого слоя*:

$$\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{pmatrix}$$

$\mathcal{E}$  представляет собой *энергию ограниченной машины Больцмана*:

$$\mathcal{E}(\mathbf{v}, \mathbf{h}, \mathbf{p}) = -\mathbf{a}^T \mathbf{v} - \mathbf{b}^T \mathbf{h} - \mathbf{v}^T \mathbf{W} \mathbf{h} \quad (1.3)$$

а  $\mathbf{p}$  представляют собой параметры ограниченной машины Больцмана:

$$\mathbf{p} = \{\mathbf{a}, \mathbf{b}, \mathbf{W}\}$$

где:

$\mathbf{a}$  — вектор-столбец *видимых смещений* сети

$\mathbf{b}$  — вектор-столбец *скрытых смещений* сети

$\mathbf{W}$  — матрица *весов* сети

Таким образом, энергия системы становится многомерной функцией параметров ограниченной машины Больцмана:

$$E = E(\mathbf{p})$$

Метод квантовых состояний нейронной сети схематически изображен на рис. 1.1.

Из вариационного принципа Ритца следует, что энергия основного состояния  $E_0$  является минимумом функционала энергии системы спинов:

$$E \geq E_0$$

Следовательно, в терминах машинного обучения задачу по нахождению вектора основного состояния можно представить в виде задачи по нахождению таких параметров  $\mathbf{p}_0$  ограниченной машины Больцмана, при которой энергия системы становится минимальной. Энергия системы принимается за *целевую функцию*, минимум которой требуется найти:

$$\mathbf{p}_0 = \arg \min_{\mathbf{p}} E(\mathbf{p}) \quad (1.4)$$

Из теоремы Колмогорова-Арнольда не известно, какое количество параметров надо использовать для точной аппроксимации многомерной функции. От числа скрытых стохастических нейронов зависит точность аппроксимации основного вектора состояния системы спинов. Принято говорить не о

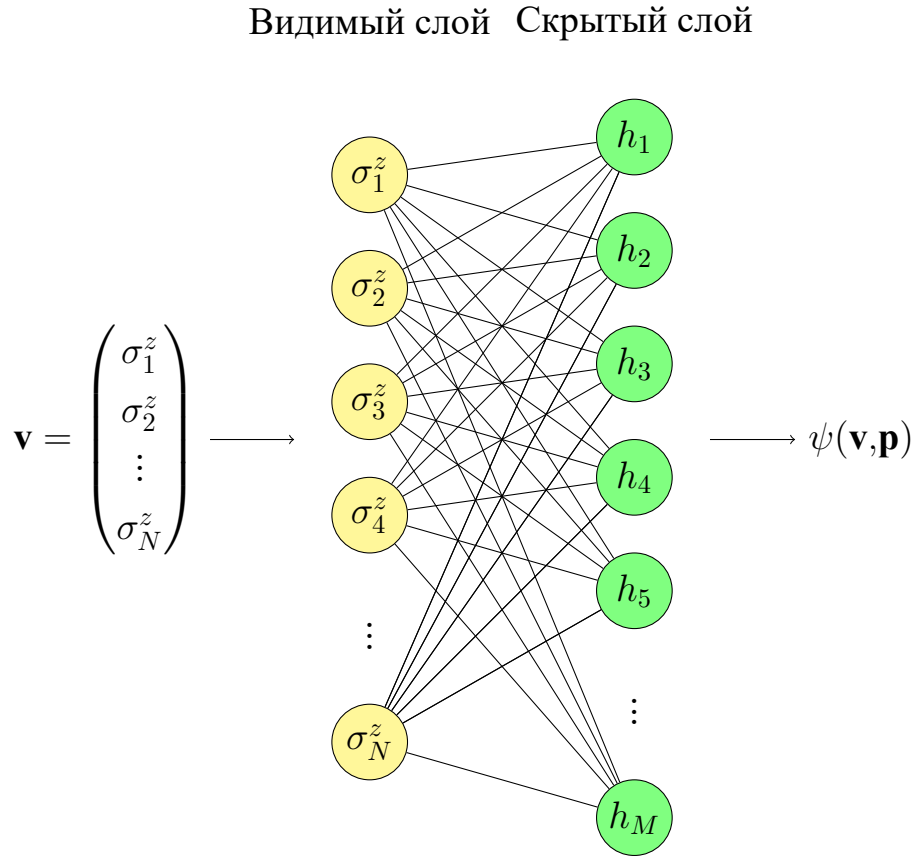


Рис. 1.1. Схематическое изображение метода квантовых состояний нейронной сети. Ограниченная машина Больцмана, изображенная на рисунке, представляет собой волновую функцию системы спинов. По данному на вход квантовому состоянию системы спинов искусственная нейронная сеть будет возвращать значение волновой функции системы в данном состоянии.

числе скрытых нейронов, а об их *плотности*:

$$\rho = \frac{M}{N}$$

В данной работе используется  $\rho = 2$ , что достаточно для получения допустимой точности для классических гамильтонианов, рассмотренных в дальнейшем.

## 1.2 Вариационный метод Монте-Карло

Как было упомянуто во введении, рассматривать полное гильбертово пространство системы достаточно большого числа спинов не представляется возможным с вычислительной точки зрения. Тогда имеет смысл ограничиться частью гильбертова пространства для решения уравнения (1.4). В качестве метода решения применим *вариационный метод Монте-Карло* (Variational Monte-Carlo или VMC) [7]. Для начала перепишем, как вычисляется значение энергии в произвольном состоянии, следующим образом:

$$\begin{aligned} E[\psi] &= \frac{\langle \psi | \hat{\mathcal{H}} | \psi \rangle}{\langle \psi | \psi \rangle} = \sum_{\mathbf{v}, \mathbf{v}'} \frac{\langle \psi | \mathbf{v} \rangle \langle \mathbf{v} | \hat{\mathcal{H}} | \mathbf{v}' \rangle \langle \mathbf{v}' | \psi \rangle}{\langle \psi | \psi \rangle} = \\ &= \sum_{\mathbf{v}} \frac{|\psi(\mathbf{v})|^2}{\langle \psi | \psi \rangle} \sum_{\mathbf{v}'} \mathcal{H}_{\mathbf{v}, \mathbf{v}'} \frac{\psi(\mathbf{v}')}{\psi(\mathbf{v})} = \mathbb{E}[E_{\text{лок}}] \quad (1.5) \end{aligned}$$

где:

$$E_{\text{лок}}(\mathbf{v}) = \sum_{\mathbf{v}'} \mathcal{H}_{\mathbf{v}, \mathbf{v}'} \frac{\psi(\mathbf{v}')}{\psi(\mathbf{v})} \quad (1.6)$$

Вариационный метод Монте-Карло состоит в приближенном вычислении энергии основного состояния посредством генерации выборки Монте-Карло. Выборка Монте-Карло должна быть такой, чтобы несмотря на потерю множества членов в вычислении математического ожидания  $\mathbb{E}[E_{\text{лок}}]$  в (1.5), можно было с достаточной точностью вычислять значение энергии системы спинов. Обозначив выборку Монте-Карло как:

$$\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}$$

вариационный метод Монте-Карло будет заключаться в следующем:

$$E = \mathbb{E}[E_{\text{лок}}] \approx \langle E_{\text{лок}} \rangle_{\mathbf{v} \in \mathbf{V}} \quad (1.7)$$

В качестве алгоритма, по которому будем генерировать выборку Монте-Карло, будем использовать *алгоритм Метрополиса-Гастингса*, суть которого продемонстрирована на рис. 1.2.

Дополнительно был использован алгоритм *термализации*, когда на каждой итерации в обучении NQS из выбранного состояния производилось определенное количество шагов, дабы сделать первое состояние случайным. В данной работе производилось  $1000N$  дополнительных шагов перед началом генерации выборки Монте-Карло. Также производилось дополнительно  $N$  шагов для каждого элемента выборки, дабы сделать два состояния в выборке более независимыми. В данной работе размер выборки составлял 10000.

В итоге получилось, что на создание выборки Монте-Карло необходимо  $11000N$  шагов. Данную выборку необходимо создавать на каждой итерации обучения. Данная часть была написана на языке программирования C++, для

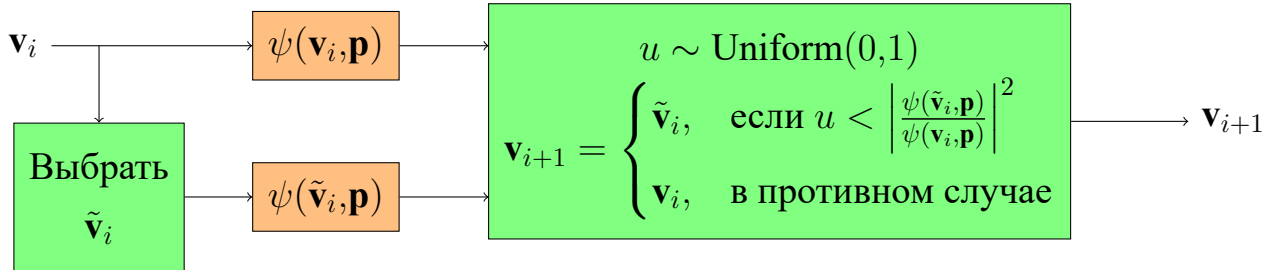


Рис. 1.2. Схематическое изображение одного шага в алгоритме Метрополиса-Гастингса. Суть этого алгоритма состоит в вычленении из (1.5) членов, вносящих наибольший вклад. В качестве меры вклада берется квадрат модуля волновой функции. Из случайно выбранного состояния  $\mathbf{v}_i$  выбираем новое состояние  $\tilde{\mathbf{v}}_i$ , посредством переворота одного или нескольких спинов. Если квадрат модуля нового состояния больше исходного, то определенно заносим данное состояние в выборку Монте-Карло. Если же квадрат модуля нового состояния меньше исходного, то заносим данное состояние в выборку Монте-Карло лишь с определенной вероятностью, равной отношению квадратов модулей нового и исходного состояния.



наибыстрейшей генерации.

В данной работе также используется *блочный анализ* (binning analysis) с целью проверки ошибки вычисления значения энергии, изложенный в статье [8]. Вычисление ошибки использовалась для проверки надежности метода оценки величины энергии системы спинов. Методы оптимизации, изложенные ниже, допускают осцилляцию энергии по ходу обучения. В ходе обучения эта осцилляция вносила несколько бóльший вклад, нежели ошибка в вычислении энергии, поэтому последняя величина в дальнейшем опускается.

### 1.3 Используемые алгоритмы оптимизации

Теперь необходимо определиться с методом оптимизации параметров ограниченной машины Больцмана. Самым распространенным методом минимизации целевой функции в машинном обучении является метод *градиентного спуска* (Gradient Descent или GD). Для демонстрации метода рассмотрим одномерный случай:

$$p_t = p_{t-1} - \alpha \frac{dE}{dp}$$

где:

$t$  — шаг обучения

$\alpha$  — скорость обучения

Если изменять параметр таким образом, то целевая функция будет уменьшаться:

$$E(p_t) \approx E(p_{t-1}) - \alpha \left( \frac{dE}{dp} \right)^2 < E(p_{t-1})$$

Таким образом достижение минимума целевой функции можно осуществить, меняя параметры следующим образом:

$$\mathbf{p}_t = \mathbf{p}_{t-1} - \alpha \nabla_{\mathbf{p}} E(\mathbf{p}_{t-1}) \quad (1.8)$$

Но данный метод останавливается в локальных минимумах, так как там градиент становится равным нулю. К тому же в некоторых случаях целевую

функцию вычислить точно не удастся, тем более ее градиент. Поэтому появилось множество модификаций метода градиентного спуска. Самым простым является так называемый метод *стохастического градиентного спуска* (Stochastic Gradient Descent или SGD) [9]. Суть его состоит в том, что целевая функция изменяется таким образом, чтобы представить ее в виде (1.7), и брать градиент от новой целевой функции.

Вычислим значение градиента от исходной целевой функции. Для этого введем:

$$\mathbf{D}_p(\mathbf{v}) = \frac{\nabla_p \psi(\mathbf{v}, \mathbf{p})}{\psi(\mathbf{v}, \mathbf{p})}$$

Тогда:

$$\begin{aligned} \nabla_p E(\mathbf{p}) &= \nabla_p \frac{\langle \psi | \hat{\mathcal{H}} | \psi \rangle}{\langle \psi | \psi \rangle} = \nabla_p \sum_{\mathbf{v}, \mathbf{v}'} \frac{\psi^*(\mathbf{v}, \mathbf{p}) \mathcal{H}_{\mathbf{v}, \mathbf{v}'} \psi(\mathbf{v}', \mathbf{p})}{\langle \psi | \psi \rangle} = \\ &= \sum_{\mathbf{v}, \mathbf{v}'} \frac{\mathbf{D}_p^*(\mathbf{v}) \psi^*(\mathbf{v}, \mathbf{p}) \mathcal{H}_{\mathbf{v}, \mathbf{v}'} \psi(\mathbf{v}', \mathbf{p})}{\langle \psi | \psi \rangle} + \sum_{\mathbf{v}, \mathbf{v}'} \frac{\psi^*(\mathbf{v}, \mathbf{p}) \mathcal{H}_{\mathbf{v}, \mathbf{v}'} \mathbf{D}_p(\mathbf{v}') \psi(\mathbf{v}', \mathbf{p})}{\langle \psi | \psi \rangle} + \\ &+ \sum_{\mathbf{v}, \mathbf{v}'} \frac{\psi^*(\mathbf{v}, \mathbf{p}) \mathcal{H}_{\mathbf{v}, \mathbf{v}'} \psi(\mathbf{v}', \mathbf{p})}{\langle \psi | \psi \rangle} \sum_{\mathbf{v}} \frac{|\psi(\mathbf{v}, \mathbf{p})|^2 (\mathbf{D}_p^*(\mathbf{v}) + \mathbf{D}_p(\mathbf{v}))}{\langle \psi | \psi \rangle} = \\ &= \sum_{\mathbf{v}} \frac{|\psi(\mathbf{v}, \mathbf{p})|^2}{\langle \psi | \psi \rangle} \mathbf{D}_p^*(\mathbf{v}) \sum_{\mathbf{v}'} \mathcal{H}_{\mathbf{v}, \mathbf{v}'} \frac{\psi(\mathbf{v}', \mathbf{p})}{\psi(\mathbf{v}, \mathbf{p})} + \\ &+ \sum_{\mathbf{v}'} \frac{|\psi(\mathbf{v}', \mathbf{p})|^2}{\langle \psi | \psi \rangle} \mathbf{D}_p(\mathbf{v}') \sum_{\mathbf{v}} \mathcal{H}_{\mathbf{v}', \mathbf{v}}^* \frac{\psi^*(\mathbf{v}, \mathbf{p})}{\psi^*(\mathbf{v}', \mathbf{p})} + 2E \operatorname{Re}(\mathbb{E}[\mathbf{D}_p]) = \\ &= 2 \operatorname{Re}(\mathbb{E}[E_{\text{лок}} \mathbf{D}_p]) + 2E \operatorname{Re}(\mathbb{E}[\mathbf{D}_p]) \end{aligned}$$

Таким образом для градиента получаем:

$$\nabla_p E(\mathbf{p}) = 2 \operatorname{Re}(\mathbb{E}[E_{\text{лок}} \mathbf{D}_p] + E \mathbb{E}[\mathbf{D}_p]) \quad (1.9)$$

А для стохастического градиента:

$$\nabla_p E(\mathbf{V}, \mathbf{p}) = 2 \operatorname{Re} \left( \left\langle (E_{\text{лок}} + \langle E_{\text{лок}} \rangle_{\mathbf{v} \in \mathbf{V}}) \mathbf{D}_p \right\rangle_{\mathbf{v} \in \mathbf{V}} \right) \quad (1.10)$$

Так как в некоторых случаях размерность гильбертова пространства будет достаточно большой, то рассматривать будем именно стохастический градиентный спуск и его модификации.

Важной модификацией является использование так называемого метода *стохастических реконфигураций* [10]. Суть его состоит в дополнении стохастического градиента матрицей стохастической реконфигурации:

$$\mathbf{S} = \langle \mathbf{D}_{\mathbf{p}}^* \otimes \mathbf{D}_{\mathbf{p}}^T \rangle_{\mathbf{v} \in \mathbf{V}} - \langle \mathbf{D}_{\mathbf{p}}^* \rangle_{\mathbf{v} \in \mathbf{V}} \otimes \langle \mathbf{D}_{\mathbf{p}}^T \rangle_{\mathbf{v} \in \mathbf{V}} \quad (1.11)$$

следующим образом:

$$\mathbf{p}_t = \mathbf{p}_{t-1} - \alpha \mathbf{S}^{-1} \nabla_{\mathbf{p}} E(\mathbf{V}, \mathbf{p}_{t-1}) \quad (1.12)$$

Если матрица стохастической реконфигурации не может быть обратима, то под  $\mathbf{S}^{-1}$  может подразумеваться псевдообращение Мура-Пенроуза. В ходе данной работы было замечено, что псевдообращение работает не эффективно. Так, помимо увеличения времени обучения, осцилляции энергии по ходу обучения значительно увеличиваются. Поэтому в данной работе используется явная регуляризация, описанная в статье [10]:

$$\mathbf{S}^{\text{reg}} = \mathbf{S} + \epsilon(t) \mathbf{I} \quad (1.13)$$

где согласно статье [5] вид  $\epsilon(t)$  выбран следующим образом:

$$\epsilon(t) = \max(\epsilon_0 b^t, \epsilon_{\min}) \quad (1.14)$$

Значение скорости обучения  $\alpha$  влияет на сходимость метода и поэтому является одним из главных параметров. Для каждого исследуемого гамильтониана значение  $\alpha$  принималось за 0,01 и менялось, в зависимости от результатов обучения для получения устойчивого значения энергии. В общем случае значение  $\alpha$  зависит от шага обучения и, согласно [9], должно удовлетворять следующим требованиям:

$$\sum_t \alpha_t^2 < \sum_t \alpha_t = \infty$$

В данной же работе  $\alpha$  принимается постоянной, дабы не попасть в локальный минимум. Конечно, это может привести к известной проблеме *переобучения*, когда некоторые параметры искусственной нейронной сети неограниченно растут. Благо есть множество методов борьбы с данным явлением, одним из которых является метод  $L_2$  регуляризации.

Метод  $L_2$  регуляризации используется в данной работе для борьбы с переобучением NQS. Суть этого метода состоит в добавлении к целевой функции добавки, которая характеризует размах весов искусственной нейронной сети. Уменьшая таким образом значения весов, можно добиться устойчивого обучения сети. Таким образом, мы будем минимизировать следующую функцию:

$$L = E + \frac{\lambda}{2} \|\mathbf{p}\|^2 \quad (1.15)$$

Тогда новый градиент будет выглядеть следующим образом:

$$\nabla_{\mathbf{p}} L = \nabla_{\mathbf{p}} E + \lambda \mathbf{p} \quad (1.16)$$

В ходе работы значение параметра  $\lambda$ , согласно [11], было принято за 0,0001. В ходе работы было замечено, что увеличение данного параметра, при постоянном значении скорости обучения, ведет к увеличению осцилляции значения энергии по ходу обучения. Поэтому, данный параметр не меняет свое значение в данной работе.

Таким образом, одним из алгоритмов минимизации, использованных в данной работе, является SGD:

---

**SGD:** стохастический градиентный спуск с методом стохастической реконфигурации и с  $L_2$  регуляризацией

---

- 1: **задать** значение параметра скорости обучения  $\alpha \in \mathbb{R}$ , значения параметров регуляризации  $\epsilon_0 \in \mathbb{R}$ ,  $b \in \mathbb{R}$ ,  $\epsilon_{\min} \in \mathbb{R}$ ,  $L_2$  регуляризационный множитель  $\lambda = 10^{-4}$
  - 2: **инициализировать** начальный шаг итерации  $t \leftarrow 0$ , случайным образом вектор параметров  $\mathbf{p}_{t=0} \in \mathbb{C}^n$
  - 3: **повторять**
  - 4:      $t \leftarrow t + 1$
  - 5:      $\nabla_{\mathbf{p}} E(\mathbf{V}, \mathbf{p}_{t-1}) \leftarrow$  Выборка Монте-Карло
  - 6:      $\mathbf{g}_t \leftarrow \mathbf{S}^{-1}(\nabla_{\mathbf{p}} E(\mathbf{V}, \mathbf{p}_{t-1}) + \lambda \mathbf{p}_{t-1})$
  - 7:      $\mathbf{p}_t \leftarrow \mathbf{p}_{t-1} - \alpha \mathbf{g}_t$
  - 8: **до тех пор, пока не пройдет достаточное число итераций**
- 

Естественно, для проверки надежности подхода, одним методом минимизации целевой функции ограничиться нельзя. Поэтому мы рассмотрим одну

из простейших, но часто используемых модификаций стохастического градиентного спуска, под названием *стохастический градиентный спуск с моментом* (Momentum) [12]. Суть данного метода состоит в том, чтобы избежать попадания искусственной нейронной сети по ходу обучения в локальный минимум, накапливая градиент по ходу обучения. Тем самым, попав в локальный минимум, у сети будет шанс выскочить из нее, обладая большим накопленным градиентом. Правило изменения параметров NQS будет выглядеть следующим образом:

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + \alpha \nabla_{\mathbf{p}} L(\mathbf{p}_{t-1})$$

$$\mathbf{p}_t = \mathbf{p}_{t-1} - \mathbf{m}_t$$

где:

$\beta_1$  — *множитель первого момента*

$\mathbf{m}$  — *вектор первого момента*

Для ограниченной машины Больцмана типичным значением  $\beta_1$  будет 0,9 [11]. В качестве начального значения вектора первого момента принимается нуль-вектор. Таким образом, алгоритм оптимизации Momentum будет выгля-

деть следующим образом:

---

**Momentum:** стохастический градиентный спуск с моментом и с  $L_2$  регуляризацией

---

- 1: **задать** значение параметра скорости обучения  $\alpha \in \mathbb{R}$ , множителя первого момента  $\beta_1 = 0,9$  и  $L_2$  регуляризационного множителя  $\lambda = 10^{-4}$
  - 2: **инициализировать** начальный шаг итерации  $t \leftarrow 0$ , случайным образом вектор параметров  $\mathbf{p}_{t=0} \in \mathbb{C}^n$  и значение вектора первого момента  $\mathbf{m}_{t=0} \leftarrow \mathbf{0}$
  - 3: **повторять**
  - 4:      $t \leftarrow t + 1$
  - 5:      $\nabla_{\mathbf{p}} E(\mathbf{V}, \mathbf{p}_{t-1}) \leftarrow$  Выборка Монте-Карло
  - 6:      $\mathbf{g}_t \leftarrow \nabla_{\mathbf{p}} E(\mathbf{V}, \mathbf{p}_{t-1}) + \lambda \mathbf{p}_{t-1}$
  - 7:      $\mathbf{m}_t \leftarrow \beta_1 \mathbf{m}_{t-1} + \alpha \mathbf{g}_t$
  - 8:      $\mathbf{p}_t \leftarrow \mathbf{p}_{t-1} - \mathbf{m}_t$
  - 9: **до тех пор, пока не пройдет достаточное число итераций**
- 

Так как мы изначально задались целью избежать застревания в локальном минимуме фиксированием скорости обучения, необходимо также рассмотреть алгоритм минимизации, который бы по ходу обучения сам выбирал ее скорость. Недавно был представлен метод *адаптивной оценки момента* (Adaptive Moment Estimation или Adam) [13]. Он явился синтезом других методов адаптивной скорости обучения AdaGrad и RMSProp [14, 15]. Суть метода состоит, помимо хранения значений предыдущих градиентов, в хранении квадратов градиентов (то есть поэлементное произведение коэффициентов градиента). Теперь правило изменения параметров NQS будет выглядеть следующим образом:

$$\begin{aligned}\mathbf{m}_t &= \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \nabla_{\mathbf{p}} L(\mathbf{p}_{t-1}) \\ \mathbf{v}_t &= \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \nabla_{\mathbf{p}} L(\mathbf{p}_{t-1})^2\end{aligned}$$

$$\begin{aligned}\tilde{\mathbf{m}}_t &= \mathbf{m}_t / (1 - \beta_1^t) \\ \tilde{\mathbf{v}}_t &= \mathbf{v}_t / (1 - \beta_2^t) \\ \mathbf{p}_t &= \mathbf{p}_{t-1} - \alpha \tilde{\mathbf{m}}_t / (\sqrt{\tilde{\mathbf{v}}_t} + \epsilon)\end{aligned}$$

где:

$\beta_2$  — множитель второго момента

$\mathbf{v}$  — вектор второго момента

$\epsilon$  — малый параметр, делающий ненулевым значение в знаменателе

В качестве значения множителя второго момента и малого параметра принимаются значения, рекомендуемые в статье [13]. В качестве начального значения вектора второго момента также принимается нуль-вектор. Таким образом, алгоритм оптимизации Adam будет выглядеть следующим образом:

---

**Adam:** адаптивная оценка момента с  $L_2$  регуляризацией

---

- 1: **здать** значение параметра скорости обучения  $\alpha \in \mathbb{R}$ , множителя первого момента  $\beta_1 = 0,9$ , множителя второго момента  $\beta_2 = 0,999$ , параметра возмущения  $\epsilon$ ,  $L_2$  регуляризационного множителя  $\lambda = 10^{-4}$
  - 2: **инициализировать** начальный шаг итерации  $t \leftarrow 0$ , случайным образом вектор параметров  $\mathbf{p}_{t=0} \in \mathbb{C}^n$ , значение вектора первого момента  $\mathbf{m}_{t=0} \leftarrow \mathbf{0}$  и вектора второго момента  $\mathbf{v}_{t=0} \leftarrow \mathbf{0}$
  - 3: **повторять**
  - 4:      $t \leftarrow t + 1$
  - 5:      $\nabla_{\mathbf{p}} E(\mathbf{V}, \mathbf{p}_{t-1}) \leftarrow$  Выборка Монте-Карло
  - 6:      $\mathbf{g}_t \leftarrow \nabla_{\mathbf{p}} E(\mathbf{V}, \mathbf{p}_{t-1}) + \lambda \mathbf{p}_{t-1}$
  - 7:      $\mathbf{m}_t \leftarrow \beta_1 \mathbf{m}_t + (1 - \beta_1) \mathbf{g}_t$
  - 8:      $\mathbf{v}_t \leftarrow \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \mathbf{g}_t^2$
  - 9:      $\tilde{\mathbf{m}}_t \leftarrow \mathbf{m}_t / (1 - \beta_1^t)$
  - 10:     $\tilde{\mathbf{v}}_t \leftarrow \mathbf{v}_t / (1 - \beta_2^t)$
  - 11:     $\mathbf{p}_t \leftarrow \mathbf{p}_{t-1} - \alpha \tilde{\mathbf{m}}_t / (\sqrt{\tilde{\mathbf{v}}_t} + \epsilon)$
  - 12: **до тех пор, пока не пройдет достаточное число итераций**
-

# Глава 2. Практическая часть

В данной главе рассматривается практическое приложение метода квантовых состояний нейронной сети к нахождению основного состояния ансамбля спинов, характеризуемых некоторыми классическими гамильтонианами. Приводятся результаты обучения искусственной нейронной сети и их анализ.

## 2.1 Исследуемые гамильтонианы

В данной работе будут рассмотрены два классических гамильтониана с известными значениями энергии основного состояния цепочки  $N = 20$  спинов и изотропная часть гамильтониана магнитного кластера  $V_{15}$ .

Первым классическим гамильтонианом является модель Изинга поперечного поля:

$$\hat{\mathcal{H}}_{\text{TFI}} = -h \sum_i \hat{\sigma}_i^x - J_z \sum_{\langle i,j \rangle} \hat{\sigma}_i^z \hat{\sigma}_j^z \quad (2.1)$$

где в данной работе  $h = J_z = 1$ .

Вторым классическим гамильтонианом является модель антиферромагнетика Гейзенберга:

$$\hat{\mathcal{H}}_{\text{AFH}} = \sum_{\langle i,j \rangle} \hat{\sigma}_i^x \hat{\sigma}_j^x + \hat{\sigma}_i^y \hat{\sigma}_j^y + J_z \hat{\sigma}_i^z \hat{\sigma}_j^z \quad (2.2)$$

где в данной работе  $J_z = 1$ .

Изотропная часть гамильтониана магнитного кластера  $V_{15}$  [16]:

$$\hat{\mathcal{H}}_{V_{15}} = \hat{\mathcal{H}}_0 + \hat{\mathcal{H}}_1 \quad (2.3)$$

$$\hat{\mathcal{H}}_0 = \frac{1}{4} J (\hat{\sigma}_1 \cdot \hat{\sigma}_2 + \hat{\sigma}_3 \cdot \hat{\sigma}_4 + \hat{\sigma}_5 \cdot \hat{\sigma}_6 + \hat{\sigma}_{10} \cdot \hat{\sigma}_{11} + \hat{\sigma}_{12} \cdot \hat{\sigma}_{13} + \hat{\sigma}_{14} \cdot \hat{\sigma}_{15})$$



$$\begin{aligned}
\hat{\mathcal{H}}_1 = & \frac{1}{4}J'(\hat{\sigma}_2 \cdot \hat{\sigma}_3 + \hat{\sigma}_4 \cdot \hat{\sigma}_5 + \hat{\sigma}_1 \cdot \hat{\sigma}_6 + \hat{\sigma}_{11} \cdot \hat{\sigma}_{12} + \hat{\sigma}_{13} \cdot \hat{\sigma}_{14} + \hat{\sigma}_{10} \cdot \hat{\sigma}_{15}) + \\
& + \frac{1}{4}J''(\hat{\sigma}_1 \cdot \hat{\sigma}_3 + \hat{\sigma}_3 \cdot \hat{\sigma}_5 + \hat{\sigma}_1 \cdot \hat{\sigma}_5 + \hat{\sigma}_{10} \cdot \hat{\sigma}_{12} + \hat{\sigma}_{12} \cdot \hat{\sigma}_{14} + \hat{\sigma}_{10} \cdot \hat{\sigma}_{14}) + \\
& + \frac{1}{4}J_1(\hat{\sigma}_2 \cdot \hat{\sigma}_7 + \hat{\sigma}_7 \cdot \hat{\sigma}_{11} + \hat{\sigma}_4 \cdot \hat{\sigma}_8 + \hat{\sigma}_8 \cdot \hat{\sigma}_{15} + \hat{\sigma}_6 \cdot \hat{\sigma}_9 + \hat{\sigma}_9 \cdot \hat{\sigma}_{13}) + \\
& + \frac{1}{4}J_2(\hat{\sigma}_1 \cdot \hat{\sigma}_7 + \hat{\sigma}_7 \cdot \hat{\sigma}_{10} + \hat{\sigma}_3 \cdot \hat{\sigma}_8 + \hat{\sigma}_8 \cdot \hat{\sigma}_{14} + \hat{\sigma}_5 \cdot \hat{\sigma}_9 + \hat{\sigma}_9 \cdot \hat{\sigma}_{12})
\end{aligned}$$

где в данной работе используются следующие параметры:

$$J = 800 \text{ K}, J' = J_1 = 350 \text{ K}, J'' = J_2 = 225 \text{ K}$$

## 2.2 Результаты обучения нейронной сети

Результаты обучения ограниченной машины Больцмана на модели Изинга поперечного поля представлены на рис. 2.1. Здесь для SGD в качестве параметров использовались:

$$\alpha = 10^{-2}, \epsilon_0 = 100, b = 0,9, \epsilon_{\min} = 10^{-4}$$

Как видно из рис. 2.2 SGD отлично подошел в качестве метода решения для данного гамильтониана, достигнув относительной погрешности менее

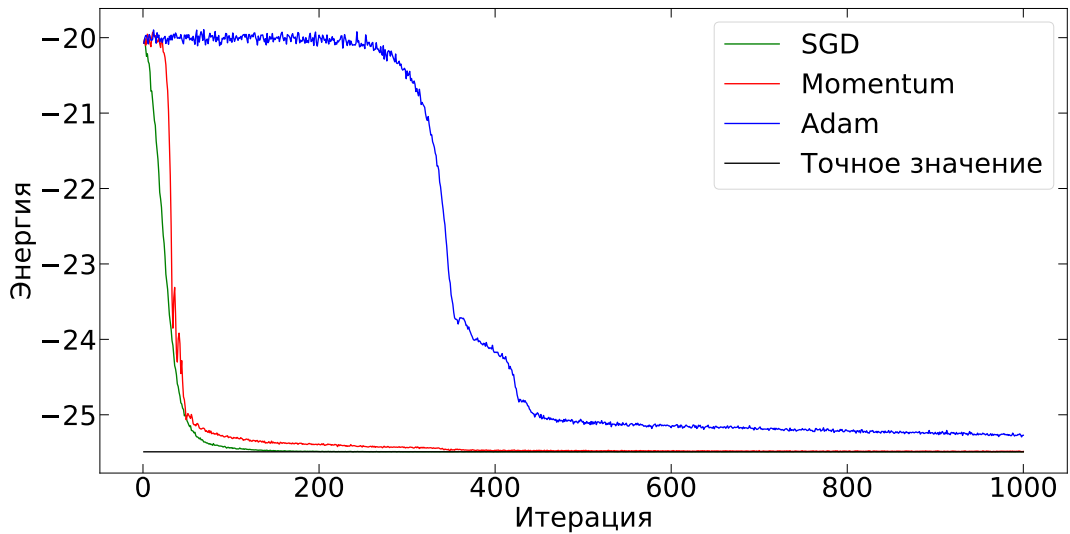


Рис. 2.1. Результат обучения на модели Изинга поперечного поля

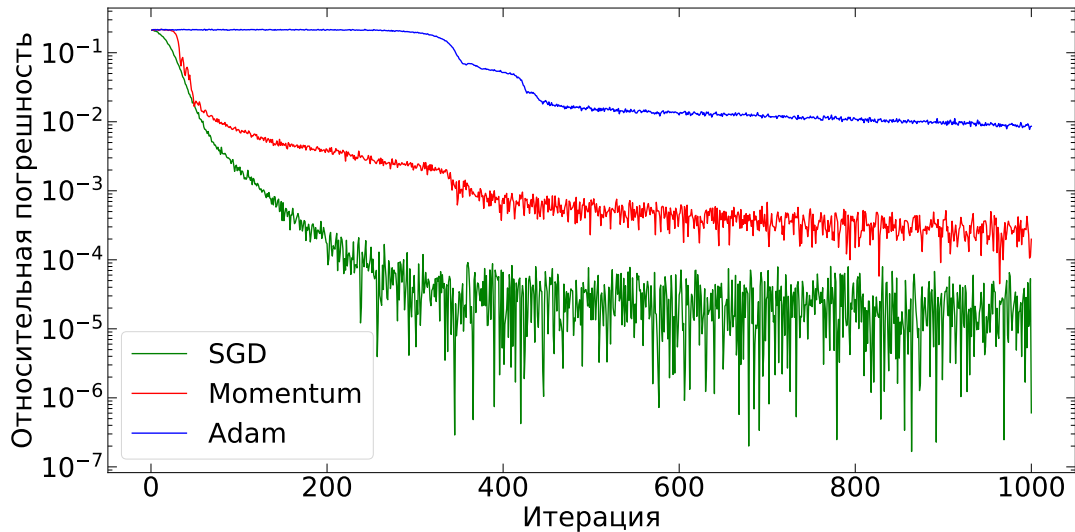


Рис. 2.2. Точность обучения на модели Изинга поперечного поля

0,01%. Модификации данного метода с такими же значениями скорости обучения не смогли похвастаться такой же точностью. И если Momentum еще показывает убедительную сходимость, то Adam смог достигнуть только 1% относительной погрешности. Здесь стоит оговориться, что, как выяснилось в дальнейшем, это результат удачного подбора регуляризационных параметров при вычислении стохастической матрицы реконфигурации. При этом Adam показывает свой характер сходимости, а именно длительное накапливание градиента, и затем стремительное повышение точности. При этом Adam более устойчив в плане параметров обучения. Таким образом, довольно медленное и устойчивое продвижение к минимуму и меньшая зависимость от выбранных параметров выгодно отличает его от других более чувствительных методов.

Результаты обучения ограниченной машины Больцмана на модели антиферромагнетика Гейзенберга представлены на рис. 2.3. Здесь для SGD в качестве параметров использовались те же регуляризационные параметры, а значение скорости обучения составило:

$$\alpha = 10^{-3}$$

Как видно из рис. 2.4 SGD по-прежнему достиг относительной погрешности менее 0,01%. Модификации данного метода обладали следующими ско-

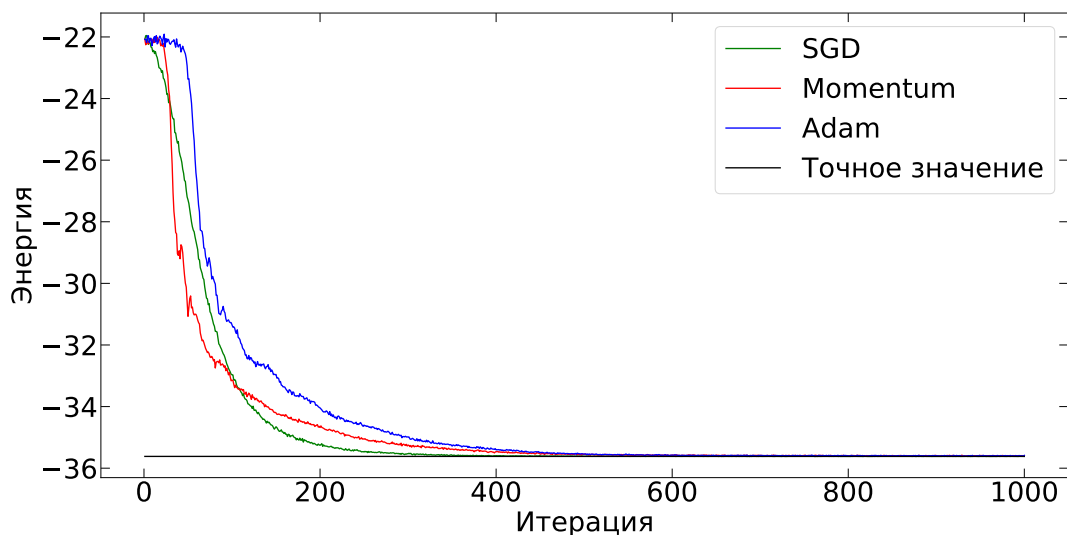


Рис. 2.3. Результат обучения на модели антиферромагнетика Гейзенберга

ростями обучения:

$$\text{Momentum} : \alpha = 5 \cdot 10^{-3}$$

$$\text{Adam} : \alpha = 0,1$$

Здесь становится заметно преимущество модификаций перед стандартным SGD. Так, Momentum и Adam уже не так сильно отстают. При этом учитывая, что скорость обучения фиксирована, становится понятно, поче-

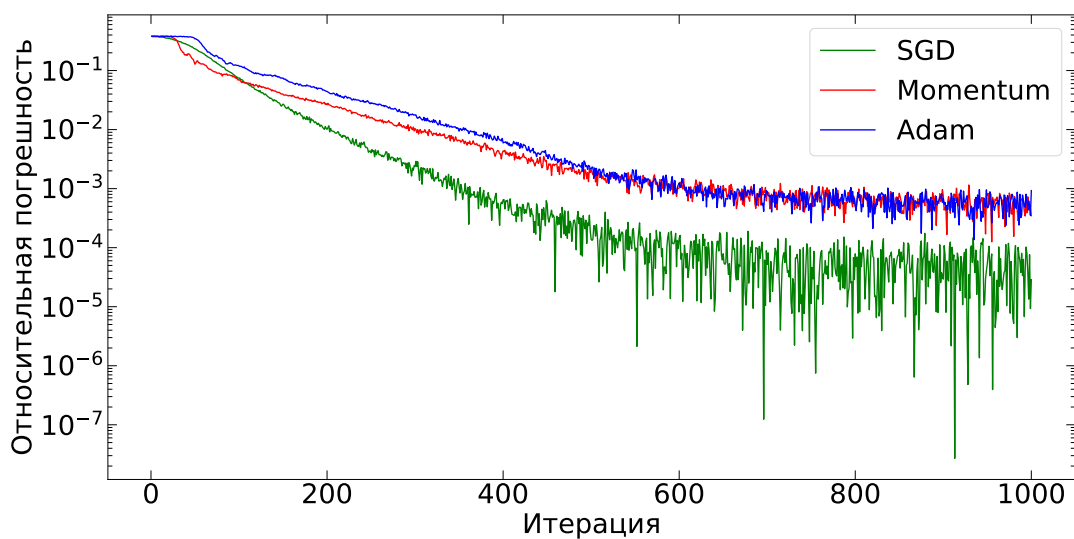


Рис. 2.4. Точность обучения на модели антиферромагнетика Гейзенберга

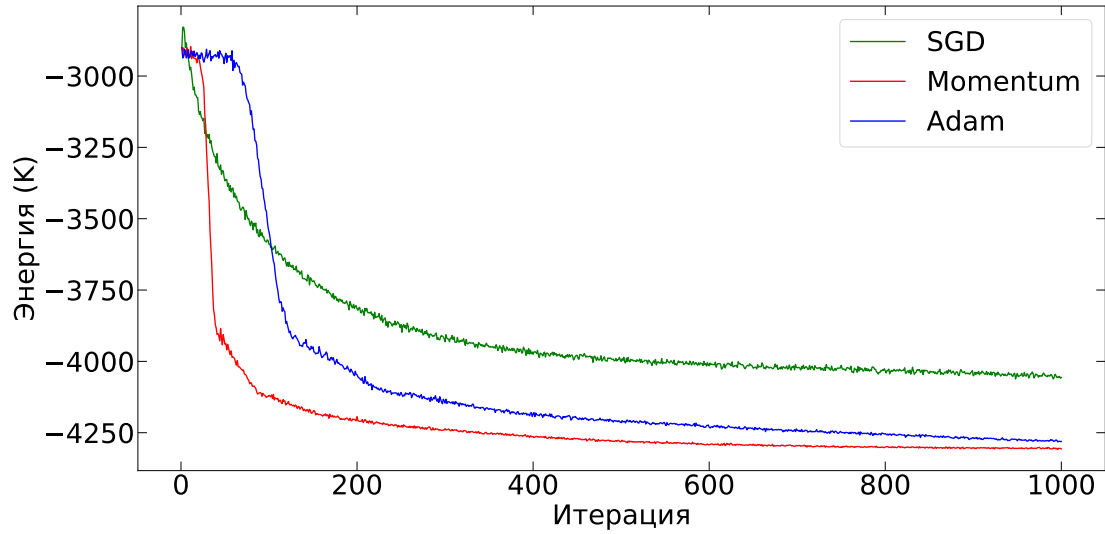


Рис. 2.5. Результат обучения на изотропной части гамильтониана магнитного кластера  $V_{15}$

му SGD не уступает лидерство. При уменьшении скорости обучения SGD бы попал в локальный минимум, в то время как Adam и Momentum, обладая большим запасом градиента, еще некоторое время двигались бы в поисках глобального минимума. Но в данной работе это не рассматривается, так как такой глубокий поиск невозможен без автоматического поиска подходящих параметров обучения. Пока же критерии поиска не выработаны, автоматизировать данную операцию не представляется выгодным по времени. Так что пока основным методом является исследование с постоянной скоростью обучения возможных кандидатов на глобальный минимум, обладающих достаточной устойчивостью при обучении, при рекомендуемых значениях параметров.

Результаты обучения ограниченной машины Больцмана на изотропной части гамильтониана магнитного кластера  $V_{15}$  представлены на рис. 2.5. Здесь для методов оптимизации в качестве параметров обучения использовались:

$$\text{SGD} : \alpha = 5 \cdot 10^{-5}, \epsilon_0 = 100, b = 0,999, \epsilon_{\min} = 20$$

$$\text{Momentum} : \alpha = 5 \cdot 10^{-5}$$

$$\text{Adam} : \alpha = 5 \cdot 10^{-4}$$

Именно для системы спинов с нетривиальными взаимодействиями про-

явилась проблема подбора параметров для простейших методов оптимизации. Старые параметры не подходят, так как с ними происходит срыв обучения. Поэтому правильность их подбора и определяет сходимость простейших методов. Данные параметры для SGD, как видно из результатов обучения, не являются оптимальными, так как обучение ограниченной машины Больцмана происходит медленно. Оптимальные параметры для нетривиальных взаимодействий уже намного труднее правильно подобрать, в то время как рекомендуемые для большинства «классических» целей параметры для модификаций метода уже дают вполне устойчивую сходимость и их изменение ее только улучшает. Таким образом, SGD с методом стохастической реконфигурации оказался очень чувствителен к усложнению картины взаимодействия системы спинов, в то время как Momentum и Adam даже при классических значениях параметров оказались менее чувствительны к данному усложнению.

# Заключение

В ходе данной работы был рассмотрен подход представления квантовых состояний системы спинов как состояний ограниченной машины Больцмана. В качестве проверки точности данного подхода была рассмотрена задача по нахождению вектора основного состояния для классических гамильтонианов модели Изинга поперечного поля и модели антиферромагнетика Гейзенберга, а также изотропной части гамильтониана магнитного кластера  $V_{15}$ , посредством вариационного метода Монте-Карло. В ходе работы в качестве методов решения поставленной задачи были использованы различные модификации метода градиентного спуска. В результате обучения ограниченной машины Больцмана получилось достичь относительной погрешности  $< 1\%$ , что указывает на наличие потенциала в данной области исследований.

Основной проблемой при решении поставленной задачи явилась собственная рассмотренным оптимизационным алгоритмам проблема подбора оптимальной скорости обучения для достижения точных результатов за относительно небольшое число шагов обучения. Данная проблема является краеугольной в машинном обучении, поэтому мы рассмотрели несколько алгоритмов ее решения. Именно за адаптивными методами обучения лежит дальнейшее развитие использования искусственных нейронных сетей в области физических исследований, так как для разных физических условий подбор параметров обучения будет затратным по времени вычислений занятием. Поэтому использование алгоритмов, самостоятельно выбирающих скорость обучения, будет уместным для различных физических задач. Естественно, для более точного результата требуется увеличить число параметров ограниченной машины Больцмана. Но тогда нужны и более надежные методы регуляризации для борьбы с переобучением искусственной нейронной сети.

Все эти результаты показывают, что необходимо в первую очередь выявить наиболее важные критерии сходимости искусственной нейронной сети в ходе обучения, дабы за несколько шагов определить диапазон параметров обучения для сходимости метода. Таким образом, необходимо использование дополнительной искусственной нейронной сети, для обучения на нахожде-

ние диапазона параметров обучения, приводящих к устойчивому результату. При этом основным критерием является быстрая обучаемость на небольшой выборке.

# Список литературы

- [1] *Carrasquilla J.* Machine learning phases of matter / J. Carrasquilla, R.G. Melko // *Neural Netw.* – 2017. – Vol. 13. – № 5. – P. 431–434.
- [2] *Butler K.T.* Machine learning for molecular and materials science / K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh // *Nature.* – 2018. – Vol. 559. – № 7715. – P. 547–555.
- [3] *Ceperley D.* Quantum monte carlo / D. Ceperley, B. Alder // *Science.* – 1986. – Vol. 231. – № 4738. – P. 555–560.
- [4] *White S.R.* Density matrix formulation for quantum renormalization groups / S.R. White // *Phys. Rev. Lett.* – 1992. – Vol. 69. – № 19. – P. 2863.
- [5] *Carleo G.* Solving the quantum many-body problem with artificial neural networks / G. Carleo, M. Troyer // *Science.* – 2017. – Vol. 355. – № 6325. – P. 602–606.
- [6] *Колмогоров А.Н.* О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения / А.Н. Колмогоров // *ДАН.* – 1957. – Т. 114. – № 5. – P. 953–956.
- [7] *McMillan W.L.* Ground state of liquid He 4 / W.L. McMillan // *Phys. Rev.* – 1965. – Vol. 138. – № 2A. – P. A442.
- [8] *Ambegaokar V.* Estimating errors reliably in Monte Carlo simulations of the Ehrenfest model / V. Ambegaokar, M. Troyer // *Am. J. Phys.* – 2010. – Vol. 78. – № 2. – P. 150–157.
- [9] *Harju A.* Stochastic gradient approximation: An efficient method to optimize many-body wave functions / A. Harju, B. Barbiellini, S. Siljamäki, R.M. Nieminen, G. Ortiz // *Phys. Rev. Lett.* – 1997. – Vol. 79. – № 7. – P. 1173.
- [10] *Sorella S.* Weak binding between two aromatic rings: Feeling the van der



- Waals attraction by quantum Monte Carlo methods / S. Sorella, M. Casula, D.Rocca // J. Chem. Phys. – 2007. – Vol. 127. – № 1. – P. 014105.
- [11] *Hinton G.E.* A practical guide to training restricted Boltzmann machines / G.E. Hinton // Neural networks: Tricks of the trade. – 2012. – P. 599–619.
- [12] *Qian N.* On the momentum term in gradient descent learning algorithms / N. Qian // Neural Netw. – 1999. – Vol. 12. – № 1. – P. 145–151.
- [13] Adam: A method for stochastic optimization – URL: <https://arxiv.org/abs/1412.6980v8> (дата обращения: 04.04.2020)
- [14] *Duchi J.* Adaptive subgradient methods for online learning and stochastic optimization / J. Duchi, E. Hazan, Y. Singer // J. Mach. Learn. Res. – 2011. – Vol. 12. – № Jul. – P. 2121–2159.
- [15] *Tieleman T.* Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude / T. Tieleman, G. Hinton // COURSERA: Neural networks for machine learning. – 2012. – Vol. 4. – № 2. – P. 26–31.
- [16] *Konstantinidis N.P.* Magnetic anisotropy in the molecular complex V 15 / N.P. Konstantinidis, D. Coffey // Phys. Rev. B. – 2002. – Vol. 66. – № 17. – P. 174426.