

Метод отжига

Лопатин А. С.

Санкт-Петербургский государственный университет¹

В работе рассмотрены реализации и свойства различных вариантов метода отжига.

1. Введение

Метод отжига (синонимы: метод обжига, метод симуляции отжига, метод модельной закалки, *simulated annealing*) — это техника оптимизации, использующая упорядоченный случайный поиск на основе аналогии с процессом образования в веществе кристаллической структуры с минимальной энергией при охлаждении.

История метода отжига начинается с 1953 года. В этом году Н. Метрополисом был разработан алгоритм симуляции установления равновесия в системе с множеством степеней свободы при заданной температуре [1]. В начале 80-х у С. Киркпатрика [2] впервые появилась идея использовать этот алгоритм не только для моделирования физических систем, но и для решения некоторых задач оптимизации.

В настоящее время метод отжига применяется для решения многих оптимизационных задач — финансовых (см. [3], [4]), компьютерной графики (см. [5], [6]), комбинаторных (см. [7]), в телекоммуникационных сетях (см. [8]), и многих других. Зачастую метод отжига используют для обучения нейронных сетей (см. [9], [10]).

Несмотря на такую широкую область применения, скорость сходимости метода отжига все еще мало изучена.

Огромным преимуществом метода отжига является свойство избежать “ловушки” в локальных минимумах оптимизируемой функции, и продолжить поиск глобального минимума. Это достигается за счет принятия не только изменений параметров, приводящих к уменьшению значения функции, но и некоторых изменений, увеличивающих ее значение, в зависимости от т. н. *температуры T* — характеристики моделируемого процесса. Чем выше температура, тем большие “ухудшающие” изменения (аналогичные случайным

¹©А. С. Лопатин, 2005

флуктуациям в нагретом веществе) допустимы, и больше их вероятность.

Еще одним преимуществом является то, что даже в условиях нехватки вычислительных ресурсов для нахождения глобального минимума, метод отжига, как правило, выдает весьма неплохое решение (один из локальных минимумов).

Л. Ингбером в [11] показано, что метод отжига и его модификации являются одним из наиболее эффективных методов случайного поиска оптимального решения для большого класса задач. В [12] Л. Ингбером проведен сравнительный анализ адаптивного метода отжига (Adaptive Simulated Annealing, ASA) и генетических алгоритмов, из которого следует, что на большинстве задач метод отжига не проигрывает генетическим алгоритмам, а на многих и выигрывает.

К настоящему времени разработано множество различных вариантов метода отжига, как общих, так и их специализаций для решения конкретных задач.

В данной работе анализируются известные на настоящий момент общие схемы отжига.

2. Алгоритм метода отжига

Метод отжига служит для поиска глобального минимума некоторой функции $f(x)$, заданной для x из некоторого пространства S , дискретного или непрерывного. Элементы множества S представляют собой состояния воображаемой физической системы (“энергетические уровни”), а значение функции f в этих точках используется как энергия системы $E = f(x)$. В каждый момент предполагается заданной *температура* системы T , как правило, уменьшающаяся с течением времени. После попадания в состояние x при температуре T , следующее состояние системы выбирается в соответствии с заданным *порождающим семейством вероятностных распределений* $\mathcal{G}(x, T)$, которое при фиксированных x и T задает случайный элемент $G(x, T)$ со значениями в пространстве S . После генерации нового состояния $x' = G(x, T)$, система с вероятностью $h(\Delta E, T)$ переходит к следующему шагу в состояние x' , в противном случае процесс генерации x' повторяется. Здесь ΔE обозначает приращение функции энергии $f(x') - f(x)$. Величина $h(\Delta E, T)$ называется *вероятностью принятия* нового состояния.

Как правило, в качестве функции $h(\Delta E, T)$ выбирается либо точное значение соответствующей физической величины

$$h(\Delta E, T) = \frac{1}{1 + \exp(\Delta E/T)},$$

либо приближенное значение

$$h(\Delta E, T) = \exp(-\Delta E/T).$$

Вторая формула используется наиболее часто. При ее использовании $h(\Delta E, T)$ оказывается больше единицы в случае $\Delta E < 0$, и тогда соответствующая вероятность считается равной 1. Таким

образом, если новое состояние дает лучшее значение оптимизируемой функции, то переход в это состояние произойдет в любом случае.

Итак, конкретная схема метода отжига задается следующими параметрами:

- Выбором закона изменения температуры $T(k)$, где k — номер шага.
- Выбором порождающего семейства распределений $\mathcal{G}(x, T)$.
- Выбором функции вероятности принятия $h(\Delta E, T)$.

Алгоритм:

- 1) Случайным образом выбирается начальная точка $x = x_0$, $x_0 \in S$. Текущее значение энергии E устанавливается в значение $f(x_0)$.
- 2) k -я итерация основного цикла состоит из следующих шагов:
 - (a) Сравнить энергию системы E в состоянии x с найденным на текущий момент глобальным минимумом. Если $E = f(x)$ меньше, то изменить значение глобального минимума.
 - (b) Сгенерировать новую точку $x' = G(x, T(k))$.
 - (c) Вычислить значение функции в ней $E' = f(x')$.
 - (d) Сгенерировать случайное число α из интервала $[0; 1]$

- (e) Если $\alpha < h(E' - E, T(k))$, то установить $x \leftarrow x', E \leftarrow E'$ и перейти к следующей итерации. Иначе повторить шаг (b), пока не будет найдена подходящая точка x' .

Известны следующие модификации этого алгоритма:

- Модификация А. На шаге 2e переход к следующей итерации происходит и в том случае, если точка x' не являлась подходящей. При этом следующая итерация начинается с точки x , но уже с новым значением температуры.
- Модификация Б. В качестве оценки точки минимума возвращается последнее значение x . Это может незначительно ускорить алгоритм в случае большой размерности S , но с небольшой вероятностью может привести к тому, что будет получено худшее решение (особенно, если температура к моменту завершения алгоритма осядет значительно больше нуля).
- Модификация В. На шаге 2b x' вычисляется рекуррентно с использованием формулы $x' = G(x', T(k))$. Изначально на шаге 1 устанавливается $x' \leftarrow x_0$. Это позволяет избежать “застывания” алгоритма, однако такая реализация теряет множество преимуществ метода отжига, т. к. не очень сильно отличается от обычного случайного поиска (особенно, если это комбинируется с вариантом 1 — в этом случае проверку $h(\Delta E, T)$ вообще можно исключить).

В следующем разделе будут рассмотрены основные схемы выбора параметров метода отжига.

3. Общие схемы метода отжига

3.1. Больцмановский отжиг

Исторически первой схемой метода отжига является т.н. схема *Больцмановского отжига*. Именно эта схема использовалась Н. Метрополисом и др. в [1] для вычисления многомерных интегралов пути в задачах статистической физики, а также С. Киркпатриком и др. в [2] для решения задачи нахождения оптимальной

разводки микросхем. В Больцмановском отжиге изменение температуры задаётся формулой

$$T(k) = \frac{T_0}{\ln(1+k)}, k > 0.$$

Семейство распределений $\mathcal{G}(x, T)$ выбирается как семейство нормальных распределений с математическим ожиданием x и дисперсией T , т. е. задаётся плотностью

$$g(x'; x, T) = (2\pi T)^{-D/2} \exp(-|x' - x|^2/(2T)),$$

где D — размерность пространства состояний. Пространство состояний предполагается метрическим.

Для Больцмановской схемы доказано (см. [1]), что при достаточно больших T_0 и общем количестве шагов K , выбор такого семейства распределений гарантирует нахождение глобального минимума.

3.2. Отжиг Коши (быстрый отжиг)

Основным недостатком Больцмановского отжига является очень медленное убывание температуры. Например, для того, чтобы понизить исходную температуру в 40 раз, требуется $e^{40} \approx 2.35 \cdot 10^{17}$ итераций, что уже вряд ли приемлемо при решении каких-либо задач. Ввиду этого Цу и Хартли в [14] предложили алгоритм, который позволяет использовать для изменения температуры схему

$$T(k) = \frac{T_0}{k} \tag{1}$$

без потери гарантии нахождения глобального минимума. Это достигается за счет использования в качестве \mathcal{G} распределений Коши с плотностью

$$g(x'; x, T) = \frac{T}{(|x' - x|^2 + T^2)^{(D+1)/2}},$$

соответствующим образом нормированных. Например, в случае $D = 1$ приходим к плотности

$$g(x'; x, T) = \frac{1}{\pi} \frac{T}{|x' - x|^2 + T^2}.$$

К сожалению, это распределение не очень удобно моделировать в пространстве размерности больше 1. Этого можно избежать, например, с помощью перемножения D одномерных распределений Коши:

$$g(x'; x, T) = \frac{1}{\pi^D} \prod_{i=1}^D \frac{T}{|x_i' - x_i|^2 + T^2},$$

но в этом случае нахождение глобального минимума гарантируется только при законе изменения температуры не быстрее чем:

$$T(k) = \frac{T_0}{k^{1/D}},$$

что гораздо медленнее схемы (1).

3.3. Сверхбыстрый отжиг

Недостатки двух предыдущих методов привели к тому, что в 1989 году американским исследователем Л. Ингбером был разработан метод сверхбыстрого отжига (Very Fast Annealing) [15]. В нем пространство S считается состоящим из D -мерных векторов $\begin{pmatrix} x_1 \\ \vdots \\ x_D \end{pmatrix}$, где $x_i \in [A_i, B_i]$. Кроме этого, температура по каждой из координат может различаться, таким образом, T также является вектором размерности D .

Семейство распределений строится следующим образом. Вводится функция

$$g_T(y) = \prod_{i=1}^D \frac{1}{2(|y_i| + T_i) \ln(1 + 1/T_i)} \equiv \prod_{i=1}^D g_{(i,T)}(y_i), \quad y_i \in [-1, 1].$$

В качестве y для получения плотности распределений \mathcal{G} используется $\frac{\Delta x}{B_i - A_i}$, таким образом, новое значение x_i' вычисляется по формуле $x_i' = x_i + z_i(B_i - A_i)$, где z_i — случайная величина с плотностью $g_{(i,T)}$ на $[-1, 1]$. При этом выходящие за границы интервала значения параметра генерируются заново (пока не произойдет попадание в интервал) или приравниваются соответствующим границам.

Такую случайную величину легко промоделировать:

$$z_i = \operatorname{sgn}(\alpha_i - \frac{1}{2}) T_i ((1 + 1/T_i)^{2\alpha_i - 1} - 1), \quad (2)$$

где α_i — независимые случайные величины, распределенные равномерно на $[0, 1]$.

Доказано (см. [15]), что закон изменения температуры

$$T_i(k) = T_{(i;0)} \exp(-c_i k^{1/D}), \quad c_i > 0$$

дает статистическую гарантию нахождения глобального минимума. Для вероятности принятия также используется отдельная шкала температуры, изменяющаяся по такому же закону. Как правило, при реализации этого метода c_i управляется двумя параметрами:

$$c_i = m_i \exp(-n_i/D).$$

Преимущества такого метода очевидны. Во-первых, экспоненциальное убывание температуры гораздо быстрее достижимого в предыдущих методах. Во-вторых, разделение размерностей может дать большой выигрыш, как и благодаря отдельным температурам (т. к. специфика конкретной задачи может сильно различать параметры), так и благодаря ускорению процесса, в случае, если не нужно менять все координаты одновременно.

Кроме того, в отличие от отжига Коши, сверхбыстрый отжиг, как было показано, допускает очень быстрое моделирование распределения \mathcal{G} независимо от размерности S .

Среди недостатков этого метода можно назвать то, что ввиду большого количества параметров иногда требуется несколько месяцев, чтобы хорошо настроить его для решения конкретной задачи.

3.4. Алгоритм Ксин Яо

Алгоритм Ксин Яо [16] был получен повторным применением идеи предыдущего алгоритма. В качестве $g_T(y)$ выбирается

$$g_T(y) = \prod_{i=1}^D g_{(i;T)}(y_i) = \prod_{i=1}^D \frac{1}{2(|y_i| + \frac{1}{\ln(1/T_i)}) \ln(1 + \ln(1/T_i))}.$$

Утверждается, что при изменении температуры по закону

$$T_i(k) = T_{(i;0)} \exp(-\exp(b_i k^{1/D})), \quad b_i > 0$$

достигается статистическая гарантия нахождения глобального минимума.

Однако, как показано в [16], увеличение скорости убывания температуры вовсе не означает ускорения в решении задачи. Более того, “размазанность” распределения приводит к тому, что метод генерирует огромное число “длинных” переходов, которые отвергаются в силу низкой вероятности их принятия.

Таким образом, несмотря на то, что этот процесс можно итерировать до бесконечности, получая законы изменения температуры вроде

$$T_i(k) = T_{(i,0)} \exp(-\exp(\exp(\dots \exp(b_i k^{1/D}) \dots))),$$

ценность таких “улучшений” представляется сомнительной. Более того, легко видеть, что в пределе это приводит к тривиальному методу случайного поиска, которым является метод отжига при $T = 0$.

Это в небольшой степени применимо и к методу сверхбыстрого отжига, так что вопрос о скорости сходимости этих методов, а также о других методах, обеспечивающих не такое быстрое убывание температуры, но бóльшую скорость сходимости, остается открытым.

Следовательно, работу [16] можно рассматривать и как некую скрытую критику метода сверхбыстрого отжига. Однако, вполне возможны задачи, на которых вторая итерация вышеописанного процесса может давать неплохие результаты, поэтому этот метод и рассмотрен в настоящей работе.

3.5. Методы “тушения”

Далеко не всегда хватает вычислительных ресурсов на поиск глобального минимума. Кроме того, зачастую достаточно достигнуть не глобально оптимального решения задачи, а достаточно близкого к нему. Методы “тушения” (simulated quenching) не гарантируют нахождения глобального минимума, но, как правило, быстро находят близкое решение, а на практике зачастую и сам оптимум.

Основная идея этих методов заключается в том, чтобы скомбинировать семейство распределений \mathcal{G} одного из предыдущих четырех методов с более быстрым законом убывания температуры.

Например, можно рассматривать нормальные распределения \mathcal{G} из Больцмановского отжига, но при этом уменьшать температуру по закону

$$T_{k+1} = cT_k.$$

Как правило, в этом случае c выбирается между 0.7 и 0.99. Такой метод очень быстро сходится, и для конкретных задач может давать весьма неплохое решение, близкое к оптимальному, в условиях реального времени.

Некоторые из методов “тушения” рассмотрены в [11] и [17]. Зачастую они основаны либо на нормальном распределении, либо на распределении для сверхбыстрого отжига. Кроме того, встречаются специальные распределения, подобранные опытным путем для решения конкретных задач.

3.6. Масштабирование в ходе отжига

Зачастую при реализации сверхбыстрого отжига в задачах с большой размерностью используется технология *масштабирования отжига* (reannealing), иногда также применяемая и к другим вариантам отжига. При использовании этой технологии периодически во время отжига производится следующая операция. Обозначим за s_i значение некоторой оценки производной целевой функции по i -й координате в точке текущего минимума:

$$s_i \geq \left| \frac{\partial f}{\partial x_i}(x_i^{min}) \right|.$$

Кроме того, пусть $s_{max} = \max_{1 \leq i \leq D} s_i$. После этого номер шага (называемый в этом алгоритме *временем отжига* (annealing-time)) и температура для каждой размерности изменяются следующим образом:

$$\begin{aligned} T'_i &= T_i(k_i)(s_{max}/s_i), \\ k'_i &= (\ln(T_{(i;0)}/T'_i)/c_i)^D. \end{aligned}$$

Таким образом, значение $T_{(i;0)}$ сохраняется, а время отжига (которое теперь может принимать не только целые значения) масштабируется согласно температуре.

Температура, используемая для расчета вероятности принятия, также масштабируется. Способ масштабирования зависит от задачи, но, как правило, коэффициент масштабирования определяется значением целевой функции в точке текущего минимума (чем меньше значение функции, тем меньше температура). Время отжига для пересчета температуры принятия также изменяется аналогичным образом.

Кроме приведенного, возможны другие способы масштабирования, зависящие от конкретной задачи. Как правило, они сводятся к другому выбору *коэффициентов чувствительности* s_i . Приведенный выше способ выбора s_i используется для нелинейных физических задач большой размерности.

Алгоритм сверхбыстрого отжига с использованием этой технологии описан в [11], [15] и [17], и носит название *адаптивного метода отжига* (adaptive simulated annealing).

4. Моделирование схем отжига

Для сравнения различных схем метода отжига моделировалось поведение их оценок. Вопрос моделирования различных схем отжига разбивается на две части:

- промоделировать соответствующее распределение,
- промоделировать схему изменения температуры.

4.1. Больцмановский отжиг

Нормальное распределение моделировалось с помощью центральной предельной теоремы. Обозначим α_i независимые реализации равномерного распределения на $[0; 1]$. Тогда $\mathbf{E}\alpha_i = 0$, $\mathbf{D}\alpha_i = 1/12$ и

$$\frac{\sum_{i=1}^n \alpha_i}{\sqrt{n/12}} \xrightarrow{n \rightarrow \infty} \mathbf{N}(0; 1).$$

Таким образом, можно сложить n независимых реализаций α_i , и разделив их сумму на $\sqrt{n/12}$, мы будем иметь достаточно хорошее приближение к нормальному распределению. Как показывает

практика, достаточно взять $n = 24$. Для моделирования требуемого $\mathbf{N}(0; T)$ достаточно домножить получившееся число на \sqrt{T} . В случае размерности, большей единицы, по каждой из координат вводились независимые возмущения с таким распределением.

Для моделирования убывания температуры использовалась непосредственно формула $T_0 / \ln k$, $k = 2, 3, \dots$

4.2. Отжиг Коши

Распределение Коши моделировалось методом обратных функций. Пусть \mathbf{X} — случайная величина, имеющая распределение Коши с плотностью

$$g(x) = \frac{1}{\pi} \frac{T}{x^2 + T^2}.$$

Вычислим $\mathbf{P}\{\mathbf{X} < y\}$:

$$\mathbf{P}\{\mathbf{X} < y\} = \int_{-\infty}^y \frac{1}{\pi} \frac{T}{x^2 + T^2} dx = \frac{1}{\pi} \left(\operatorname{arctg} \left(\frac{y}{T} \right) + \frac{\pi}{2} \right).$$

Следовательно, если α — реализация равномерно распределенной на $[0; 1]$ случайной величины, то величина

$$T \operatorname{tg} \left(\pi \alpha - \frac{\pi}{2} \right) \tag{3}$$

имеет требуемое распределение.

В случае задач с размерностью $D > 1$ использовалось произведение D одномерных распределений Коши, т. е. вносилось возмущение по каждой из координат, вычисленное по формуле (3) для независимых α_i .

Для температуры в этом случае используется формула

$$T(k) = \frac{T_0}{k^{1/D}},$$

причем в случае $D = 1$ она вычислялась как T_0/k , в случае $D = 2$ — как T_0/\sqrt{k} для увеличения скорости работы программ.

4.3. Сверхбыстрый отжиг

Для сверхбыстрого отжига использовалась также полученная методом обратных функций формула (2). Закон изменения температуры также имел отличия при $D = 1$ и $D = 2$. Если при $D = 2$ отличие, аналогично предыдущему случаю, сводилось к вычислению $k^{1/2}$ как \sqrt{k} , то при $D = 1$ возможна более существенная оптимизация, т. к.

$$\frac{T(k+1)}{T(k)} = \frac{e^{-c_i/(k+1)}}{e^{-c_i/k}} = e^{-c_i} = \text{const.}$$

Значение этой константы вычислялось заранее, и при каждом уменьшении температуры производилось умножение на нее вместо вычисления экспоненты и деления, что примерно в 200 раз быстрее.

Для предотвращения деления на 0 в формулах температура была ограничена снизу значением 10^{-4000} .

4.4. Методы “тушения”

Кроме вышеописанных методов, были смоделированы четыре метода тушения:

- Больцмановский отжиг с уменьшением температуры как в методе Коши,
- Больцмановский отжиг с экспоненциальным уменьшением температуры ($T_{k+1} = cT_k$, $c = 0.99$),
- метод Коши с экспоненциальным уменьшением температуры ($T_{k+1} = cT_k$, $c = 0.99$),
- сверхбыстрое тушение.

Сверхбыстрое тушение получается из метода сверхбыстрого отжига с использованием следующих формул:

$$T_i(k) = T_{(i;0)} \exp(-c_i k^{Q/D}), \quad c_i > 0,$$
$$c_i = m_i \exp(-n_i/D),$$

где Q — так называемый *множитель тушения*. При тестировании множитель тушения брался равным 2.

5. Сравнение схем отжига на задаче о нахождении точки минимума функции со случайной помехой

В этом разделе приводятся результаты сравнения реализованных схем отжига. В качестве исследуемой функции выбрана

$$f(x_1, \dots, x_{30}) = \sum_{i=1}^{30} x_i^4 + \alpha,$$

где α — случайная помеха, распределенная равномерно на $[0; 1]$.

Множество минимизации задавалось условиями $-1000 \leq x_i \leq 1000$.

Выбор параметров и начальной точки

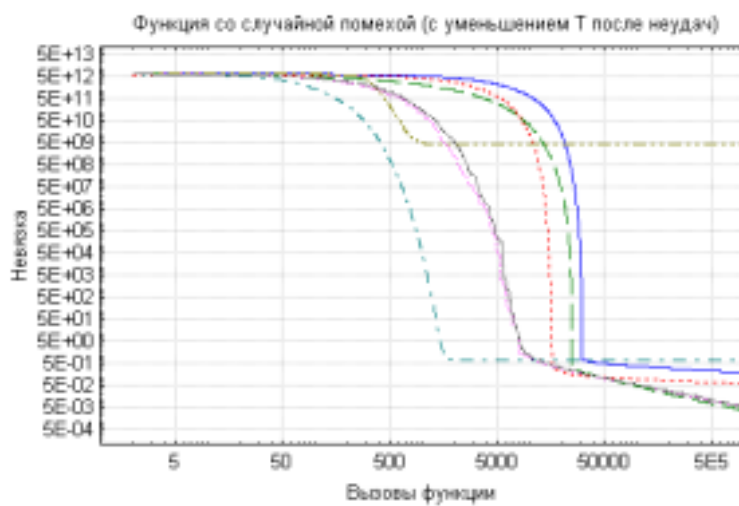
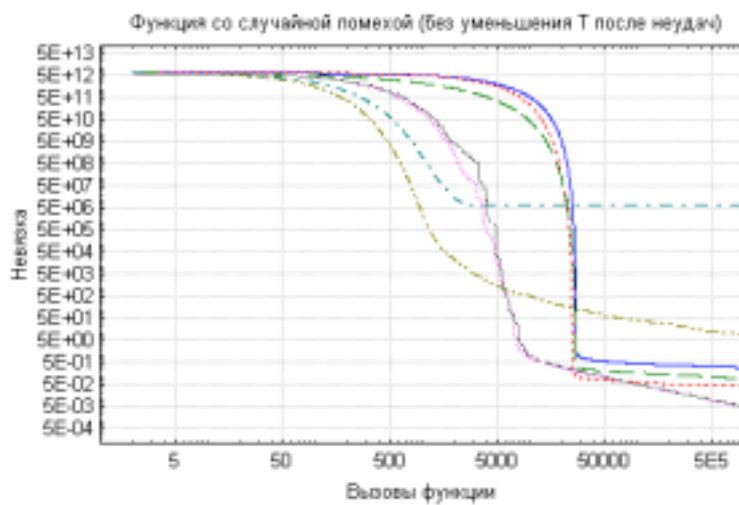
Начальная точка выбиралась равномерно распределенной на множестве минимизации. Для сверхбыстрого отжига использовались $m_i = 1$, $n_i = -160$, для сверхбыстрого тушения — $m_i = 1$, $n_i = -80$. Температура по всем координатам совпадала, стартовые значения подбирались опытным путем. Ниже приводится таблица стартовых температур:

Метод	T_0
Больцм. отжиг без уменьшения T после неудач	1
Больцм. отжиг с уменьшением T после неудач	1
Отжиг Коши без уменьшения T после неудач	0.03
Отжиг Коши с уменьшением T после неудач	0.05
Сверхбыстрый отжиг без уменьшения T	0.001
Сверхбыстрый отжиг с уменьшением T	10^{-4}
Больцм. тушение (по Коши) без уменьшения T	450
Больцм. тушение (по Коши) с уменьшением T	1 000
Больцм. тушение ($T_{k+1} = cT_k$) без уменьшения T	500
Больцм. тушение ($T_{k+1} = cT_k$) с уменьшением T	40 000
Тушение Коши ($T_{k+1} = cT_k$) без уменьшения T	10
Тушение Коши ($T_{k+1} = cT_k$) с уменьшением T	800
Сверхбыстрое тушение без уменьшения T	0.5
Сверхбыстрое тушение с уменьшением T	10^{-4}

Результаты тестирования

Каждая схема запускалась в двух вариантах — классическом, а также в варианте с уменьшением температуры и после неудач (вариант 2.А). Для получившихся 14 схем приводятся графики невязки, полученной в результате усреднения по 100 запускам метода.

На первом графике сравниваются методы в классическом варианте, на втором — в варианте 2.А. На графике отображено изменение невязки по мере вычисления значений функции в новых точках.



Ниже приводится таблица обозначений для графиков, приведенных для тестов:

- Вольфмановский отжиг
- Отжиг Коши
- Сверхбыстрый отжиг
- Вольфмановское тушение (убывание T по Коши)
- Вольфмановское тушение (экспоненциальное убывание T)
- Тушение по Коши-1
- Сверхбыстрое тушение

Анализ результатов

Видно, что различные методы отжига хорошо работают с функциями с помехой, за исключением некоторых методов тушения. Известно, что скорость сходимости любого метода, основанного только на значениях функции с невырожденными неубывающими помехами, не может быть по порядку быстрее, чем корень из числа итераций. Ниже приводится таблица невязки при использовании метода сверхбыстрого отжига в зависимости от числа вызовов функций:

Число итераций	Невязка
10000	0.763129571893273058
100000	0.043839954318724353
1000000	0.004662248394661734
10000000	0.000547701824698814
100000000	0.000056790943476683
1000000000	0.000007776614600166

При тестировании использовалось усреднение по десяти запускам.

Непосредственно из приведенной таблицы видно, что скорость убывания невязки достаточно высока, и, скорее всего, совпадает с оптимальной теоретической оценкой. К сожалению, доказательство этого факта для метода отжига до сих пор не найдено.

6. Заключение

Метод отжига является весьма эффективным алгоритмом случайного поиска глобального минимума. Версия алгоритма сверхбы-

строго отжига сходится значительно быстрее других методов, поэтому в практических задачах ее использование, видимо, наиболее целесообразно.

Пока мало изучены возможности метода отжига для решения других задач, сводящихся к оптимизационным. Надеюсь, что в ближайшее время будут получены теоретические результаты относительно скорости сходимости метода отжига. К сожалению, на настоящий момент единственные известные результаты касаются статистической гарантии сходимости (см. [13], [14], [15], [18]).

Список литературы

- [1] *Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., and Teller E.* Equation of State Calculations by Fast Computer Machines // *J. Chemical Physics*. 21. 6. June. 1953. P. 1087–1092.
- [2] *Kirkpatrick S., Gelatt Jr. C. D., and Vecchi M. P.* Optimization by Simulated Annealing // *Science*. 220. 1983. P. 671–680.
- [3] *Ingber L., Mondescu R. P.* Optimization of Trading Physics Models of Markets // *IEEE Trans. Neural Networks*. 12(4). 2001. P. 776–790.
- [4] *Ingber L., Wilson J. K.* Statistical mechanics of financial markets: Exponential modifications to Black-Scholes // *Mathematical Computer Modelling*. 31(8/9). 2000. P. 167–192.
- [5] *Forman M. C., Aggoun A., McCormick M.* Simulated Annealing for Optimisation and Characterisation of Quantisation Parameters in Integral 3D Image Compression // *The Institute of Mathematics and its Applications*. Horwood. 2000. P. 393–413.
- [6] *Forman M. C.* Compression of Integral Three Dimensional Television Pictures / Ph. D. Thesis at De Montfort University Leicester. 2000. United Kingdom.
- [7] *Jeong C., Kim M.* Fast Parallel Simulated Annealing for Traveling Salesman Problem on SIMD Machines with Linear Interconnections // *Parallel Computing*. 17. 1991. P. 221–228.

- [8] *Yao X.* Call Routing by Simulated Annealing // International Journal of Electronics. Oct. 1995.
- [9] *Cohen B.* Training Synaptic Delays in a Recurrent Neural Network / Thesis submitted towards the degree "Master of Science" at Tel-Aviv University. 1994.
- [10] *Уоссермен Ф.* Нейрокомпьютерная техника: Теория и практика / Перевод на русский язык, Ю. А. Зуев, В. А. Точенов. 1992.
- [11] *Ingber L.* Simulated Annealing: Practice versus theory // Mathematical and Computer Modelling. 18(11). 1993. P. 29–57.
- [12] *Ingber L., Rosen B.* Genetic Algorithms and Very Fast Simulated Reannealing: A Comparison // Mathematical and Computer Modelling. 16(11). 1992. P. 87–100.
- [13] *Geman S., Geman D.* Stochastic relaxation, Gibbs distribution and the Bayesian restoration in images // IEEE Trans. Patt. Anal. Mac. Int. 6(6). 1984. P. 721–741.
- [14] *Szu H. H., Hartley R. L.* Fast Simulated Annealing // Physical Letters A. 122. 1987. P. 157–162.
- [15] *Ingber L.* Very fast simulated re-annealing // Mathematical and Computer Modelling. 12. 1989. P. 967–973.
- [16] *Yao X.* A New Simulated Annealing Algorithm // International Journal of Computer Mathematics. 56. 1995. P. 161–168.
- [17] *Ingber L.* Adaptive simulated annealing (ASA): Lessons learned // Journal "Control and Cybernetics". 1995.
- [18] *Yao X.* Simulated Annealing with Extended Neighbourhood // International Journal of Computer Mathematics. 40. 1991. P. 169–189.