© 2013 г. Е. А. Нурминский, д-р физ.-мат. наук,
(Дальневосточный федеральный университет, Владивосток, Россия),
Д. Тьен, д-р фил.,
(Университет Чарльза Стюрта, Батхерст, Австралия)

МЕТОД СОПРЯЖЕННЫХ СУБГРАДИЕНТОВ С ОГРАНИЧЕННОЙ ПАМЯТЬЮ[1]

В работе представлен метод решения выпуклых задач недифференцируемой оптимизации, использующий основную идею классического метода сопряженных градиентов и совпадающий с ним в случае квадратичных функций. Основное отличие от ранее рассмотренных аналогов состоит в фиксированном заранее ограничении на объем используемой памяти, независящим от требуемой точности получаемого решения. Численные эксперименты показывают практически линейную скорость сходимости этого алгоритма.

1. Введение

В этой работе рассматривается выпуклая задача недифференцируемой оптимизации

$$\min_{x \in E} f(x) = f_\star, \tag{1}$$

где E является конечномерным евклидовым пространством с обычным скалярным произведением xy и соответствующей нормой $\|x\| = \sqrt{xx}$. Предполагается, что задача (1) вполне определена в том смысле, что ее решение существует. Основная цель работы заключается в том, чтобы предложить для таких задач метод типа сопряженных градиентов с фиксированной априори границей на используемую память, доказать его сходимость и представить некоторые обнадеживающие результаты вычислительных экспериментов.

Наиболее общие методы решения задачи (1) используют так называемые субградиентные оракулы, обеспечивающие возможность вычисления в произвольной точке значения целевой функции $f(x)$ и некоторого субградиента g из субдифференциального множества $\partial f(x)$. Простейшим из таких методов является субградиентный алгоритм

$$x^{k+1} = x^k - \lambda_k g^k,\ g^k \in \partial f(x^k),\ k = 0, 1, \ldots \tag{2}$$

который активно исследовался, начиная с пионерских работ Н.З. Шора [1], Б.Т. Поляка [2]. В наиболее общем случае было показано, что (2) сходится к решению (1) при весьма слабых условиях, если шаговые множители λ_k удовлетворяют условиям "расходимости ряда": $\sum_k \lambda_k = \infty, \lambda_k \to +0$. Однако, численные эксперименты и теоретический анализ показали что это правило выбора шаговых множителей приводит как правило к медленной сходимости. Поэтому практически сразу начался поиск более эффективных алгоритмов, в частности многочисленные варианты квадратично-линейных алгоритмов, использующих кусочно-линейные модели

[1] Работа выполнена при финансовой поддержке РФФИ (грант №13-07-12010).

негладких целевых функций и квадратичные поправки, повышающие, по замыслу, точность этих аппроксимаций [5], [17], [8]. Весьма эффективными на практике оказались идеи переменной метрики [6], из последних идей в этой области можно упомянуть сглаживание в комбинации с оптимальными градиентными схемами гладкой оптимизации [9], расщепление на подпространства гладкости-негладкости (так называемые UV-алгоритмы [10]).

Вместе с тем, начиная с работ [3,4], исследовались м аналоги метода сопряженных градиентов [11, 12], [13]. Эти исследования развивались в основном в следующих двух направлениях. В первом из них, восходящих к начальным работах Ph.Wolfe [3, 4], алгоритмы использовали накапливаемый пакет субградиентов, полученных на некоторой предыстории текущей итерации. Этот пакет субградиентов использовался для попытки построения направления убывания целевой функции как решения задачи

$$\|p^k\|^2 = \min_{\substack{p \in \mathrm{co}\{g^0, g^1, \dots, g^k\} \\ g^i \in \partial f(x^i), i = 0, 1, \dots, k}} \|p\|^2 \tag{3}$$

где $x^0, x^1, \dots, x^k$ представляет собой историю предыдущих итераций (пробных и рабочих шагов). Здесь для простоты обозначений рассматривается предыстория, начинающаяся с нулевой (начальной) итерации.

Далее, в зависимости от различных условий, происходит либо рестарт алгоритма, что можно рассматривать как смену начальной точки, либо дальнейшее накопление информации с помощью пробных шагов, либо делается рабочий шаг вида

$$x^{k+1} = x^k - \lambda_k p^k, \tag{4}$$

где λ_k определяется с помощью точной или приближенной одномерной оптимизации. Проблема с этим подходом заключается в том, что объем накапливаемого пакета субградиентов определяется оценками точности решения, полученного в данном цикле, и при повышении точности неограниченно возрастает. Помимо увеличения затрат памяти и роста объема обрабатываемых данных, с соответствующим достаточно быстрым ростом объем вычислений, это приводит и к нежелательным алгоритмическим последствиям: некоторый неудачный пробный шаг может усложнить поиск хорошего направления убывания на протяжении многих последующих итераций.

Второе направление можно рассматривать как приближенное решение задачи (3) с использованием аналогов формулы Полака-Рибьера для построения сопряженных направлений см., например, [11, 12], где использовались постоянные шаговые множители и постоянные весовые коэффициенты. Вычислительные эксперименты, проведенные с этими алгоритмами, показали, однако, прогрессивный рост количества пробных шагов в ситуации, например, отсутствия острого "минимума". Соответственно сходимость алгоритма при этом сильно замедлялась в терминах временных вычислительных затрат. Для гладких экстремальных задач эти методы тем не менее показали неплохие вычислительные результаты и под названием SR-методов (shortest residuals) активно исследуются и в настоящее время [14], [15, 16].

В настоящей работе для решения задач выпуклой недифференцируемой оптимизации производится в некотором смысле объединение этих двух подходов: в предлагаемом варианте метода сопряженных субградиентов сохраняются преимущества метода Ph.Wolfe, заключающиеся в синтетическом использовании субградиентого пакета, а с другой стороны ограничивается объем вычислительных затрат на обработку этого пакета. Последнее достигается заданным apriori ограничением на размер пакета, по достижении которого происходит рестарт алгоритма.

2. Метод сопряженных субградиентов с ограниченной памятью

Мы рассматриваем метод сопряженных субградиентов с ограниченной памятью как правило построения последовательности $x^k, k = 0, 1, \ldots$, сходящейся при определенных условиях к решению задачи (1). Это правило использует пакет субградиентов, накапливаемый и модифицируемый в ходе работы алгоритма. В общем виде этот пакет обозначается через $G(z, s, t)$ и состоит из инициализирующего вектора z и субградиентов $g^i \in \partial f(x^i), i = s, s+1, \ldots, t$, вычисленных на итерациях $s, s+1, \ldots, t$:

$$G(z, s, t) = \{z, g^i \in \partial f(x^i), i = s, s+1, \ldots, t\ \}.$$

Инициализирующий вектор z используется для передачи информации при рестартах алгоритма. Для упрощения обозначений мы обозначим через $G_{co}(z, s, t)$ выпуклую оболочку конечного множества $G(z, s, t)$.

Пакет $G(z, s, t)$ состоит не более чем из $N + 1$ векторов, где N — фиксированный параметр метода. В алгоритме задается также управляющая последовательность величин $\delta_k \to +0, k = 0, 1, \ldots$, являющихся некоторым эквивалентом точности выполнения условий оптимальности.

Алгоритм строит последовательность итераций, прерываемых моментами рестарта, во время которых происходит очистка пакета $G(\cdot, \cdot, \cdot)$ от накопленных субградиентов и изменение инициализирующего вектора. Рестарт алгоритма выполняется при выполнении по крайней мере одного условия: либо количество субградиентов в пакете достигает максимального размера N, либо норма кратчайшего вектора в выпуклой оболочке пакета $G_{co}(\cdot, \cdot, \cdot)$ падает ниже текущей оценки выполнения условий оптимальности.

В введенных обозначениях метод имеет следующий вид.

Инициализация Установить начальные значения счетчика рестартов $r = 0$, счетчика итераций $t = 0$, определить начальный момент рестарта $t_r = 0$. Задать максимальный размер N пакета субградиентов, последовательность оценок точности $\{\delta_k\}$ и начальную точку x^0. Вычислить $g^0 \in \partial f(x^0)$, координирующий вектор z^0 установить равным g^0 и положить начальный пакет $G_0 = G(z^0, 0, 0)$ равным $\{g^0, g^0\}$.

При условии, что проделано t итераций, за которые произведено r рестартов и последний из них произошел в момент $t_r < t$, очередная $t + 1$-ая итерация выполняется следующим образом:

$t + 1$**-ая итерация**

Шаг 1. Решить задачу поиска элемента минимальной евклидовой нормы

$$\min_{p \in G_{co}(z^r, t_r, t)} \|p\|^2 = \|p^t\|^2. \qquad (5)$$

Если $\|p^t\| > \delta_r$, то перейти к **Шагу 2.**

Если $\|p^t\| \leqslant \delta_r$ то осуществить рестарт алгоритма с увеличением требований к точности выполнения условий оптимальности:

- Увеличить счетчик рестартов $r = r + 1$ и полностью ре-инициализировать пакет субградиентов:

$$t_r = t,\ z^r = g^t \in \partial f(x^t),\ G(z^r, t, t) = \{g^t\}; \tag{6}$$

- Повторить Шаг 1.

Шаг 2. Решить одномерную задачу

$$\min_{\lambda} f(x^t - \lambda p^t) = f(x^t - \lambda_t p^t) = f(x^{t+1}) \leqslant f(x^t) \tag{7}$$

и выбрать $g^{t+1} \in \partial f(x^{t+1})$ такой, что $g^{t+1} p^t = 0$. Условие оптимальности для этой задачи гарантирует существование такого g^{t+1} даже для $\lambda_t = 0$,.

При нахождении точки x^{t+1} методом дихотомии, то есть как предел вложенных интервалов $[x^t - \lambda_k^- p^t, x^t - \lambda_k^+ p^t],\ k = 0, 1, \dots$ с $g_-^k p^t = \alpha_k^- < 0$, а $g_+^k p^t = \alpha_k^+ > 0$, где $g_-^k \in \partial f(x^t - \lambda_k^- p^t)$, а $g_+^k \in \partial f(x^t - \lambda_k^+ p^t)$, такой вектор может быть найден как предел последовательности $\bar{g}^k = \gamma_k g_-^k + (1 - \gamma_k) g_+^k,\ k = 0, 1, \dots$ где $\gamma_k = \dfrac{\alpha_k^+}{\alpha_k^+ - \alpha_k^-} \in [0, 1]$ и обращает $\bar{g}^k p^t$ в нуль. По полунепрерывности сверху субдифференциального отображения ∂f вектор $g^{t+1} = \lim_{k\to\infty} \bar{g}^k \in \partial f(x^{t+1})$, где предел возможно будет необходимо брать по произвольной сходящейся подпоследовательности.

Шаг 3. Пополнить множество $G(z^r, t_r, t)$ вектором g^{t+1}:

$$G(z^r, t_r, t+1) = \{G(z^r, t_r, t), g^{t+1}\},$$

увеличить значение счетчика итераций $t \to t + 1$ и перейти к **Шагу 4.**

Шаг 4. Этот шаг осуществляет рестарт алгоритма при достижении ограничений на объем используемой памяти, без изменений счетчика рестартов и текущей оценки точности решения. Если $t - t_r \geqslant N$, изменить координирующий вектор $z^r = p^t$, переопределить момент последнего рестарта $t_r = t$, инициализировать пакет субградиентов $G(z^r, t_r, t) = \{z^r, g^t\}, g^t \in \partial f(x^t)$. и перейти к **Шагу 1.**

3. Сходимость алгоритма

Для исследования сходимости алгоритма мы используем условия сходимости, подробно рассмотренные в [18]. С точки зрения этих условий алгоритм решения оптимизационной задачи представляет собой некое правило для построения последовательности приближенных решений $\{x^k\}$, которая должна сходится к некоторому желаемому множеству $X_\star$. Само это множество обычно определяется соответствующими условиями оптимальности.

Слабая форма сходимости (сходимость по подпоследовательности) гарантируется при выполнении следующих условий:

A1 Последовательность $\{x^k\}$ ограничена.

A2 Существует непрерывная функция $W(x) : E \to \mathbb{R}$ такая, что если у $\{x^k\}$ есть предельная точка $x' \notin X_\star$ то у этой последовательности есть другая предельная точка x'' такая, что $W(x'') < W(x')$.

При выполнении этих условий у последовательности $\{x^k\}$ существует предельная точка $x^\star \in X_\star$.

Сильная форма сходимости $\{x^k\}$ к множеству $X_\star$ имеет место при выполнении несколько более сильных условий:

B1 Последовательность $\{x^k\}$ ограничена.

B2 Для произвольной подпоследовательности $\{x^{k_t}\} \to x' \notin X_\star$ при $t \to \infty$ существует $\epsilon > 0$ такое, что для любого t существует момент выхода из -окрестности x':

$$m_t = \inf\{m : \|x^{k_t} - x^m\| > \epsilon\} < \infty. \tag{8}$$

B3 Существует непрерывная функция $W(x) : E \to \mathbb{R}$ такая, что

$$\limsup_{t\to\infty} W(x^{m_t}) < \lim_{t\to\infty} W(x^{k_t}) = W(x') \tag{9}$$

для всех подпоследовательностей $\{x^{k_t}\}, \{x^{m_t}\}$ отвечающих условию **B2**.

B4 Множество $W_\star = \{W(x^\star), x^\star \in X_\star\}$ такого, что $\mathbb{R} \setminus W_\star$ всюду плотно.

B5 Если $\{x^{k_t}\} \to x^\star \in X_\star$ то $\|x^{k_t+1} - x^{k_t}\| \to 0$ при $t \to \infty$.

При выполнении этих условий все предельные точки $\{x^k\}$ принадлежат $X_\star$ [18].

Используя приведенные условия сходимости, можно доказать следующее утверждение.

Теорема 1. *Пусть f конечна, сильно выпукла и лебегово множество $\{x : f(x) \leqslant f(x^0)\}$ ограничено. Тогда $\{x^t\}$ сходится к единственному решению задачи 1.*

Доказательство. Для применения условий **B1-B5** прежде всего определим множество $X_\star$ как точку $x^\star$, удовлетворяющую необходимым, а данном случае и достаточным, условиям оптимальности $0 \in \partial f(x^\star)$.

В силу монотонности метода все элементы последовательности $\{x^t\}$ принадлежат ограниченному множеству $\{x : f(x) \leqslant f(x^0)\}$ так что условие **B1** тривиальным образом выполнено. Далее предположим, что условие **B2** не выполнено, то есть вся последовательность $\{x^t\}$ сходится к некоторой точке $x' \notin X_\star$, где соответственно $0 \notin \partial f(x')$. В силу полунепрерывности сверху ∂f и конечности f существует достаточно малое $\epsilon > 0$ такое, что для некоторых $0 < \gamma < \Gamma < \infty$ имеет место оценки $\gamma \leqslant \|g\| \leqslant \Gamma$ для любых $g \in \partial f(x), \|x - x'\| \leqslant 4\epsilon$. Для простоты обозначений мы опускаем зависимость γ, Γ от x' и ϵ, поскольку они в последующих рассуждениях фиксированы.

Покажем, что при сделанном предположении возникает бесконечная последовательность $k = 0, 1, \ldots$ таких итераций, для которых $\|z^k\| \leqslant \delta_k \to 0$.

Действительно, если это не так, то существует $\bar{k}$ такое, что $\|p^t\| \geqslant \delta_{\bar{k}}$ для всех $t > t_{\bar{k}}$. Без потери общности можно считать t настолько большим, что $\|x^t - x'\| \leqslant \epsilon$. Предположение об ограниченности $\bar{k}$ по сути дела означает, что все последующие рестарты будут происходить в соответствии лишь с **Шагом 4,** по достижению максимального размера пакета векторов $G(\cdot)$.

Тогда для $t_{\bar{k}} < t \leqslant t_{\bar{k}} + N$

$$\|z^t\|^2 = \min_{z \in G_{co}(z^{\bar{k}},\ t_{\bar{k}}, t)} \|z\|^2 \leqslant \min_{\mathrm{co}\{z^{\bar{k}}, g^{t_{\bar{k}}+1}\}.} \|z\|^2 .$$

Заключительный минимум может быть легко оценен сверху следующим образом:

$$\min_{\lambda \in [0,1]} \|\lambda z^{\bar{k}} + (1-\lambda) g^{t_{\bar{k}}+1}\|^2 = \min_{\lambda \in [0,1]} \{\lambda^2 \|z^{\bar{k}}\|^2 + (1-\lambda)^2 \|g^{t_{\bar{k}}+1}\|^2\} = \lambda_\star \|z^{\bar{k}}\|^2, \tag{10}$$

где

$$\lambda_\star = \frac{\|g^{t_{\bar{k}}+1}\|^2}{\|z^{\bar{k}}\|^2 + \|g^{t_{\bar{k}}+1}\|^2.}$$

решает (10).

Учитывая то, что $\|z^{\bar{k}}\|$ глобально ограничено некоторой константой $C \geqslant \|g\|, g \in \partial f(z), f(z) \leqslant f(x^0)$, а $\|g^{t_{\bar{k}}+1}\| \geqslant \gamma$, для $\lambda_\star$ легко получить следующую оценку

$$\lambda_\star \leqslant 1/(1 + \gamma^2/C^2) = \theta < 1,$$

что означает, что $\|z^k\|^2, k = \bar{k}, \bar{k}+1, \ldots$, убывает как минимум со скоростью геометрической прогрессии со знаменателем θ, и, следовательно, стремится к 0, что противоречит исходному предположению.

Это противоречие означает, что

а) либо $\{x^t\} \to x^\star$ (в этом случае теорема доказана);

б) либо $\{x^t\} \to x' \notin X^\star$, но при этом $\|z^k\| \to 0$ при $k \to \infty$;

в) либо для любой предельной точки $x' \notin X_\star$ последовательность $\{x^t\}$ покидает любую ее достаточно малую окрестность бесконечное число раз.

Из **и)** следует, что по крайней мере в этом случае условие **B2** выполнено. Для того, чтобы окончательно убедиться в том, что условие **B2** выполнено безусловно остается показать, что случай **b)** исключается.

Для этого будем считать $\epsilon > 0$ настолько малым, что множество $\tilde{G} = \mathrm{co}\{\partial f(x), \|x - x'\| \leqslant 4\epsilon\}$ строго отделимо от нуля, то есть существует вектор p, $\|p\| = 1$ и $\delta > 0$ такие, что $pg \geqslant \delta$ для всех $g \in \tilde{G}$. Поскольку для достаточно больших t точки $\|x^t - x'\| \leqslant \epsilon$, то и с момента некоторого рестарта с полным обновлением (6) и $z^r \in \tilde{G}$, то есть $pz^r \geqslant \delta > 0$, что исключает $z^r \to 0$.

Далее мы покажем, что условие **B3** также выполнено, но предварительно обеспечим выполнение условий **B4, B5**. Для этого мы определим традиционный для выпуклых задач индикатор сходимости $W(x) = \|x - x^\star\|^2$, где $x^\star$ — единственный в силу сильной выпуклости элемент множества $X_\star$. Заметим, что это автоматически обеспечивает выполнение **B4**.

Легко показать, что последовательность $\|x^{t+1} - x^t\| \to 0$ при $t \to \infty$, что даже сильнее, чем **B5**. Действительно, элементы последовательности $\{x^t\}$ связаны соотношением $x^{t+1} = x^t - \alpha_t p^t$ где шаговый множитель α_t выбран так, что существует субградиент $\bar{g}^{t+1} \in \partial f(x^{t+1}$ такой, что

$$\bar{g}^{t+1} p^t = 0 = g^{t+1}(x^{t+1} - x^t). \tag{11}$$

Поскольку $f(x^{t+1}) \leqslant f(x^t)$ то в силу ограниченности $f(x^t) \to \bar{f}$ при $t \to \infty$.

Предполагая, что $x^t \to x'$, $x^{t+1} \to x''$, $\bar{g}^{t+1} \to \bar{g}$, отметим, что $f(x') = f(x'') = \bar{f}$ и $\bar{g} \in \partial f(x'')$ в силу полунепрерывности сверху $df(x)$. Переходя в (11) к пределу и используя сильную выпуклость с некоторой константой сильной выпуклости $\sigma > 0$

$$f(x^t) - f(x^{t+1}) \geqslant \bar{g}^{t+1}(x^t - x^{t+1}) + \sigma\|x^t - x^{t+1}\|^2 = \sigma\|x^t - x^{t+1}\|^2 \geqslant 0$$

получаем

$$0 = f(x') - f(x'') = \sigma\|x' - x''\|^2 \geqslant 0$$

или $\|x^{t+1} - x^t\| \to 0$ при $t \to \infty$, что доказывает **B5.**

Переходя к доказательству выполнения условия **B3**, обозначим через $\{m_k\}$ и $\{n_k\}, k = 0, 1, \ldots$ последовательности индексов, такие, что $x^{n_k} \to x' \neq x^\star$ и существует $\epsilon > 0$ такое, что $\|x^{m_k} - x^{n_k}\| > \epsilon$ и $\|x^t - x^{n_k}\| \leqslant \epsilon$ для всех $n_k \leqslant t < m_k$. Такое можно считать сколь угодно малым с тем, чтобы выполнялись все предыдущие оценки.

По построению последовательность $\{x^{n_k}\}$ сходится к x', а последовательность $\{x^{m_k}\}$ — это последовательность первых выходов из ϵ-окрестностей соответствующих точек x^{n_k}.

Пусть $q_k < n_k$ будет максимальным индексом, не превосходящим n_k, таким, что $G(\cdot,\cdot,\cdot)$ был ре-инициализирован и p_k — минимальный индекс, превосходящий n_k когда $G(\cdot,\cdot,\cdot)$ был ре-инициализирован следующий раз. Независимо от вида рестарта $p_k - n_k \leqslant p_k - q_k \leqslant N$ и, следовательно,

$$\|x^{p_k} - x^{n_k}\| \leqslant \sum_{t=n_k}^{p_k-1} \|x^{t+1} - x^t\| \leqslant N \sup_{t \geqslant n_k} \|x^{t+1} - x^t\| \to 0$$

при $k \to \infty$.

Поэтому $x^{p_k} \to x'$ и, следовательно, $\lim_{k\to\infty} W(x^{p_k}) = \lim_{k\to\infty} W(x^{n_k}) = W(x')$. При этом доказательство того, что

$$\lim_{k\to\infty} W(x^{p_k}) > \limsup_{k\to\infty} W(x^{m_k}) \tag{12}$$

эквивалентно доказательству выполнения **B3.**

Для доказательства (12) заметим, что для всех t таких, что $p_k \leqslant t < m_k$ вектора $p^t \in \mathrm{co}\{\partial f(x), \|x - x'\| \leqslant 4\epsilon\} = \tilde{G}$ для достаточно малого $\epsilon > 0$.

Произвольное $p \in \tilde{G}$ в силу теоремы Каратеодори может быть представлено как $p = \sum_{i=1}^{n+1} \lambda_i g^i$, где $g^i \in \partial f(y^i), \|y^i - x^\star\| \leqslant 4\epsilon$, $\lambda_i \geqslant 0, \sum_{i=1}^{n+1} \lambda_i = 1$. В силу выпуклости

$0 < \gamma < f(y^i) - f(x^\star) \leqslant g^i(y^i - x^\star)$ что по непрерывности может быть преобразовано в

$$0 < \gamma/2 \leqslant g^i(x' - x^\star). \tag{13}$$

Умножая (13) на λ_i и суммируя, получаем $p(x' - x^\star) \geqslant \gamma/2$.

Рассмотрим теперь t такое, что $p_k \leqslant t < m_k$ и

$$W(x^{t+1}) - W(x^t) = \|x^t - \alpha_t p^t - x^\star\|^2 - \|x^t - x^\star\|^2 = -\alpha_t p^t(x^t - x^\star) + \alpha_t^2 \|p^t\|^2.$$

Снова по непрерывности $p^t(x^t - x^\star) \geqslant \gamma/4$ и, следовательно,

$$W(x^{t+1}) - W(x^t) \leqslant -\alpha_t \gamma/2 + \alpha_t^2 \|p^t\|^2. \tag{14}$$

Поскольку $\alpha_t \|p^t\| \to 0$, но $\|p^t\| \geqslant \kappa > 0$ ($\tilde{G}$ может быть строго отделено от 0 в силу выпуклости), то $\alpha_t \to 0$ и это означает, что мы можем игнорировать последнее слагаемое в (14). Поэтому для $p_k \leqslant t < m_k$

$$W(x^{t+1}) - W(x^t) \leqslant -\alpha_t \gamma/4,$$

что после суммирования дает

$$W(x^{m_k}) - W(x^{p_k}) \leqslant -\gamma \sum_{t=p_k}^{m_k-1} \alpha_t/4 < 0.$$

Так как

$$\epsilon/2 < \|x^{m_k} - x^{p_k}\| \leqslant \sum_{t=p_k}^{m_k-1} \alpha_t \|p^t\| \leqslant C \sum_{t=p_k}^{m_k-1} \alpha_t$$

то $\sum_{t=p_k}^{m_k-1} \alpha_t \geqslant \epsilon/2C$ и

$$W(x^{m_k}) - W(x^{p_k}) \leqslant -\gamma\epsilon/8C < 0.$$

Переход к пределу при $k \to \infty$ доказывает **B3** и, следовательно, доказывает сходимость алгоритма. Интересно отметить, что сильная выпуклость целевой функции в задаче (1) играла в доказательстве в общем-то техническую роль и может быть убрана путем незначительного усложнения как алгоритма, так и доказательства. Предположение же о конечности объема пакета N и механизм полной ре-инициализации (6) играют в доказательстве существенную роль, обеспечивая возможность "забывать" возможно неудачную предысторию поиска. Вместе с тем, полная потеря информации при (6) играет, возможно, и негативную роль, заставляя тратить определенное количество обращений к субградиентному оракулу для накопления достаточно представительного пакета G. Поиск приемлемого компромисса в этом направлении представляется интересным направлением исследований.

4. Связь с методом сопряженных градиентов

В теории квазиньютоновских алгоритмов известно, что некоторые варианты алгоритмов Бройденовского типа представляют собой коррекции квазиньютоновских матрицы, имеющие минимальную матричную норму (в частности, норму Фробениуса). Интересно отметить, что тот принцип построения направления поиска, как элемента с минимальной евклидовой нормой, который мы применили в предлагаемом алгоритме, в классическом случае сильно выпуклых квадратичных целевых функций также дает традиционный алгоритм сопряженных градиентов [21].

Действительно, пусть $\{g^1, g^2, \ldots, g^k\}$ — совокупность градиентов сильно выпуклой квадратичной целевой функции, полученных на предыдущих k итерациях, которые мы будем считать взаимно ортогональными и полученными в результате одномерной минимизации по соответствующим сопряженным направлениям $p^1, p^2, \ldots, p^k$. Соответствующая проблема P_k поиска направления поиска выглядит как

$$(15)\qquad \min_{\sum_{i=1}^k \lambda_i = 1,\ \lambda_i \geqslant 0,\ i=1,2,\ldots,k.} \tfrac{1}{2}\|\textstyle\sum_{i=1}^k \lambda_i g^i\|^2 = \min_{\sum_{i=1}^k \lambda_i = 1,\ \lambda_i \geqslant 0,\ i=1,2,\ldots,k.} \tfrac{1}{2}\textstyle\sum_{i=1}^k \lambda_i^2 \|g^i\|^2$$

Если $g^i \geqslant 0, i = 1, 2, \ldots, n$ взаимно ортогональны, то решения задач P_k и P_{k+1} представляют собой сопряженные вектора. Чтобы продемонстрировать это, рассмотрим решение P_{k+1}, которое определяется условиями оптимальности в (15)

$$(16)\qquad \lambda_i\|g^i\|^2 + \theta = 0, i = 1, 2, \ldots, k+1; \theta \sum_{i=1}^{k+1} \|g^i\|^{-2} = -1.$$

Мы игнорируем ограничения неотрицательности λ_i, поскольку они, как показано ниже, выполняются автоматически.

Из (16) следует, что

$$\lambda_j = \|g^j\|^{-2}(\sum_{i=1}^{k+1} \|g^i\|^{-2})^{-1}, j = 1, 2, \ldots, k+1 \geqslant 0$$

и, следовательно, условия неотрицательности действительно выполнены автоматически.

Обозначим

$$\sigma_{k+1} = \sum_{i=1}^{k+1} \|g^i\|^{-2} = \sigma_k + \|g^{k+1}\|^{-2}.$$

Тогда

$$p^{k+1} = \sum_{i=1}^{k+1} \lambda_i g^i = \sum_{i=1}^{k} \lambda_i g^i + \lambda_{k+1} g^{k+1} =$$

$$\sum_{i=1}^{k} \|g^i\|^{-2}(\sigma_k + \|g^{k+1}\|^{-2})^{-1} g^i + \|g^{k+1}\|^{-2}(\sigma_k + \|g^{k+1}\|^{-2})^{-1} g^{k+1} =$$

$$(\sigma_k + \|g^{k+1}\|^{-2})^{-1}(g^{k+1} + \|g^{k+1}\|^{-2} \sum_{i=1}^{k} \|g^i\|^{-2} g^i) =$$

$$\theta_{k+1}(g^{k+1} + \|g^k\|^{-2}\|g^{k+1}\|^2(\|g^k\|^2 \sum_{i=1}^{k} \|g^i\|^{-2} g^i)) =$$

$$\theta_k(g^{k+1} + \|g^k\|^{-2}\|g^{k+1}\|^2 z^k) = \theta_k(g^{k+1} + \mu_{k+1} p^k),$$

где $\mu_{k+1} = \|g^{k+1}\|^2/\|g^k\|^2$ и, следовательно, p^{k+1} с точностью до масштабирующего множителя θ_k определено по классической формуле Полака-Рибьера. Отсюда следует сопряженность p^{k+1} к $p^1, p^2, \ldots, p^k$. Поскольку масштабирующий множитель не играет никакой роли в последующей одномерной минимизации по направлению p^{k+1}, градиент g^{k+2} будет ортогонален всем предыдущим градиентам и результирующая последовательность итераций будет совпадать с методом сопряженных градиентов.

5. Численные эксперименты

Для практической демонстрации вычислительных свойств предложенного метода рассмотрим результаты численного экспериментов с хорошо известной кусочно-квадратичной тестовой функцией `maxqfg`. Эта функция имеет вид

$$f(x) = \max_{1 \leqslant k \leqslant 5} \phi_k(x),$$

где $\phi_k(x) = xA_kx - b^kx$, а $A^{(k)}, k = 1, 2, \ldots, 5$ представляют собой симметричные положительно определенные матрицы 10×10,

$$A_{ij}^{(k)} = \begin{cases} \exp(\min(i,j)/\max(i,j))\cos(ij)\sin(k), & i \neq j, \\ i|\sin(k)|/10 + \sum\limits_{l=1,2,\ldots,10, l\neq i} |A_{il}^{(k)}| & i = j, \end{cases} \qquad i, j = 1, 2, \ldots, 10,$$

$$b_i^k = \exp(i/k)\sin(ik), \quad i = 1, 2, \ldots, 10, \quad k = 1, 2, \ldots, 5.$$

Эта задача хорошо демонстрирует трудности решения кусочно-квадратичных задач выпуклой негладкой оптимизации, поскольку в ней сочетаются как проблемы негладкого характера (разрывность градиентов), так и гладкой оптимизации (овражность лебеговых множеств). Попытки решить эту проблема при помощи простых субградиентных алгоритмов встречаются с серьезными трудностями, поскольку здесь не выполняется условие острого минимума и субдифференциальное множество в оптимальной точке имеет пустую внутренность, в оптимуме активны только 3 из 5 функций $\phi_k(x)$ Это означает, что в подпространстве, ортогональном линейной оболочке субдифференциала функция ведет себя квадратично и у нас имеет место вырожденность в терминах отношения "вписанного и описанного шара" для субдифференциальных множеств. Несмотря на это метод сопряженных субградиентов с ограниченной памятью как показано на Рис. 1 демонстрирует фактическую линейную сходимость скорость сходимости для функции `maxquad`, практически недостижимую для субградиентных алгоритмов. Несмотря на все усилия, обычный субградиентный метод смог достигнуть относительную точность всего 10^{-2} лишь после 5000 итераций.

Стоит заметить, что метод сопряженных субградиентов слегка улучшил предыдущий рекорд для этой функции. С точностью до 16 значащих цифр оптимальное значение в этой задаче равно -0.8414083345821985 и достигается в точке, приведенной в Табл. 1.

Стоит обратить также внимание на горизонтальные участки графика Рис. 1, которые соответствуют нулевым решениям задачи одномерной минимизации (7). Во время этих итераций происходит фактически сбор информации для поиска направления убывания целевой функции и замечательной особенностью алгоритма является то, что количество пробных шагов практически не изменяется на всем диапазоне

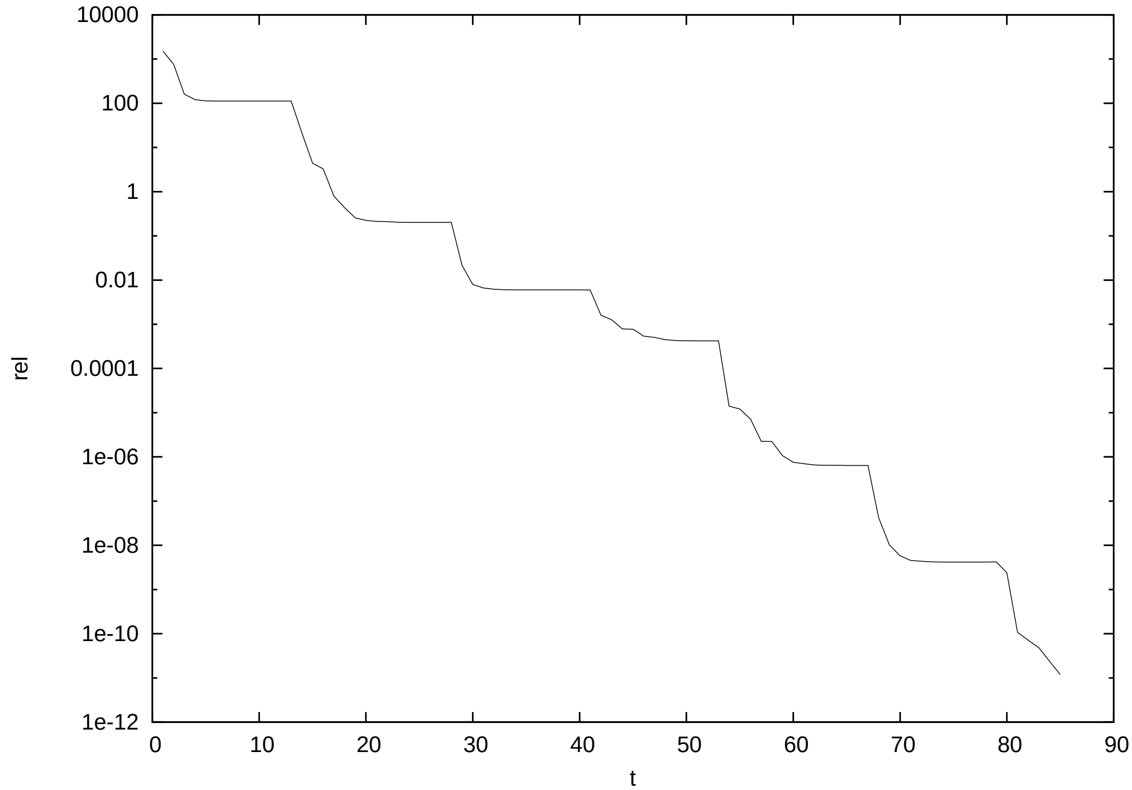

Рис. 1. Сходимость метода сопряженных субградиентов для кусочно-квадратичной функции `maxquad`. Показана зависимость относительной точности нахождения минимума $\texttt{rel} = (f(x^k) - f_\star)/|f_\star|$ от числа итераций `t` алгоритма.

Таблица 1. Точка минимума для `maxquad`

1	0.1262565919226512	6	0.2783995015309495
2	0.0343783011310847	7	-0.0742186640960634
3	0.0068571878440697	8	-0.1385240462792682
4	-0.0263606695458208	9	-0.0840312187567561
5	-0.0672949264854349	10	-0.0385803073994817

точности и не возрастает по мере приближения к экстремуму и соответствующему ухудшению обусловленности вычислительного процесса.

Заключение

В настоящей работе удалось провести полное теоретическое обоснование аналога метода сопряженных градиентов с априорно ограниченной памятью для задач выпуклой недифференцируемой оптимизации. Предложенный алгоритм лишен двух основных недостатков предложенных ранее аналогов: неограниченнных требований к памяти и/или значительного роста количества пробных итераций для поиска направления улучшения целевой функции по мере приближения к экстремуму. Вычислительный эксперимент с тестовой задачей недифференцируемой оптимизации, представляющей существенные сложности для субградиентных алгоритмов, показал вполне обнадеживающие результаты, что позволяет надеятся и на практическое применение полученных результатов.

СПИСОК ЛИТЕРАТУРЫ

1. *Шор Н.З.* Применение метода градиентного спуска для решения сетевой транспортной задачи. // Материалы научн. семинара по теорет. и прикл. вопр. кибернетики и исследования операций.- Киев: Науч. совет по кибернетике АН УССР, 1962.- Вып. 1.- С. 9-17.

2. *Поляк Б.Т.* Один общий метод решения экстремальных задач // Докл. АН СССР.-1967.- 174.-№1.- С.33-36.

3. *Wolfe Ph.* Note on a method of conjugate subgradients for minimizing nondifferentiable functions // Mathematical Programming. 1974. V. 7(1). 380–383.

4. *Wolfe Ph.* A method of conjugate subgradients for minimizing nondifferentiable functions // Nondifferentiable Optimization. Mathematical Programming Studies. 1975. V. 3. Berlin, Heidelberg: Springer. 145–173.

5. *Lemarechal C.* An extension of Davidon methods to non-differentiable problems // Mathematical Programming Study.- 3.- 1975.- 95-109.

6. *Shor N.Z., Kiwiel K.C., Ruszcayǹski A.* Minimization methods for non-differentiable functions, Springer-Verlag, New York, 1985.

7. *Brannlund U., Kiwiel K.C., Lindberg P.O.* A descent proximal level bundle method for convex nondifferentiable optimization // Operations Research Letters, 17(3), 1995, 121-126

8. *Lemarechal C., Nemirovskii A., Nesterov Ju.* New variants of bundle methods // Mathematical Programming Volume 69, Numbers 1-3, 1995, 111-147

9. *Nesterov Yu.E.* Smooth minimization of non-smooth functions // Mathematical Programming.-2005.-103.-1.-127-152

10. *Mifflin R., Sagastizábal C.* A VU-algorithm for convex minimization // Mathematical Programming Volume 104, Numbers 2-3, 2005, 583-608

11. *Коннов И. В.* Метод типа сопряженных субградиентов для минимизации функционалов // Исслед. по прикл. матем..-12.-1984.-59–62

12. *Konnov I.V.* Combined Relaxation Methods for Variational Inequalities, Springer-Verlag, Berlin, Heidelberg, New York, 2001, ISBN 3-540-67999-5, 181 pp.

13. *Pytlak R.* On the convergence of conjugate gradient algorithms // IMA J. Numer. Anal. 14, 1994, 443-460

14. *Pytlak R., Tarnawski T.* On the method of shortest residuals // J. Optim. Theory Appl. 133, 2007, 99-110

15. *Dai Y.H., Yuan Y.* Global convergence of the method of shortest residuals // Numer. Math. 82, 1999, 581-598

16. *Dai Y.H.* Convergence of conjugate gradient methods with constant stepsizes // Optimization Methods & Software Vol. 26, No. 6, December 2011, 895-909

17. *Hiriart-Urruty J.-B., Lemarechal C.* Convex analysis and minimization algorithms II. Advanced theory and bundle methods Springer-Verlag, 1993.

18. *Нурминский Е.А.* Численные методы выпуклой оптимизации М.: Наука, 1991.-168 с.

19. *Nurminski E.A.* Envelope step-size control for iterative algorithms based on Fejer processes with attractants // Optimization Methods and Software, 1029-4937, v. 25(1), 2010, 97–108.

20. *Нурминский Е.А.* Фейеровские алгоритмы с адаптивным шагом // Журн. вычисл. матем. и матем. физикиі .- т.51, вып. 5.- 2011.-791-801

21. *Hesten M.R., Stiefel E.* Methods of conjugate gradients for solving linear systems // Journal of Research of the National Bureau of Standards 49 (6), 1952, Research Paper 2379 409–436.

22. *Lemarechal C., Mifflin R.* Nonsmooth optimization // Oxford.: Pergamon Press. – 1978. – 186 p.

23. *Lemarechal C. Numerical experiments in nonsmooth optimization* // In: Progress VPN nondifferentiable optimization /Ed. E. A. Nurminski. CP-82-58. International Institute for Applied System Analysis: Laxenburg, Austria, 1982. – P. 61-84.