
ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 517.23

БАД-МЕТОДОЛОГИЯ И ДИФФЕРЕНЦИРОВАНИЕ СЛОЖНОЙ ФУНКЦИИ¹⁾

© 2023 г. А. Ф. Албу¹, А. Ю. Горчаков^{1,2}, В. И. Zubov^{1,*}

¹ 119333 Москва, ул. Вавилова, 40, ФИЦ ИУ РАН, Россия

² 141700 М. о., Долгопрудный, Институтский пер., 9, МФТИ, Россия

*e-mail: vladimir.zubov@mail.ru

Поступила в редакцию 27.04.2022 г.
Переработанный вариант 27.04.2022 г.
Принята к публикации 17.09.2022 г.

Сравниваются разные подходы к вычислению градиента сложной функции многих переменных, такие как использование точных, аналитически выведенных формул; использование формул, полученных с помощью методологии быстрого автоматического дифференцирования; использование стандартных программных пакетов, реализующих идеи методологии быстрого автоматического дифференцирования. Сравнение подходов осуществляется на примере сложной функции, представляющей энергию системы атомов, потенциал взаимодействия которых – потенциал Терсоффа. В качестве критерия сравнения используется компьютерное время, необходимое для вычисления градиента функции. Результаты показывают превосходство методологии быстрого автоматического дифференцирования по сравнению с подходом, использующим аналитические формулы. Стандартные пакеты вычисляют градиент функции примерно за то же время, что и при использовании формул методологии быстрого автоматического дифференцирования. Библ. 15. Фиг. 1. Табл. 5.

Ключевые слова: быстрое автоматическое дифференцирование, многошаговый процесс, пакеты стандартных программ.

DOI: 10.31857/S0044466923010039, EDN: LEJFIO

ВВЕДЕНИЕ

В настоящее время методы математического моделирования и вычислительной математики все глубже проникают в различные области науки, техники и производства. Они занимают особое место среди современных методов решения прикладных задач. В связи с интенсификацией производства в последние десятилетия особенно возрос интерес к проблемам оптимизации, большую актуальность приобрели задачи оптимального управления сложными динамическими системами.

Можно выделить несколько различных направлений в разработке численных методов решения задач оптимального управления, которые существенно отличаются друг от друга. Прежде всего следует упомянуть прямые методы, основанные на спуске в пространстве управлений. Ряд исследований связан с непрямыми методами, в которых с помощью принципа максимума Л.С. Понтрягина исходная задача оптимального управления сводится к краевой (см. [1]). В [2] Р.П. Федоренко предложил подход, который базируется на использовании идей метода линеаризации. Большой цикл работ по численным методам оптимального управления был выполнен Н.Н. Моисеевым (см. [3]) и его учениками.

Другой подход численного решения задач оптимального управления базируется на идеях нелинейного программирования: задачи оптимального управления с условиями на управляющую функцию и с фазовыми ограничениями сводятся к задачам нелинейного программирования. Такой подход оказался чрезвычайно эффективным. Обычно такие задачи решаются градиентными методами минимизации целевого функционала. Но для этого необходимо уметь точно вычислять градиент целевой функции.

¹⁾ Работа выполнена при частичной финансовой поддержке РФФИ (проекта № 21-71-30005).

Простейший способ вычисления градиента достаточно сложной функции — определение компонент градиента с помощью метода конечных разностей. Однако этот подход достаточно трудоемок и не обеспечивает получение точного градиента. В [4] раскрыты проблемы, возникающие при использовании метода конечных разностей для вычисления градиента целевой функции в задачах оптимального управления сложными системами.

Второй способ вычисления градиента функционала — использование сопряженной задачи. Здесь представляется естественным сформулировать исходную задачу оптимального управления в “непрерывном” пространстве, сформулировать сопряженную задачу также в “непрерывном” пространстве, определить градиент функционала в “непрерывном” случае, и только после этого осуществлять аппроксимацию функционала, прямой и сопряженной задач. В результате этих действий получают формулы для определения градиента целевого функционала в дискретном варианте. Однако здесь необходимо учитывать то, что получаемый дискретный градиент функционала неточен. Его значение зависит от способа дискретизации как прямой и сопряженной задач, так и от способа дискретизации самого градиента. Эта неточность может быть уменьшена, если уменьшить погрешности указанных аппроксимаций, или даже полностью ликвидирована за счет выбора “согласованных” аппроксимаций прямой и сопряженной задач и “согласованной” с ними аппроксимации градиента. Эти “согласованные” аппроксимации можно случайно угадать, но указать способ определения “согласованных” аппроксимаций, используя данный подход, по-видимому, невозможно. Если выбрать неправильный дискретный вариант сопряженной задачи, получается неверное значение градиента. С такими случаями приходилось встречаться на практике (см., например, [5]).

Другой путь связан с формулировкой рассматриваемой задачи оптимального управления в конечномерном пространстве и с определением градиента функционала сразу в дискретном виде. Если дискретный градиент определять с помощью метода множителей Лагранжа (сопряженных переменных), то такой путь известен как обобщенная методология быстрого автоматического дифференцирования (БАД-методология).

БАД-методология значительно превосходит по эффективности аналитическое дифференцирование (точнее, использование аналитически полученных формул) и вычисление производных с помощью метода конечных разностей. В работах Ю.Г. Евтушенко [6–8] предложен общий подход к дифференцированию сложных функций. Там показано, что БАД-методология дает возможность с единой позиции рассмотреть самые разнообразные задачи. Например, из общих формул дифференцирования, приведенных в этих работах, несложно получить формулы БАД-методологии для определения градиента функций многих переменных. При этом вычисление значения функции представляется как многошаговый процесс, в котором введены новые фазовые переменные. Эти фазовые переменные являются функциями независимых переменных, по которым ищутся производные заданной функции. Показано, что при определенных условиях использование БАД-методологии позволяет получить точное значение градиента дискретной задачи оптимального управления (дискретного целевого функционала при наличии дискретных связей) за время, не превосходящее утроенного времени вычисления самой функции.

БАД-методология эффективно использовалась при решении многих теоретически сложных прикладных задач, имеющих большое значение для народного хозяйства. Например, она была применена при решении задач из области материаловедения, при исследовании задачи моделирования и управления процессом сварки материалов, задачи управления процессом кристаллизации металла в литейном деле.

В настоящей работе проводится сравнение разных подходов к вычислению градиента сложной функции многих переменных. Сравнимаемыми подходами являются: использование точных, аналитически выведенных формул; использование формул, полученных с помощью БАД-методологии; использование стандартных программных пакетов, реализующих идеи БАД-методологии. Сравнение подходов осуществляется на примере сложной функции, представляющей энергию системы атомов, потенциал взаимодействия которых — потенциал Терсоффа.

1. СРАВНЕНИЕ РАЗЛИЧНЫХ МЕТОДОВ ДИФФЕРЕНЦИРОВАНИЯ ФУНКЦИЙ

Функция $f(u)$ называется элементарной функцией, если она может быть представлена как конечная суперпозиция основных элементарных функций и арифметических действий. Будем считать, что дифференцируемая скалярная функция $f(u)$ векторного аргумента $u \in R^r$ задана явно. Задача состоит в вычислении частных производных функции $f(u)$ по компонентам век-

тора u . В вычислительной практике часто встречаются функции, для которых точное вычисление частных производных представляет собой непростую задачу.

В таком случае можно воспользоваться несколькими способами нахождения частных производных функции $f(u)$:

1. *Приближенное численное дифференцирование.* Если $f(u)$ — функция r -мерного вектора u , то при использовании самой простой формулы численного дифференцирования необходимо по меньшей мере $r + 1$ раз вычислить значение самой функции $f(u)$, и r раз произвести вычитание и деление. Результаты таких расчетов будут точными только для линейных функций; во всех остальных случаях градиент определяется с некоторой погрешностью. Ниже будет показано, что при вычислении частных производных энергии группы атомов (потенциал взаимодействия атомов — потенциал Терсоффа) по специфичным параметрам потенциала Терсоффа с помощью этого метода необходимо проводить дополнительные исследования, связанные с выбором приращения каждого параметра потенциала, причем при каждом его новом значении.

2. *Аналитическое дифференцирование.* Здесь имеется в виду использование выведенных аналитических выражений для вычисления частных производных функции $f(u)$. Следует также отметить, что в современной компьютерной литературе такой подход нередко называют прямым методом автоматического дифференцирования, или аналитическим дифференцированием. Если вектор u — вектор большой размерности, то при применении этой техники возникают большие трудности.

3. *Использование формул БАД (БАД-методология)* (см. [6]–[8]). При использовании этого метода вычисление значения функции $f(u)$ записывается в виде многошагового процесса в каноническом виде, формулируется сопряженная задача в каноническом виде и с их помощью выводятся аналитические выражения для вычисления частных производных рассматриваемой функции.

4. *Использование стандартных пакетов программ автоматического дифференцирования.*

а) Используются пакеты программ для генерации кода вычисления производных по коду вычисления функции (source-to-source AD tool). В качестве примера такой программы можно привести TAPENADE (см. [9]).

б) Используются пакеты программ, перегружающих операторы в коде вычисления функции. В результате такой перезагрузки код вычисляет не только значение функции, но и значение ее производных (Operator Overloading AD tool). В качестве примера такой программы можно привести CoDiPack и Adept (см. [10], [11]).

Ниже в этом разделе приводится сравнение дифференцирования сложной функции с помощью формул БАД-методологии (способ 3) и с помощью аналитически полученных выражений для ее частных производных (способ 2).

Для того чтобы воспользоваться техникой БАД (способ 3), необходимо представить вычисление элементарной функции $f(u)$ как многошаговый процесс. С этой целью введем вектор $z \in R^k$ фазовых переменных (вектор вспомогательных, промежуточных значений функции):

$$z_1 = F(1, Z_1, U_1), \quad z_2 = F(2, Z_2, U_2), \quad \dots, \quad z_k = F(k, Z_k, U_k), \quad (1.1)$$

где Z_i — набор элементов z_j , которые входят в правую часть равенства $z_i = F(i, Z_i, U_i)$, а U_i — набор элементов u_j , которые входят в правую часть этого равенства. Последовательность (1.1) строится таким образом, чтобы процесс был явным, и последняя компонента z_k совпала бы со значением функции $f(u)$, т.е. $z_k = f(u)$.

Предполагаем, что все функции $F(i, Z_i, U_i)$ дифференцируемы по всем компонентам векторов z и u . Введем вектор сопряженных переменных $p \in R^k$, который определим как решение следующей системы линейных алгебраических уравнений:

$$p_l = \sum_{q \in \bar{Q}_l} F_{z_q}^T(q, Z_q, U_q) p_q, \quad \bar{Q}_l = \{i: z_l \in Z_i\}, \quad l = \overline{1, k}. \quad (1.2)$$

Тогда, согласно формулам БАД-методологии (см. [8]), компоненты градиента функции многих переменных $f(u)$ вычисляются по формуле

$$\frac{\partial f}{\partial u_m} = \sum_{q \in \bar{K}_m} F_{u_m}^T(q, Z_q, U_q) p_q, \quad \bar{K}_m = \{i: u_m \in U_i\}, \quad m = \overline{1, r}. \quad (1.3)$$

Таким образом, мы сначала последовательно выполняем вычисления в соответствии с k -шаговым процессом и запоминаем значения переменных z и u (прямой ход), а потом идем в обратную сторону и рассчитываем сопряженные переменные p (обратный ход).

Что касается вычисления градиента функции методом 2, то здесь сопряженные переменные не определяются и не используются. Учитывая тот факт, что сложная функция $f(u)$ представляет собой суперпозицию более простых функций, представим вычисление ее значения в виде k -шагового алгоритма (1.1).

Тогда производные $\frac{\partial f}{\partial u_m}$, $m = \overline{1, r}$, вычисляются аналитически последовательно следующим образом:

$$\frac{\partial z_1}{\partial u_m}, \quad \frac{\partial z_2}{\partial u_m} = \frac{\partial z_2}{\partial z_1} \frac{\partial z_1}{\partial u_m}, \quad \frac{\partial z_3}{\partial u_m} = \frac{\partial z_3}{\partial z_2} \frac{\partial z_2}{\partial u_m}, \quad \dots, \quad \frac{\partial z_k}{\partial u_m} = \frac{\partial z_k}{\partial z_{k-1}} \frac{\partial z_{k-1}}{\partial u_m} = \frac{\partial f}{\partial u_m}, \quad m = \overline{1, r}. \quad (1.4)$$

Работу с формулами (1.2), (1.3) и (1.4) мы вначале проиллюстрируем на простом примере. Пусть задана элементарная функция $f(u) = f(u_1, u_2) = e^{u_1 + (u_2)^2} + \sin((u_1)^3 + u_2)^2$. Для применения формул БАД-методологии (метод 3) строим многошаговый процесс ее вычисления следующим образом:

$$z_1 = u_1 + (u_2)^2, \quad z_2 = e^{z_1}, \quad z_3 = (u_1)^3 + u_2, \quad z_4 = (z_3)^2, \quad z_5 = \sin(z_4), \quad z_6 = z_2 + z_5. \quad (1.5)$$

Используя формулы (1.2) и (1.3), имеем

$$p_1 = e^{z_1} p_2, \quad p_2 = p_6, \quad p_3 = 2z_3 p_4, \quad p_4 = \cos(z_4) p_5, \quad p_5 = p_6, \quad p_6 = 1, \quad (1.6)$$

$$\frac{\partial f}{\partial u_1} = p_1 + 3(u_1)^2 p_3, \quad \frac{\partial f}{\partial u_2} = 2u_2 p_1 + p_3. \quad (1.7)$$

Сопряженные переменные вычисляются в обратном порядке: $p_6, p_5, \dots, p_1$.

Для того чтобы получить частные производные этой функции без использования сопряженной задачи (с помощью формул (1.4), метод 2), введем обозначения $\tilde{z}_i = \frac{\partial z_i}{\partial u_1}$, $\tilde{\tilde{z}}_i = \frac{\partial z_i}{\partial u_2}$. Значения $\tilde{z}_i$ и $\tilde{\tilde{z}}_i$ вычисляются по формулам

$$\tilde{z}_1 = 1, \quad \tilde{z}_2 = \tilde{z}_1 e^{z_1}, \quad \tilde{z}_3 = 3(u_1)^2, \quad \tilde{z}_4 = 2z_3 \tilde{z}_3, \quad \tilde{z}_5 = \tilde{z}_4 \cos z_4, \quad \tilde{z}_6 = \tilde{z}_2 + \tilde{z}_5, \quad (1.8)$$

$$\tilde{\tilde{z}}_1 = 2u_2, \quad \tilde{\tilde{z}}_2 = \tilde{\tilde{z}}_1 e^{z_1}, \quad \tilde{\tilde{z}}_3 = 1, \quad \tilde{\tilde{z}}_4 = 2z_3 \tilde{\tilde{z}}_3, \quad \tilde{\tilde{z}}_5 = \tilde{\tilde{z}}_4 \cos z_4, \quad \tilde{\tilde{z}}_6 = \tilde{\tilde{z}}_2 + \tilde{\tilde{z}}_5. \quad (1.9)$$

Тогда $\frac{\partial f}{\partial u_1} = \tilde{z}_6$, $\frac{\partial f}{\partial u_2} = \tilde{\tilde{z}}_6$.

В случае метода 2 вместо одного многошагового процесса (1.6), который требуется для вычисления сопряженных переменных, необходимо выполнить вычисления по $r = 2$ (количество независимых переменных) многошаговым процессам: (1.8) и (1.9). Следовательно, время, необходимое для вычисления градиента с помощью формул БАД, примерно в r раз меньше времени, которое необходимо для вычисления этого градиента с помощью формул (1.4). Эффективность использования техники БАД возрастает, если растет количество независимых переменных.

Вычислительные эксперименты, касающиеся сравнения способов определения производных функции многих переменных, были также проведены и для достаточно сложной и практически важной функции, которая определяет энергию группы атомов, потенциал взаимодействия кото-

рых — потенциал Терсоффа. В этом случае энергия E взаимодействия атомов некоторой кристаллической решетки вычисляется по формуле

$$E = \sum_{i=1}^I \sum_{\substack{j=1 \\ j \neq i}}^I V_{ij},$$

где

$$V_{ij} = f_c(r_{ij})(V_R(r_{ij}) - b_{ij}V_A(r_{ij})),$$

$$f_c(r) = \begin{cases} 1, & r < R - R_{\text{cut}} = R_m, \\ \frac{1}{2} \left(1 - \sin \left(\frac{\pi(r - R)}{2R_{\text{cut}}} \right) \right), & R_m < r < R_p, \\ 0, & r > R + R_{\text{cut}} = R_p, \end{cases}$$

$$V_{ij}^R = V_R(r_{ij}) = \frac{D_e}{S-1} \exp(-\beta \sqrt{2S}(r_{ij} - r_e)), \quad V_{ij}^A = V_A(r_{ij}) = \frac{SD_e}{S-1} \exp\left(-\beta \sqrt{\frac{2}{S}}(r_{ij} - r_e)\right),$$

$$b_{ij} = \left(1 + (\gamma \varsigma_{ij})^\eta\right)^{-\frac{1}{2\eta}}, \quad \varsigma_{ij} = \sum_{\substack{k=1 \\ k \neq i, j}}^I f_c(r_{ik}) g_{ijk} \omega_{ijk}, \quad \omega_{ijk} = \exp(\lambda^3 \tau_{ijk}),$$

$$\tau_{ijk} = (r_{ij} - r_{ik})^3, \quad g_{ijk} = g(\theta_{ijk}) = 1 + \left(\frac{c}{d}\right)^2 - \frac{c^2}{d^2 + (h - \cos \theta_{ijk})^2}.$$

Здесь I — количество атомов в рассматриваемой группе; R и R_{cut} — параметры, определяющиеся из полученных экспериментально геометрических характеристик вещества; $x_i = (x_{1i}, x_{2i}, x_{3i})$ — координаты i -го атома; $r_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2}$ — расстояние между атомами с номерами i и j ; θ_{ijk} — угол, образованный векторами, соединяющими атом i с атомами j и k соответственно. Косинус этого угла вычисляется по формуле $q_{ijk} = \cos(\theta_{ijk}) = \frac{r_{ij}^2 + r_{ik}^2 - r_{jk}^2}{2r_{ij}r_{ik}}$.

Потенциал Терсоффа зависит от десяти параметров ($m = 10$), специфичных для моделируемого вещества ($D_e, r_e, \beta, S, \eta, \gamma, \lambda, c, d, h$), и от координат атомов $x = (x_1, x_2, \dots, x_I)^T$. Указанные специфические параметры вещества, как правило, неизвестны. Для решения задач параметрической идентификации потенциалов межатомного взаимодействия в последнее время все большую актуальность приобретает применение различных методов оптимизации. В связи с этим возникает необходимость определения точного значения градиента функции энергии E по специфическим для моделируемого вещества параметрам потенциала Терсоффа. Формулы для вычисления компонент этого градиента были получены в [12] с помощью БАД-методологии.

Оптимизационные задачи решаются при определенном, фиксированном положении базисных атомов рассматриваемой кристаллической структуры. Решив задачу идентификации параметров, нет уверенности, что положения базисных атомов будут соответствовать минимуму потенциальной энергии системы. Поэтому следующим шагом исследований является оптимизация по координатам частиц, расставляющая частицы в положения, соответствующие минимуму суммарной потенциальной энергии рассматриваемой системы атомов. На этом этапе возникает необходимость вычисления градиента функции энергии по координатам атомов. Формулы для вычисления этого градиента также были получены в [12] на основе формул БАД-методологии (метод 3).

Ниже мы построим вычислительный процесс для определения указанного выше градиента с помощью метода 2 (аналитического дифференцирования) и сравним его с тем, который получен с помощью метода 3.

Функция энергии E представляет собой достаточно сложную суперпозицию более простых функций. Вычисление самой функции было представлено в [12] в виде некоторого многошагового процесса.

Пусть $\bar{u}$ и $\bar{z}$ – векторы, имеющие координаты

$$\bar{u}^T = [u_1, u_2, \dots, u_{10}]^T, \quad \bar{z}^T = [z_1, z_2, \dots, z_{17}]^T,$$

где

$$\begin{aligned} u_1 &= D_e, \quad u_2 = r_e, \quad u_3 = \beta, \quad u_4 = S, \quad u_5 = \eta, \\ u_6 &= \gamma, \quad u_7 = \lambda, \quad u_8 = c, \quad u_9 = d, \quad u_{10} = h; \\ z_1 &= \left\{ z_1^{ijk} = \sqrt{(x_{1i} - x_{1k})^2 + (x_{2i} - x_{2k})^2 + (x_{3i} - x_{3k})^2} \right\}, \\ z_2 &= \left\{ z_2^{ijk} = \sqrt{(x_{1j} - x_{1k})^2 + (x_{2j} - x_{2k})^2 + (x_{3j} - x_{3k})^2} \right\}, \\ z_3 &= \left\{ z_3^{ijk} = q_{ijk} = \frac{(z_{13}^{ij})^2 + (z_1^{ijk})^2 - (z_2^{ijk})^2}{2z_{13}^{ij}z_1^{ijk}} \right\}, \\ z_4 &= \left\{ z_4^{ijk} = f_c(z_1^{ijk}) \right\}, \\ z_5 &= \left\{ z_5^{ijk} = g_{ijk} = 1 + \left(\frac{u_8}{u_9} \right)^2 - \frac{(u_8)^2}{(u_9)^2 + (u_{10} - z_3^{ijk})^2} \right\}, \\ z_6 &= \left\{ z_6^{ijk} = \tau_{ijk} = (z_{13}^{ij} - z_1^{ijk})^3 \right\}, \\ z_7 &= \left\{ z_7^{ijk} = \omega_{ijk} = \exp((u_7)^3 z_6^{ijk}) \right\}, \\ z_8 &= \left\{ z_8^{ijk} = f_c(r_{ik})g_{ijk}\omega_{ijk} = z_4^{ijk}z_5^{ijk}z_7^{ijk} \right\}, \\ z_9 &= \left\{ z_9^{ij} = \varsigma_{ij} = \sum_{k=1, k \neq i, j}^I z_8^{ijk} \right\}, \\ z_{10} &= \left\{ z_{10}^{ij} = \gamma\varsigma_{ij} = u_6 z_9^{ij} \right\}, \\ z_{11} &= \left\{ z_{11}^{ij} = (\gamma\varsigma_{ij})^\eta = (z_{10}^{ij})^{u_5} \right\}, \\ z_{12} &= \left\{ z_{12}^{ij} = b_{ij} = (1 + z_{11}^{ij})^{-\frac{1}{2u_5}} \right\}, \\ z_{13} &= \left\{ z_{13}^{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2} \right\}, \\ z_{14} &= \left\{ z_{14}^{ij} = V_{ij}^R = \frac{u_1}{u_4 - 1} \exp(-u_3 \sqrt{2u_4}(z_{13}^{ij} - u_2)) \right\}, \\ z_{15} &= \left\{ z_{15}^{ij} = V_{ij}^A = \frac{u_1 u_4}{u_4 - 1} \exp\left(-u_3 \sqrt{\frac{2}{u_4}}(z_{13}^{ij} - u_2)\right) \right\}, \\ z_{16} &= \left\{ z_{16}^{ij} = f_c(z_{13}^{ij}) \right\}, \\ z_{17} &= \left\{ z_{17}^{ij} = V_{ij} = z_{16}^{ij}(z_{14}^{ij} - z_{12}^{ij} z_{15}^{ij}) \right\}. \end{aligned}$$

Каждая компонента z_l вектора $\bar{z}$ состоит из множества других компонент, а процесс вычисления функции энергии E строится следующим образом. Для всех фиксированных пар индексов (i, j) , таких что $i = \overline{1, I}$, $j = \overline{1, I}$ и $j \neq i$, вычисляются значения z_{13}^{ij} . Далее для всех таких пар индексов (i, j) и для всех индексов $k = \overline{1, I}$, таких что $k \neq i$ и $k \neq j$, вычисляются значения $z_1^{ijk}, z_2^{ijk}, \dots, z_8^{ijk}$. Компоненты z_1^{ijk} используются для вычисления компонент z_3^{ijk} . При этом набор

индексов k , которые используются для вычисления компонент z_1^{ijk} , зависит от конкретного индекса j компоненты z_3^{ijk} . Из этих соображений мы присваиваем компонентам z_1^{ijk} и индекс j , хотя явно они от него не зависят. Аналогично, и компоненты z_2^{ijk} не зависят от индекса j , но мы им присваиваем этот индекс.

Далее для всех фиксированных пар индексов (i, j) , таких что $i = \overline{1, I}$, $j = \overline{1, I}$ и $j \neq i$, вычисляем компоненты $z_9^{ij}, z_{10}^{ij}, z_{11}^{ij}, z_{12}^{ij}, z_{14}^{ij}, z_{15}^{ij}, z_{16}^{ij}, z_{17}^{ij}$. Компоненты z_{13}^{ij} также не зависят от третьего индекса (k) , но они используются для вычисления компонент z_3^{ijk} , поэтому вычисляются раньше.

Наконец, энергия системы атомов вычисляется по формуле

$$E = E(u, x) = \sum_{i=1}^I \sum_{\substack{j=1 \\ j \neq i}}^I z_{17}^{ij}, \quad (1.10)$$

где фазовые переменные $z_1, z_2, \dots, z_{17}$ определяются указанным выше многошаговым процессом.

Для получения аналитических выражений производных функции E по координатам атомов воспользуемся структурой функции E , которая представлена формулой (1.10). Для всех фиксированных пар индексов (l, m) , таких что $l = 1, 2, 3$ и $m = \overline{1, I}$, введем следующие обозначения:

$\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_{17}$, где

$$\tilde{z}_s = \left\{ \tilde{z}_s^{ijk} : \tilde{z}_s^{ijk} = \frac{\partial z_s^{ijk}}{\partial x_{lm}} \right\}, \quad s = \overline{1, 8}, \quad \tilde{z}_s = \left\{ \tilde{z}_s^{ij} : \tilde{z}_s^{ij} = \frac{\partial z_s^{ij}}{\partial x_{lm}} \right\}, \quad s = \overline{9, 17}.$$

Указанные выше величины вычисляются по следующим формулам:

— для всех $l = 1, 2, 3$ и $i, j, k, m = \overline{1, I}$

$$\begin{aligned} \tilde{z}_1^{ijk} = \frac{\partial z_1^{ijk}}{\partial x_{lm}} &= \begin{cases} \frac{x_{lm} - x_{lk}}{z_1^{mjk}}, & m = i, \quad k \neq m, \\ \frac{x_{lm} - x_{li}}{z_1^{ijm}}, & m = k, \quad i \neq m, \\ 0 & \text{в других случаях,} \end{cases} \\ \tilde{z}_2^{ijk} = \frac{\partial z_2^{ijk}}{\partial x_{lm}} &= \begin{cases} \frac{x_{lm} - x_{lk}}{z_2^{imk}}, & m = j, \quad k \neq m, \\ \frac{x_{lm} - x_{lj}}{z_2^{ijm}}, & m = k, \quad j \neq m, \\ 0 & \text{в других случаях,} \end{cases} \\ \tilde{z}_3^{ijk} &= \frac{z_1^{ijk} (z_{13}^{ij})^2 \tilde{z}_{13}^{ij} + z_{13}^{ij} (z_1^{ijk})^2 \tilde{z}_1^{ijk} - 2z_1^{ijk} z_{13}^{ij} z_2^{ijk} \tilde{z}_2^{ijk} +}{2(z_{13}^{ij})^2 (z_1^{ijk})^2} + \\ &+ \frac{-(z_1^{ijk})^3 \tilde{z}_{13}^{ij} + z_1^{ijk} (z_2^{ijk})^2 \tilde{z}_{13}^{ij} - (z_{13}^{ij})^3 \tilde{z}_1^{ijk} + z_{13}^{ij} (z_2^{ijk})^2 \tilde{z}_1^{ijk}}{2(z_{13}^{ij})^2 (z_1^{ijk})^2}; \\ \tilde{z}_4^{ijk} &= \begin{cases} 0 & \text{в других случаях,} \\ -\pi \cos \left(\frac{\pi(z_1^{ijk} - R)}{2R_{\text{cut}}} \right) \frac{1}{4R_{\text{cut}}}, & R - R_{\text{cut}} \leq z_1^{ijk} \leq R + R_{\text{cut}}, \end{cases} \end{aligned}$$

$$\tilde{z}_5^{ijk} = \frac{-2(u_8)^2 (u_{10} - z_3^{ijk}) \tilde{z}_3^{ijk}}{\left((u_9)^2 + (u_{10} - z_3^{ijk})^2\right)^2}; \quad \tilde{z}_6^{ijk} = 3(z_{13}^{ij} - z_1^{ijk})^2 (\tilde{z}_{13}^{ij} - \tilde{z}_1^{ijk});$$

$$\tilde{z}_7^{ijk} = (u_7)^3 \exp\left((u_7)^3 z_6^{ijk}\right) \tilde{z}_6^{ijk}; \quad \tilde{z}_8^{ijk} = \tilde{z}_4^{ijk} \tilde{z}_5^{ijk} \tilde{z}_7^{ijk} + \tilde{z}_4^{ijk} \tilde{z}_5^{ijk} \tilde{z}_7^{ijk} + \tilde{z}_4^{ijk} \tilde{z}_5^{ijk} \tilde{z}_7^{ijk};$$

$$\tilde{z}_9^{ij} = \sum_{\substack{k=1 \\ k \neq i, j}}^I \tilde{z}_8^{ijk}; \quad \tilde{z}_{10}^{ij} = \tilde{z}_9^{ij} u_6; \quad \tilde{z}_{11}^{ij} = \tilde{z}_{10}^{ij} u_5 (z_{10}^{ij})^{u_5-1}; \quad \tilde{z}_{12}^{ij} = -\frac{1}{2u_5} \tilde{z}_{11}^{ij} (1 + z_{11}^{ij})^{-\frac{1}{2u_5}-1};$$

— для всех $l = 1, 2, 3$ и $i, j, m = \overline{1, I}$

$$\tilde{z}_{13}^{ij} = \frac{\partial z_{13}^{ij}}{\partial x_{lm}} = \begin{cases} \frac{x_{lm} - x_{lj}}{z_{13}^{mj}}, & m = i, \quad j \neq m, \\ \frac{x_{lm} - x_{li}}{z_{13}^{im}}, & m = j, \quad i \neq m; \\ \tilde{z}_{14}^{ij} = -u_3 \sqrt{2u_4 z_{14}^{ij} \tilde{z}_{13}^{ij}}, & 0 \text{ в других случаях,} \end{cases}$$

$$\tilde{z}_{15}^{ij} = -u_3 \sqrt{2/u_4} z_{15}^{ij} \tilde{z}_{13}^{ij};$$

$$\tilde{z}_{16}^{ij} = \begin{cases} \frac{-\pi \cos\left(\frac{\pi(z_{13}^{ij} - R)}{2R_{\text{cut}}}\right)}{4R_{\text{cut}}}, & R - R_{\text{cut}} \leq z_{13}^{ij} \leq R + R_{\text{cut}}, \\ 0 & \text{в других случаях.} \end{cases}$$

$$\tilde{z}_{17}^{ij} = \tilde{z}_{16}^{ij} z_{14}^{ij} - \tilde{z}_{16}^{ij} z_{12}^{ij} z_{15}^{ij} + z_{16}^{ij} \tilde{z}_{14}^{ij} - z_{16}^{ij} \tilde{z}_{12}^{ij} z_{15}^{ij} - z_{16}^{ij} z_{12}^{ij} \tilde{z}_{15}^{ij}.$$

Порядок вычисления этих производных такой же, как и для фазовых переменных $z_1, z_2, \dots, z_{17}$, а формулы для вычисления частных производных функции энергии по координатам атомов имеют вид

$$\frac{\partial E(u, x)}{\partial x_{lm}} = \sum_{i=1}^I \sum_{\substack{j=1 \\ j \neq i}}^I \tilde{z}_{17}^{ij}, \quad l = 1, 2, 3 \text{ и } m = \overline{1, I}.$$

Проведена большая серия вычислительных экспериментов, в которых менялись размерность материала (рассматривались трехмерная и двумерная модели) и количество атомов в фрагменте материала. Если рассматривается трехмерная модель ($l = 1, 2, 3$), то для каждого m -го атома вычисления по многошаговому процессу $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_{17}$ необходимо проводить 3 раза, а если рассматривается двумерная модель, то 2 раза. Все численные эксперименты показали, что в первом случае (трехмерная модель) градиент функции энергии по координатам атома, определенный с помощью формул БАД-методологии (см. [12]), вычисляется в 3 раза быстрее, чем с помощью многошагового процесса $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_{17}$. Когда используется двумерная модель материала, эффективность формул БАД-методологии возрастает в 2 раза.

Возникает естественный вопрос: не проще ли при вычислении производных этой сложной функции воспользоваться приближенным численным дифференцированием? Ниже показывается, с какими трудностями приходится сталкиваться при использовании такого метода.

В [13] рассмотрена задача идентификации параметров потенциала Терсоффа применительно к однокомпонентному кристаллу кремния. Для определения начальных приближений в указанной работе допустимое множество $X = [u, \bar{u}] = \{u \in R^{10} : \underline{u}_i \leq x_i \leq \bar{u}_i\}$ выбиралось таким образом, чтобы заведомо содержались все допустимые значения параметров, а именно:

$$\underline{u} = (0.5; 0.5; 0.5; 0.5; 0.1; 5 \times 10^{-8}; 0.5; 10000; 1; -2);$$

$$\bar{u} = (10; 5; 5; 5; 2; 3 \times 10^{-6}; 3; 200000; 30; -0.1).$$

Таблица 1. ($i = 1, u_1 = 1$)

Δ	10^{+0}	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}	10^{-13}
Абсолютное отклонение	1.0×10^{-14}	1.0×10^{-14}	5.1×10^{-11}	4.6×10^{-9}	1.3×10^{-7}	8.9×10^{-6}	9.7×10^{-4}	4.4×10^{-2}

Таблица 2. ($i = 3, u_3 = 1$)

Δ	10^{+0}	10^{-2}	10^{-4}	10^{-6}	10^{-7}	10^{-8}	10^{-10}	10^{-12}
Абсолютное отклонение	3.1×10^0	3.5×10^{-2}	3.5×10^{-4}	3.5×10^{-6}	3.7×10^{-7}	5.4×10^{-7}	4.7×10^{-5}	1.7×10^{-3}

Таблица 3. ($i = 6, u_6 = 10^{-7}$)

Δ	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-11}	10^{-12}	10^{-13}
Абсолютное отклонение	$1.3 \times 10^{+6}$	$1.2 \times 10^{+6}$	$3.0 \times 10^{+5}$	$4.1 \times 10^{+3}$	$4.0 \times 10^{+1}$	$3.5 \times 10^{+0}$	$7.5 \times 10^{+1}$	$7.5 \times 10^{+2}$

В табл. 1–3 представлены результаты сравнения вычисления градиента потенциала Терсоффа на основе формул БАД-методологии и вычисления этого градиента с помощью метода конечных разностей для трех компонент градиента (по параметрам $u_1 = D_e$, $u_3 = \beta$, $u_6 = \gamma$). Все входные параметры задачи выбирались из указанного выше диапазона допустимых значений. Величина Δ в табл. 1–3 указывает на значения приращения параметра u_i , которые были использованы при вычислении градиента с помощью метода конечных разностей.

В табл. 1 сравниваются значения первой компоненты градиента энергии по параметру $u_1 = D_e$ при $D_e = 1$. Значение этой компоненты градиента, вычисленное с помощью БАД-методологии (точное значение), равно $\frac{\partial E(u)}{\partial u_1} = -7.079590296064$.

В табл. 2 сравниваются значения третьей компоненты градиента энергии по параметру $u_3 = \beta$ при $\beta = 1$. Значение этой компоненты градиента, вычисленное с помощью БАД-методологии (точное значение), равно $\frac{\partial E(u)}{\partial u_3} = 8.117021018694$.

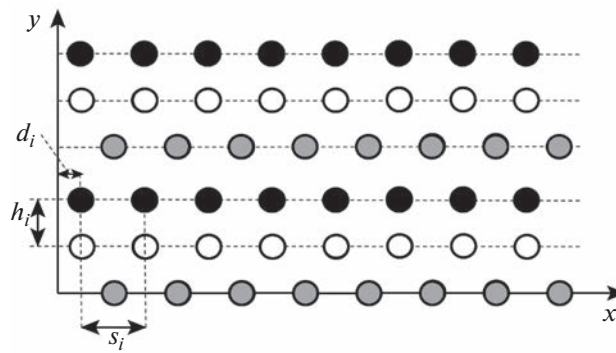
В табл. 3 сравниваются значения шестой компоненты градиента энергии по параметру $u_6 = \gamma$ при $\gamma = 1.0 \times 10^{-7}$. Значение этой компоненты градиента, вычисленное с помощью БАД-методологии (точное значение), равно $\frac{\partial E(u)}{\partial u_6} = 1.288320939087 \times 10^{+6}$.

Как следует из табл. 1, при вычислении первой компоненты градиента наиболее подходящим является выбор Δ из интервала $[10^{-8}; 1.0]$, что составляет от 10^{-7} до 100% от начального приближения $u_1 = 1.0$.

Из табл. 2 следует, что при вычислении третьей компоненты градиента наиболее подходящим является выбор $\Delta \approx 10^{-7}$. Это составляет примерно $10^{-5}\%$ от начального приближения $u_3 = 1.0$.

Из табл. 3 следует, что при вычислении шестой компоненты градиента наиболее подходящим является выбор $\Delta \approx 10^{-11}$. Это составляет примерно $10^{-2}\%$ от начального приближения $u_6 = 10^{-7}$.

На основании сказанного можно сделать вывод о том, что вычисление градиента энергии системы атомов с помощью метода конечных разностей связано с определенными трудностями. В случае использования этого метода необходимо проводить исследования, связанные с выбором приращения каждого параметра потенциала, причем при каждом его новом значении. В противоположность этому, алгоритм, использующий формулы БАД, позволяет автоматически получить с машинной точностью значение этого градиента.



Фиг. 1. Двумерная модель материала.

2. СРАВНЕНИЕ МЕТОДОВ ОПРЕДЕЛЕНИЯ ПРОИЗВОДНЫХ ДЛЯ ПЛОСКОЙ МОДЕЛИ МНОГОСЛОЙНОГО КУСОЧНО-ОДНОРОДНОГО МАТЕРИАЛА

В качестве вычислительного эксперимента для сравнения методов определения производных с помощью указанных выше способов и с помощью стандартных пакетов программ рассмотрена предложенная в [14] плоская модель многослойного кусочно-однородного материала. Материал представляется в виде периодической кусочно-однородной многослойной структуры, в рамках которой могут отличаться типы атомов в различных слоях. Модель накладывает ряд ограничений на структуру слоев:

- каждый слой состоит из одинаковых атомов, в разных слоях могут находиться различные атомы;
- расстояния между соседними атомами в одном слое одинаковы, но в разных слоях они могут отличаться;
- в рассматриваемой системе выделяется группа из K параллельных слоев, которая периодически повторяется в направлении оси y (см. фиг. 1);
- число атомов в каждом слое и общее число слоев считаются потенциально неограниченными.

На фиг. 1 представлен пример модели, в которой повторяется группа из трех слоев. Каждый слой состоит из атомов своего вида.

В рамках данной модели положение атомов определяется следующими параметрами:

h_i , $i = 1, \dots, K$, — расстояние между слоем с номером i и предыдущим слоем,

d_i , $i = 1, \dots, K$, — смещение первого атома в слое i с положительной абсциссой относительно нулевой отметки,

s_i , $i = 1, \dots, K$, — расстояние между атомами в слое i .

Совокупность значений перечисленных параметров будем называть *конфигурацией*. Ниже рассматривается ситуация, когда материал находится в ненагруженном состоянии. Требуется определить конфигурацию, соответствующую минимальной энергии взаимодействия атомов, входящих в моделируемый фрагмент материала.

Оптимизационная задача состоит в минимизации энергии $E(u, w)$ группы атомов (потенциал взаимодействия атомов — потенциал Терсоффа), расположенных на K соседних слоях. Параметрами оптимизационной задачи являются переменные $w = (h_1, d_1, s_1, \dots, h_K, d_K, s_K)$, составляющие вектор размерности $r = 3K$.

Пусть (для наглядности) $x_i = x_{1i}$ — первая координата i -го атома двумерной структуры, $y_i = x_{2i}$ — его вторая координата. Если i -й атом является атомом с порядковым номером j на k -м слое, то

$$x_i = d_k + (j-1)s_k; \quad y_i = \begin{cases} 0, & k = 1, \\ \sum_{m=2}^k h_m, & k = 2, \dots, K. \end{cases}$$

Соответствующие производные вычисляются по формулам (см. разд. 1)

$$\frac{\partial E}{\partial h_k} = \sum_{i=1}^I \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial h_k}, \quad \frac{\partial E}{\partial d_k} = \sum_{i=1}^I \frac{\partial E}{\partial x_i} \frac{\partial x_i}{\partial d_k}, \quad \frac{\partial E}{\partial s_k} = \sum_{i=1}^I \frac{\partial E}{\partial x_i} \frac{\partial x_i}{\partial s_k},$$

где $\frac{\partial E}{\partial x_i}$ и $\frac{\partial E}{\partial y_i}$ – производные функции энергии по первой и второй координатам атома соответственно.

Расчет производных энергии $E(u, w)$ по параметрам $w = (h_1, d_1, s_1, \dots, h_K, d_K, s_K)$ осуществлялся как с использованием аналитически выведенных формул, так и с помощью имеющихся пакетов программ.

Работа выполнялась с использованием инфраструктуры Центра коллективного пользования “Высокопроизводительные вычисления и большие данные” (ЦКП “Информатика”) ФИЦ ИУ РАН (г. Москва). Испытательный стенд состоял из следующих средств вычислительной техники:

- Intel Core i7 4770 3.4 Гц, 8 Гб RAM – 4-х ядерный процессор архитектуры Haswell с поддержкой технологий Hyper-Threading (два вычислительных протока на ядро, всего восемь логических ядер), AVX2 – Advanced Vector Extensions (в контексте решаемой задачи – обработка данных в формате с плавающей запятой в группах длиной 256 бит, т.е. для чисел с двойной точностью одновременно обрабатывается вектор из четырех чисел), и FMA – Fused Multiply-Add – выполнение совмещенной операции умножения-сложения вида $a = a + b * c$. Использовался компилятор GNUC++ 6.4.0.

- 2x Intel Xeon E5-2683 V4 2.1 Гц, 512 Гб RAM – двухпроцессорный сервер, каждый процессор содержит 16 ядер с поддержкой Hyper-Threading, AVX2 и FMA (всего 64 логических ядра). Использовался компилятор GNUC++ 7.2.0.

В работе использовались хорошо зарекомендовавшие себя ранее (см. [15]) пакеты CoDiPack 1.6.0 (см. [10]) и Adept 1.1 (см. [11]).

Сравнивались следующие методы вычисления градиента сложной функции $E(u, w)$:

1) использование выведенных аналитически выражений для вычисления частных производных этой функции (метод **Hand1**),

2) использование выведенных аналитически формул БАД-методологии для вычисления частных производных этой функции (метод **Hand2**),

3) использование прямого метода автоматического дифференцирования, реализованного в пакете CoDiPack с типом данных RealForwardVecGen (метод **CoDiPack_kr**),

4) использование прямого метода автоматического дифференцирования, реализованного в пакете CoDiPack с типом данных RealForward (метод **CoDiPack_rk**),

5) использование прямого метода автоматического дифференцирования, реализованного в пакете CoDiPack с типом данных RealForward, распараллеленный с помощью директивы OpenMP – #pragma omp for (метод **CoDiPack_rk**),

6) использование обратного метода автоматического дифференцирования, реализованного в пакете Adept ([11]) (метод **Adept**).

Для методов **Hand1**, **CoDiPack_kr**, **CoDiPack_rk** и **CoDiPack_rk_omp** использовались опции оптимизации компилятора O2 и Ofast. Для методов **Hand2** и **Adept** только опция O2.

Расчеты проводились для фиксированного набора специфических параметров потенциала Терсоффа, но для различного количества слоев и атомов:

- 4 слоя и 94 атома, размерность r вектора w равна 12 (распределение атомов по слоям 28, 19, 28, 19),

- 8 слоев и 188 атомов, размерность r вектора w равна 24 (распределение атомов по слоям 28, 19, 19, 28, 28, 19, 19, 28),

- 12 слоев и 204 атома, размерность r вектора w равна 36 (распределение атомов по слоям 15, 19, 19, 15, 15, 19, 19, 15, 15, 19, 19, 15).

При расчете градиента пакетом CoDiPack применялись типы данных RealForward и RealForwardVecGen. Тип данных RealForward состоит из двух чисел – значения переменной и ее производной по выбранной переменной. Для расчета градиента функции необходимо запустить вычисление функции и производной в цикле r раз. Тип данных RealForwardVecGen состоит из $r + 1$ чисел – значения переменной и вектора ее частных производных. При расчете градиента цикл

Таблица 4. Время (секунды) вычисления функции и градиента (одновременно с функцией) для прямых методов

Компьютер	IntelCore i7			IntelXeon		
Слой*атомы	4*94	8*188	12*204	4*94	8*188	12*204
Функция O2	0.25	0.94	1.17	0.25	0.98	1.23
Функция Ofast	0.23	0.92	1.16	0.29	1.14	1.43
Hand1 O2	25.45	256.58	360.98	20.91	138.09	186.02
Hand1 Ofast	25.89	218.89	325.18	22.96	154.43	210.42
CoDiPack_kr O2	7.23	62.85	122.88	12.19	99.66	189.23
CoDiPack_kr Ofast	1.66	17.61	72.91	2.68	23.14	134.45
CoDiPack_rk O2	7.75	55.71	99.62	8.76	66.04	125.05
CoDiPack_rk Ofast	5.55	43.05	82.95	5.88	37.92	71.77
CoDiPack_rk_omp O2	1.91	11.30	22.89	1.21	6.93	6.65
CoDiPack_rk_omp Ofast	1.64	10.00	20.55	0.72	5.10	4.75

Таблица 5. Время (секунды) вычисления функции и градиента (одновременно с функцией) для обратных методов

Компьютер	Intel Core i7			Intel Xeon		
Слой*атомы	4*94	8*188	12*204	4*94	8*188	12*204
Hand 2	2.66	23.03	29.61	3.95	54.28	68.44
Adept	1.25	5.88	8.03	2.43	10.51	13.60

по r становится внутренним, но функция рассчитывается один раз. Использование типа данных RealForwardVecGen позволяет задействовать технологию AVX2 и ускоряет вычисления от 1.5 до 4 раз без каких-либо трудозатрат; достаточно указать компилятору опцию оптимизации —Ofast.

В табл. 4 приводится сравнение прямых методов дифференцирования **Hand1**, **CoDiPack_kr**, **CoDiPack_rk** в однопоточном исполнении и **CoDiPack_rk_omp**, распараллеленный с помощью директивы OpenMP — #pragma omp for. Как видно из таблицы, время вычисления функции имеет примерно квадратичную зависимость от количества атомов.

Время вычисления производной методом **Hand1** изменяется на 10–17% в зависимости от установленных опций оптимизации, при этом для IntelCore i7 опция Ofast ускоряет расчет, а для IntelXeon наоборот замедляет.

Метод **CoDiPack_kr** при опции O2 оказывается в 3–4 раза быстрее метода **Hand1** с той же опцией. Связано это с тем, что при реализации методов **Hand1** и **Hand2** задача оптимизации кода не ставилась, а пакетные методы оптимизировались. Сравнивая влияние опций O2 и Ofast, мы видим существенное ускорение решения задачи при $r = 12$ и несколько уменьшающееся ускорение для больших размерностей. Для малой размерности задачи преимущества технологии AVX2 раскрываются полностью.

Метод **CoDiPack_rk**, как можно увидеть, не так сильно чувствителен к опциям оптимизации компилятора. Вычислительно он более сложен, так как значение функции рассчитывается r раз. Но его применение целесообразно при распараллеливании процесса расчета градиента на несколько ядер. Эксперименты с методом **CoDiPack_rk_omp** это показывают.

В табл. 5 приведено время вычисления градиента и функции обратными методами **Hand2** и **Adept**. Как видно из таблицы, время вычисления функции и градиента мало зависит от размерности задачи. Исключение составляет **Hand2** для задачи с $r = 12$ (4*94). Связано это с небольшим количеством памяти, при котором вычисления выполняются в основном с использованием кэш-шей процессора. Как и в случае с прямыми методами оптимизация кода **Hand2** не проводилась, с чем и связано большее время вычислений. Также необходимо обратить внимание на тот факт, что вычисления прямым методом **CoDiPack_rk_omp** на компьютере IntelXeon оказываются более эффективными, чем вычисления обратным методом **Adept**, за счет параллельного исполнения.

3. ЗАКЛЮЧЕНИЕ

Проведенные исследования, некоторые результаты которых представлены в настоящей работе, показывают, что при вычислении градиента сложной функции наиболее эффективным оказалось использование методологии быстрого автоматического дифференцирования. При этом обе версии БАД-методологии (и “ручная”, и “пакетная”) показали свое значительное преимущество перед другими методами (при отсутствии распараллеливания). С точки зрения затрат машинного времени обе версии БАД-методологии примерно равноценны (в случае оптимизации “ручной” версии), опережая одна другую на отдельных задачах.

Преимущество “пакетной” версии БАД-методологии в том, что для ее работы необходимо задать алгоритм вычисления только самой функции. Все остальные этапы (построение алгоритма решения сопряженной задачи, построение формул для вычисления градиента) выполняются пакетом автоматически.

Что касается “ручной” версии БАД-методологии, то в этом случае необходимо вручную построить многошаговый процесс вычисления функции, с помощью канонических формул БАД-методологии построить многошаговый процесс решения сопряженной задачи и вывести формулы для вычисления градиента функции. Безусловно, все это требует серьезных интеллектуальных затрат исследователя. Однако при этом у исследователя появляется возможность учесть специфику конкретной задачи и сделать алгоритм вычисления градиента функции более эффективным. К преимуществам “ручной” версии БАД-методологии следует отнести и то, что эта версия гораздо слабее, чем “пакетная” версия, зависит от параметров используемого компьютера.

СПИСОК ЛИТЕРАТУРЫ

1. Крылов И.А., Черноусько Ф.Л. О методе последовательных приближений для решения задач оптимального управления // Ж. вычисл. матем. и матем. физ. 1962. Т. 2. № 6. С. 669–683.
2. Федоренко Р.П. Приближенное решение задач оптимального управления. М.: Наука, 1978.
3. Моисеев Н.Н. Численные методы в теории оптимальных систем. М.: Наука, 1971.
4. Албу А.Ф., Зубов В.И. Вычисление градиента функционала в одной задаче оптимального управления, связанной с кристаллизацией металла // Ж. вычисл. матем. и матем. физ. 2009. Т. 9. № 1. С. 51–75.
5. Евтушенко Ю.Г., Засухина Е.С., Зубов В.И. О численном подходе к оптимизации решения задачи Бюргерса с помощью граничных условий // Ж. вычисл. матем. и матем. физ. 1997. Т. 37. № 12. С. 1449–1458.
6. Айда-Заде К.Р., Евтушенко Ю.Г. Быстрое автоматическое дифференцирование // Матем. моделирование. 1989. Т. 1. С. 121–139.
7. Evtushenko Y.G. Computation of exact gradients in distributed dynamic systems // Optimizat. Meth. and Software. 1998. V. 9. P. 45–75.
8. Евтушенко Ю.Г. Оптимизация и быстрое автоматическое дифференцирование. Научное издание. ВЦ им. А.А. Дородницына РАН. Москва. 2013. 144 с.
9. Hascoet L., Pascual V. The Tapenade automatic differentiation tool: principles, model, and specification // ACM Transact. on Math. Software (TOMS). 2013. T. 39. № 3. С. 1–43.
10. Albring T. et al. An aerodynamic design framework based on algorithmic differentiation // ERCOFTAC Bull. 2015. V. 102. P. 10–16.
11. Hogan R.J. Fast reverse-mode automatic differentiation using expression templates in C++ // ACM Transact. on Math. Software (TOMS). 2014. V. 40. № 4. P. 26–42.
12. Албу А.Ф. Применение быстрого автоматического дифференцирования для вычисления градиента потенциала Терсоффа // Информ. технологии и вычисл. системы. 2016. Т. № 1. С. 43–49.
13. Абгарян К.К., Посыпкин М.А. Программный комплекс для решения задач параметрической идентификации потенциалов межатомного взаимодействия // Inter. J. of Open Inform. Tech. 2014. Т. 2. № 10. С. 14–19.
14. Евтушенко Ю.Г., Лурье С.А., Посыпкин М.А., Соляев Ю.О. Применение методов оптимизации для поиска равновесных состояний двумерных кристаллов // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 12. С. 2032–2041.
15. Горчаков А.Ю. О программных пакетах быстрого автоматического дифференцирования // Информ. технологии и вычисл. системы. 2018. № 1. Р. 30–36.