ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Октябрь—декабрь 2007. Серия 1. Том 14, № 4. 57–75

УДК 519.7

О СЛОЖНОСТИ ГРАДИЕНТА РАЦИОНАЛЬНОЙ ФУНКЦИИ*)

И. С. Сергеев

Из метода Баура—Штрассена следует, что $L(\bigtriangledown f) \leqslant 4L(f)$, где $L(f)$ — сложность реализации рациональной функции f схемами над арифметическим базисом, а $\bigtriangledown f$ — градиент функции f. Показано, что $L(\bigtriangledown f) \leqslant 3L(f)+n$, где n — число переменных функции f. Кроме того, получены оценки глубины схемы для градиента.

Введение

Рассматривается реализация рациональных функций над некоторым полем F схемами над арифметическим базисом

$$B_A = \{x \pm y, xy, x/y\} \cup \{ax \mid a \in F\}.$$

Сложность реализации системы функций $f_1, \ldots, f_k$ будем обозначать через $L(f_1, \ldots, f_k)$, а глубину — через $D(f_1, \ldots, f_k)$.

В. Баур и Ф. Штрассен [7] в 1983 г. показали, что по схеме, реализующей функцию f, можно построить схему, реализующую f и $\bigtriangledown f$ с не более чем в 4 раза большей сложностью или, если игнорировать аддитивные операции и скалярные умножения, с не более чем в 3 раза большей. Похожий метод ранее был предложен в [14], а в [4] аналогичный результат был получен для более общей задачи, однако в обоих случаях указывались менее точные оценки.

Результат Баура—Штрассена уточнялся в последующих работах [15, 13, 2]. В частности, в [15, 2] теорема Баура—Штрассена обобщена на более широкий класс функций, чем рациональные (фактически, на класс всех элементарных функций). В [13] отмечено, что глубина схемы для градиента, которая строится в методе Баура—Штрассена, превосходит глубину схемы для самой функции в $O(1)$ раз.

*)Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект 05–01–00994), программы поддержки ведущих научных школ РФ (проект НШ–5400.2006.1) и программы фундаментальных исследований Отделения математических наук РАН «Алгебраические и комбинаторные методы математической кибернетики» (проект «Синтез и сложность управляющих систем»).

Объединяя результаты [7, 13], получаем, что для рациональной функции $f(x_1, \ldots, x_n)$ справедливы соотношения:

$$L(f, \bigtriangledown f) \leqslant 4L(f); \qquad D(f, \bigtriangledown f) = O(D(f)). \tag{1}$$

В доказательствах [7, 13] неявно используется так называемый принцип транспозиции. В [2] он сформулирован и применён явно. В интерпретации [2] строится схема, реализующая дифференциал функции df и имеющая помимо входных переменных $x_1, \ldots, x_n$ также дифференциальные входы $dx_1, \ldots, dx_n$. После этого результат о сложности градиента получается автоматически применением принципа транспозиции. Отметим, что вывод основного результата Баура и Штрассена (о мультипликативной сложности градиента) методом [2] является элементарным. В [2] получены оценки

$$L(f, df) \leqslant 4L(f); \qquad D(f, df) \leqslant 3D(f),$$

из которых следует (1). Результаты относительно схем для дифференциалов без изменения переносятся на вектор-функции.

Используя анализ изменения глубины схемы при транспозиции, основанный на методах минимизации глубины схем [5, 11] (см. также [1]), можно получить оценку

$$D(f, \bigtriangledown f) \leqslant 7D(f).$$

Здесь и далее мы пользуемся соглашением, согласно которому оценка глубины приводится для той же схемы, которая доставляет оценку сложности. Вообще, для глубины реализации градиента справедливо очевидное соотношение $D(\bigtriangledown f) \leqslant D(df)$, так как схему для произвольной частной производной функции f можно получить из схемы для дифференциала при подаче на соответствующий дифференциальный вход константы 1, а на остальные дифференциальные входы — константы 0.

Оценка сложности метода Баура—Штрассена следует из соотношения $L(f, df) \leqslant 4M + 2A + 2S$, где M, A, S — соответственно число мультипликативных, аддитивных элементов и элементов скалярного умножения в минимальной схеме, реализующей f. Ниже будет показано, что $L(f, df) \leqslant 3M + 3A + 2S + n$, где n — число аргументов функции f, и установлены следующие оценки:

$$L(f, df) \leqslant 3L(f) + n; \quad D(f, df) \leqslant 2D(f) + 1;$$
$$L(f, \bigtriangledown f) \leqslant 3L(f) + n; \quad D(f, \bigtriangledown f) \leqslant 6D(f) + 4.$$

Иногда вместо базиса B_A удобнее рассматривать базис $B'_A = B_A \cup \{-x - y\}$. Такой базис использовался, например, в [2]. Все излагаемые ниже результаты справедливы как для схем над базисом B_A, так и для схем над базисом B'_A. Доказательства приводятся только для базиса B_A, однако они легко переносятся на случай базиса B'_A.

Изложение построено следующим образом. В § 1 даётся формулировка и доказательство принципа транспозиции. В § 2 выводится следствие о дифференциале и градиенте. Материал этих двух параграфов существенно пересекается с работой [2]. Подробное рассмотрение указанных вопросов вызвано необходимостью более детального анализа глубины. В § 3 излагается базовый метод, в § 4 приводится уточнение, соединяющее метод § 3 с методом Баура—Штрассена (под методом Баура—Штрассена далее, как правило, будет подразумеваться интерпретация из [2]).

§ 1. Принцип транспозиции

История многократно переоткрывавшегося принципа транспозиции восходит, в том числе, к работе О. Б. Лупанова [6], в которой используется принцип транспозиции для вентильных схем. В наиболее общей формулировке, принадлежащей Фидуччиа [9], принцип транспозиции распространяется на схемы для билинейных преобразований. Для целей настоящей работы достаточным является линейный вариант принципа транспозиции, излагаемый ниже. Более полный исторический обзор содержится в [2, 3, 8].

Пусть схема V над линейным базисом

$$B_2 = \{x \pm y\} \cup \{ax \mid a \in F\}$$

реализует некоторое линейное (m, n)-преобразование $\mathbf{y} = C\mathbf{x}$, где $\mathbf{x} = (x_1, \ldots, x_n)^T$ и $\mathbf{y} = (y_1, \ldots, y_m)^T$ — соответственно векторы входов и выходов, а C — матрица преобразования.

Определим операцию *приведения* схемы V к схеме над базисом

$$B_\infty = \{x_1 \pm \ldots \pm x_k \mid k \in \mathbb{N}\} \cup \{ax \mid a \in F\}$$

следующим образом. *Неприводимой аддитивной подсхемой* схемы V назовём связную подсхему, состоящую только из аддитивных элементов и ведущих в них дуг, имеющую единственный выход (т. е. никакие внутренние элементы подсхемы не имеют дуг, ведущих к элементам вне подсхемы, и не являются выходами схемы V) и не содержащуюся строго ни в какой другой удовлетворяющей перечисленным требованиям подсхеме.

Пусть такая подсхема содержит k аддитивных элементов, тогда её можно заменить одним $(k+1)$-входовым аддитивным элементом из базиса B_∞, реализующим ту же функцию. Схему над B_∞, полученную указанным преобразованием всех неприводимых аддитивных подсхем схемы V, назовём *приведённой*. *Высотой* схемы V назовём глубину её приведённой схемы (это определение согласуется с определением высоты для формул из [5]).

Обратное преобразование схемы над B_∞ к схеме над B_2 можно выполнить методом Ложкина [5] (см. также [1]), в котором многовходовые элементы заменяются соответствующими бинарными деревьями из аддитивных элементов, сбалансированными по глубине (т. е. если входы расположены на глубинах $d_1, \ldots, d_k$, то выход расположен на глубине $\lceil \log_2(2^{d_1} + \ldots + 2^{d_k}) \rceil$). Покажем, что такая замена возможна.

Рассмотрим произвольный элемент e, реализующий k-местную функцию $\varphi \in B_\infty$, со входами на глубинах $d_1, \ldots, d_k$. Без ограничения общности будем считать, что все входы различны. Построим корневое бинарное дерево, ориентированное к корню, с k листьями, расположенными на глубинах $d_1, \ldots, d_k$, и корнем на глубине $\lceil \log_2(2^{d_1} + \ldots + 2^{d_k}) \rceil$ (известно, что такое дерево существует, см., например, [1]). Листья построенного дерева ассоциируем со входами элемента e и опишем процедуру размещения элементов базиса B_2 в вершинах дерева, в результате которой схема будет реализовывать функцию φ. Будем говорить, что функция ψ *входит в* φ *со знаком «+»* (*«−»*), если функция $\varphi - \psi$ (соответственно, $\varphi + \psi$) существенно не зависит от аргументов ψ. Будем размещать элементы в вершинах дерева последовательно от листьев к корню по следующему правилу. Пусть входы некоторой вершины уже реализуют функции ψ_1 и ψ_2. Если эти функции входят в φ с одинаковыми знаками, то в данной вершине размещается элемент суммы и реализуется функция $\psi_1 + \psi_2$. Если ψ_1 и ψ_2 входят в φ с разными знаками, то в вершину помещается элемент разности, который реализует $\psi_1 - \psi_2$, когда ψ_1 входит в φ со знаком «+», и $\psi_2 - \psi_1$, когда ψ_1 входит в φ со знаком «−». Несложно проверить корректность описанной процедуры. В результате получается схема над B_2, реализующая φ.

Эквивалентной сложностью (*глубиной*) схемы над B_∞ назовём сложность (глубину) схемы, получающейся заменой многовходовых аддитивных элементов реализующими те же функции сбалансированными деревьями из двухвходовых элементов. Заметим, что в силу определения операции приведения сложность схемы над B_2 совпадает с эквивалентной сложностью приведённой схемы.

Лемма 1. *Пусть схема глубины h над B_∞ имеет r дуг и p аддитивных элементов. Тогда её эквивалентная сложность равна $r-p$, а эквивалентная глубина каждого выхода не превосходит $h+\lceil\log_2 Q\rceil$, где Q — число различных ориентированных путей, соединяющих этот выход со входами схемы.*

Доказательство. Пусть s — число элементов скалярного умножения в схеме. Тогда число дуг, входящих в аддитивные элементы, равно $r-s$. При описанном выше преобразовании k-входовый элемент с входящими в него k дугами заменяется деревом из $k-1$ элементов и k дуг. Таким образом, число аддитивных элементов в эквивалентной схеме равно $r-s-p$, и вместе с s скалярными элементами равно $r-p$. Доказательство оценки для глубины можно найти в [1].

Мы будем использовать принцип транспозиции в специальной формулировке.

Лемма 2 (принцип транспозиции). *Пусть схема V над базисом B_2 реализует линейное (m,n)-преобразование с матрицей C, и пусть $L(V)$ — её сложность. Тогда можно построить схему V', реализующую преобразование с матрицей C^T, сложность которой удовлетворяет соотношению*

$$L(V')-L(V)\leqslant m-1.$$

При этом наборы скалярных элементов схем V и V' различаются только некоторым числом элементов умножения на -1. Кроме того, если обозначить через $D(V)$ и $H(V)$ глубину и высоту схемы V, а через $G(V)$ скалярную глубину приведённой схемы (максимальное число скалярных элементов с ветвлением выхода в цепочке между входом и выходом схемы), то также выполнено неравенство

$$D(V')-D(V)\leqslant H(V)+G(V)+\lceil\log_2 m\rceil+1.$$

Доказательство. Пусть схема W над B_∞ является приведённой к схеме V. Каждой дуге схемы W поставим в соответствие элемент из F (*метку*) по следующему правилу: если дуга является входом аддитивного элемента, то метка равна 1 в случае, когда вход участвует в вычисляемой линейной комбинации со знаком «плюс», и -1, когда со знаком «минус». Дуга, ведущая к элементу скалярного умножения, снабжается меткой, совпадающей со значением скалярного множителя. *Проводимостью* пути, соединяющего две вершины в схеме, назовём произведение меток его дуг (проводимость пути между совпадающими вершинами считается равной 1). *Функция проводимости* упорядоченной пары вершин

определяется как сумма проводимостей всех путей, ведущих из первой вершины во вторую (если таких путей нет, то функция полагается равной нулю).

Такая схема с помеченными дугами, функционирующая как *вентильная схема* в смысле [2], также реализует линейное преобразование с матрицей $C = (c_{ij})$. Это значит, что функция проводимости между входом x_j и выходом y_i построенной схемы равна c_{ij}, т. е. матрица проводимости схемы совпадает с матрицей C.

Следуя работе [2], введём понятие *транспонированной схемы*. Изменим ориентацию дуг схемы W, не изменяя меток. Полученную схему обозначим через W'. Очевидно, что матрица проводимости такой схемы совпадает с C^T. Далее опишем алгоритм преобразования этой схемы к схеме (которую и будем называть транспонированной) с корректным соответствием между метками дуг и функциональными элементами. Вначале элементы схемы W' считаются неопределёнными, однако про них известно, что они являются аддитивными или скалярными, в зависимости от реализуемых ими функций до изменения ориентации.

1) В схему W' добавляются дуги с единичными метками, ведущие от входов y_i к элементам, которые прежде были выходами схемы W. Число дуг схемы увеличивается на m.

2) На место входов схемы W помещаются аддитивные элементы-выходы новой схемы. Таким образом, число аддитивных элементов в схеме увеличивается на n.

3) В схему добавляются новые аддитивные элементы-выходы x_j. Каждый добавленный в п. 2 элемент соединяется дугой, отмеченной единицей, с новым выходом схемы. Число дуг и аддитивных элементов схемы изменяется на одну и ту же величину.

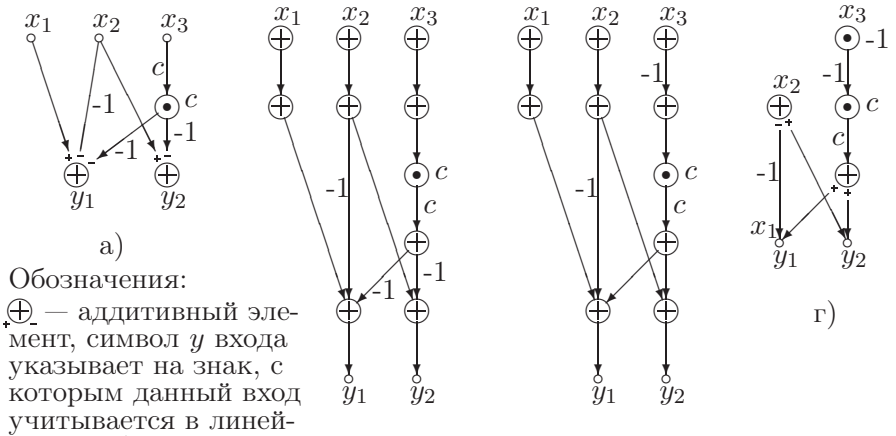
4) Если элемент скалярного умножения схемы W' имеет один вход, то метки входящей и исходящей из него дуг следует поменять местами. Иначе, если у скалярного элемента образуется пучок, состоящий не менее чем из двух входов, то в схему вставляется аддитивный элемент, имеющий данный пучок входов. Выход добавленного элемента соединяется со входом рассматриваемого скалярного элемента, добавленная дуга отмечается меткой скалярного множителя. Исходящей из скалярного элемента дуге присваивается метка 1. Функция, которую реализовывал данный скалярный элемент в схеме W, сохраняется. На данном шаге число дуг и аддитивных элементов в схеме увеличилось на одну и ту же величину.

После выполнения шагов 1–4 разность между числом дуг и числом аддитивных элементов схемы изменилась в точности на $m - n$. Неопре-

делёнными остаются аддитивные элементы схемы. Метки входящих в них дуг могут принимать только значения ± 1 в силу п. 4. Шаги 5 и 6 выполняются только тогда, когда характеристика поля F отлична от двух.

5) Следующая процедура применяется к аддитивным элементам последовательно, от входов к выходам схемы. Если метки всех дуг, входящих во внутренний аддитивный элемент схемы W', равны -1, то эти метки заменяются на 1, и также инвертируются метки исходящих из данного элемента дуг, за исключением тех, которые ведут к элементам скалярного умножения: в таком случае метка входящего в скалярный элемент дуги остается без изменений, а указанная процедура инвертирования применяется к исходящему из скалярного элемента ребру.

После выполнения шага 5 в схеме не остаётся аддитивных элементов, за исключением, может быть, некоторых выходов, имеющих все входящие дуги с меткой -1.



Обозначения:
$\oplus^{+}_{-}$ — аддитивный элемент, символ y входа указывает на знак, с которым данный вход учитывается в линейной комбинации;
$\odot\, c$ — элемент скалярного умножения на c.

Рис. 1

6) Выходы схемы с меткой входящей дуги -1 заменяются элементами скалярного умножения на -1. Таким образом, из схемы удаляется некоторое число аддитивных элементов. Ниже будет показано, что оно не превосходит $n-1$.

7) Одновходовые аддитивные элементы схемы (с меткой входящей дуги 1 в силу п. 6) удаляются из схемы вместе с ведущими в них дугами. Исходящие из удалённых элементов дуги соединяются с элементом (или входом схемы) на другом конце удалённой дуги. Число дуг и аддитивных элементов схемы уменьшается на одну и ту же величину. Оставшиеся ад-

дитивные элементы реализуют функции из B_∞ в соответствии с метками входящих в них дуг.

Пример 1. На рис. 1*а* изображена схема, реализующая линейное преобразование

$$y_1 = x_1 - x_2 - cx_3, \quad y_2 = x_2 - cx_3.$$

На рис. 1*б* и 1*в* показаны схемы, получающиеся из нее при транспонировании после выполнения шагов 4 и 5 соответственно. Окончательный вид схемы, реализующей транспонированное преобразование

$$x_1 = y_1, \quad x_2 = y_2 - y_1, \quad x_3 = -c(y_1 + y_2),$$

приведён на рис. 1*г* (единичные метки дуг на рисунках не указаны).

Индукцией по сложности схемы W покажем, что после выполнения шага 5 хотя бы один выход схемы W' имеет входящую дугу с меткой 1. Действительно, для схем единичной сложности это утверждение справедливо. Предположим, что оно выполняется для схем сложности $r-1$, и рассмотрим произвольную схему w сложности r. Пусть схема w_1 получается из схемы w удалением одного элемента (с ведущими к нему дугами), входами которого являются только входы схемы. Будем считать этот элемент аддитивным — рассуждение для случая скалярного элемента совершенно аналогично. Рассмотрим схему w_1', полученную из w_1 после выполнения шагов 1–5 алгоритма транспонирования. По предположению, хотя бы один выход схемы w_1' имеет входящую дугу с меткой 1. Рассмотрим также схему w', получаемую из w в алгоритме транспонирования в тот момент, когда процедура шага 5 выполнена для всех аддитивных элементов, за исключением элемента e, соответствующего отсутствующему в схеме w_1, и зависящих от него элементов (к последним относятся элементы, в которые ведут дуги из указанного, а также соединённые с этими элементами выходы схемы w'). При этом по построению имеется соответствие между дугами, ведущими к выходам схемы w_1', и дугами схемы w' трех типов: (1) входы элемента e, (2) входы элементов, расположенных между элементом e и выходами схемы w', не соединённые с e, (3) дуги, ведущие к не зависящим от e выходам схемы w'. Соответствие включает, в частности, совпадение меток. Возможный вид такого соответствия изображен на рис. 2: показана часть схемы w', содержащая выходы; метки дуг, ведущих к выходам схемы w_1', обозначены через a_i (дуги с метками a_2, a_3, a_4 имеют тип (1), с метками a_1, a_5, a_6 — тип (2), с метками a_7, a_8, a_9 — тип (3)). Доказательство завершается рассмотрением трех случаев. Если дуга с меткой 1, соединённая с выходом

w'_1, соответствует дуге типа (1), то метки дуг, соединённых с e, не инвертируются на шаге 5. Следовательно, найдется выход схемы w', который соединён с e цепочкой из двух дуг с единичными метками; тогда метка входящей в него дуги останется равной 1 по завершении шага 5 (одна из исходящих из e дуг имеет метку 1). Если аналогичное соответствие выполняется для ребра типа (2), то у выхода, соединённого с элементом, к которому оно ведёт, метка входящей дуги останется единичной. Случай дуги типа (3) не нуждается в комментарии.

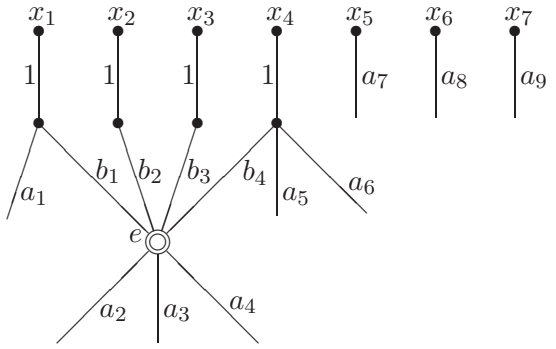


Рис. 2

Матрица проводимости схемы W' не изменялась при выполнении шагов 1–7. Поэтому построенная схема (как схема над базисом B_∞) реализует преобразование с матрицей C^T. Согласно лемме 1 её эквивалентная сложность не превосходит $L(V)+(m-n)+(n-1) = L(V)+m-1$ (иначе говоря, можно построить схему V' над B_2 такой сложности).

Оценим глубину схемы W'. Перед выполнением алгоритма транспонирования глубина совпадала с глубиной схемы W, равной $H(V)$. На каждом из шагов 2 и 3 глубина схемы увеличивается на единицу, на шаге 4 — увеличивается не более чем на $G(V)$, а на шаге 7 уменьшается по меньшей мере на единицу, так как из схемы неизбежно удаляются элементы, соответствующие выходам схемы W, расположенным на максимальной глубине. Используя лемму 1, эквивалентную глубину схемы W' (глубину схемы V') можно оценить как $H(V)+G(V)+1+\lceil\log_2 Q\rceil$, где Q — общее число путей, соединяющих входы и выходы схемы W'. Число путей не изменяется при транспонировании и оно одинаково для схем V, W, V', W'. В силу очевидного соотношения $D(V) \geqslant \lceil\log_2(Q/m)\rceil$ окончательно имеем

$$\begin{aligned} D(V')-D(V) &\leqslant H(V)+G(V)+1+\lceil\log_2 Q\rceil-\lceil\log_2(Q/m)\rceil \\ &\leqslant H(V)+G(V)+\lceil\log_2 m\rceil+1. \end{aligned}$$

Лемма 2 доказана.

Следствие 1. *Сложность преобразований с транспонированными матрицами удовлетворяет соотношению*

$$1 - n \leqslant L(C^T) - L(C) \leqslant m - 1,$$

а если сложность умножений на -1 *равна нулю* (*это заведомо так в случае поля характеристики два*), *то справедливо классическое соотношение* $L(C^T) - L(C) = m - n$.

Принцип транспозиции является достаточно универсальным инструментом. Он, в частности, используется в некоторых методах построения аддитивных цепочек (см., например, [3]), реализации автоморфизмов конечных полей (см., например, [12]), интерполяции многочленов (см., например, [9]). На принципе транспозиции основан метод построения схемы для градиента рациональной функции [7]. Доказательство, в котором этот принцип формулируется и применяется явно, было предложено в работе [2].

§ 2. Дифференциал и градиент

Рассмотрим схему над арифметическим базисом, реализующую df и обладающую тем свойством, что у каждого мультипликативного элемента не более чем один вход зависит от дифференциалов, причём у элемента деления — только числитель (такие схемы называются *линеаризованными*, см. [2]). В данной схеме выделим подсхему, элементы которой реализуют функциональные выражения, зависящие от дифференциалов (как и в [2], эту подсхему назовём *линейной*, а дополнительную к ней — *скалярной*). Мультипликативные элементы линейной подсхемы интерпретируются как элементы скалярного умножения, т. е. умножения на не зависящую от дифференциалов функцию (множество рациональных функций над полем F также является полем). Следующая лемма фактически содержится в [2].

Лемма 3. *Пусть* $L(V)$ *и* $D(V)$ — *сложность и глубина линеаризованной схемы* V, *реализующей дифференциал функции* $f(x_1, \ldots, x_n)$, $D'(V)$ — *глубина её скалярной подсхемы, а* $H(V)$ *и* $G(V)$ — *высота и скалярная глубина приведённой линейной подсхемы соответственно. Тогда*

$$L(\bigtriangledown f) \leqslant L(V); \qquad D(\bigtriangledown f) \leqslant D(V) + D'(V) + H(V) + G(V) + 1.$$

Доказательство. Пусть l и d — сложность и глубина линейной подсхемы Λ схемы V (очевидно, что $d \leqslant D(V)$). Согласно лемме 2 транспонированная к Λ схема (операции производятся в поле рациональных

функций над F) имеет эквивалентные сложность и глубину не выше l и $d + H(V) + G(V) + 1$ соответственно.

Такая схема при подаче на вход константы 1 вычисляет в точности $\bigtriangledown f$. Набор её скалярных элементов, не считая умножений на -1, совпадает с набором скалярных элементов схемы Λ и может быть реализован скалярной подсхемой схемы V. Присоединение скалярной подсхемы увеличивает глубину схемы для градиента не более чем на $D'(V)$. Лемма 3 доказана.

В методе из [2] $D(V), H(V) \leqslant 3D(f)$ и $D'(V), G(V) \leqslant D(f)$, откуда выводится оценка глубины схемы (которая кроме градиента реализует и саму функцию)

$$D(f, \bigtriangledown f) \leqslant 8D(f) + 1.$$

Более аккуратная оценка приведена во введении.

§ 3. Основной метод

Теорема 1. *Пусть f — рациональная функция n переменных. Тогда*

$$L(f, df) \leqslant 3L(f) + n; \qquad D(f, df) \leqslant 2D(f) + 1;$$

$$L(f, \bigtriangledown f) \leqslant 3L(f) + n; \qquad D(f, \bigtriangledown f) \leqslant 6D(f) + 4.$$

Эти оценки справедливы и для вектор-функций.

Доказательство. По существу отличие предлагаемого подхода к построению схемы от метода Баура—Штрассена сводится к различной реализации мультипликативных операций. В методе Баура—Штрассена дифференциалы $d(uv)$ и $d(u/v)$ вычисляются следующим образом:

$$d(uv) = (u \cdot dv) + (v \cdot du); \qquad d(u/v) = (du - (u/v) \cdot dv)/v.$$

Однако можно рассмотреть более симметричный способ:

$$d(uv) = (du/u + dv/v) \cdot (uv); \qquad d(u/v) = (du/u - dv/v) \cdot (u/v),$$

который, если применять его непосредственно, требует четыре операции (с дифференциалами) вместо трёх в первоначальном способе. Однако внутри достаточно большой схемы экономия достигается за счёт того, что (1) данный мультипликативный элемент $\circ$ реализует также дифференциальную функцию $d(u \circ v)/(u \circ v)$, которую можно использовать, если выход элемента подается на вход другого мультипликативного элемента; (2) если выход данного мультипликативного элемента не подаётся

на вход аддитивного элемента, либо на выход схемы, то заключительное умножение на $u \circ v$ выполнять не нужно.

Схема, которая одновременно реализует f и df, строится из схемы, реализующей f, при помощи следующей процедуры, аналогичной [7, 2]. Без ограничения общности считаем, что в схеме для f нет пар, соединённых дугой элементов скалярного умножения (каждую такую пару можно заменить парой скалярных элементов, расположенных на одном уровне, не изменяя сложность и не увеличивая глубину схемы).

Для мультипликативного элемента схемы, вычисляющего $u \circ v$, реализуются du/u и dv/v (для тех входов, которые являются либо выходами аддитивных элементов, либо входами схемы), сумма или разность du/u и dv/v, которая умножается на $u \circ v$ (если выход элемента присоединён ко входу аддитивного элемента или является выходом схемы, либо это выполняется для скалярного элемента, входом которого является рассматриваемый мультипликативный элемент). Для аддитивного элемента просто вычисляется соответствующая функция от дифференциалов $d(u \pm v) = du \pm dv$. Для элемента скалярного умножения u на $a \in F$, который является входом аддитивного элемента либо выходом схемы, вычисляется дифференциал $d(au) = a \cdot du$. Если выход данного скалярного элемента присоединён ко входу мультипликативного элемента, а вход является выходом аддитивного элемента либо входом схемы, то в схему добавляется элемент, вычисляющий du/u.

Обозначим через A, M, S соответственно число аддитивных, мультипликативных и скалярных элементов в схеме для f, а через m число выходов схемы (для вектор-функции). Сложность построенной схемы складывается из:

$A + M + S$ элементов исходной схемы;

не более $A + n$ элементов, вычисляющих выражения вида du/u (такие элементы соответствуют либо входам схемы, либо её аддитивным элементам);

не более $A+M$ элементов суммирования дифференциалов (по одному такому элементу приходится на каждый двухвходовый элемент исходной схемы);

не более $\min\{M + S, 2A + m\}$ элементов заключительных умножений для мультипликативных элементов суммарно с элементами умножения дифференциалов на скаляры поля F (с одной стороны, их число ограничено числом мультипликативных и скалярных элементов в схеме для f, а с другой — суммой числа входов аддитивных элементов и числа выходов схемы).

Окончательно

$$L(f, df) \leqslant 3M + 3A + 2S + n \leqslant 3L(f) + n.$$

Для оценки глубины заметим, что глубина элемента, вычисляющего $d(u \circ v)/(u \circ v)$, не более чем вдвое превосходит глубину соответствующего элемента $\circ$ в исходной схеме. Аналогично, глубина элемента, вычисляющего дифференциал выхода аддитивного или скалярного элемента исходной схемы, не более чем вдвое превосходит глубину последнего. Наконец, если выход схемы является мультипликативным элементом, вычисляющим $u \circ v$, то глубина вычисления $d(u \circ v)$ в схеме для дифференциала на единицу больше, чем глубина вычисления $d(u \circ v)/(u \circ v)$. Окончательно получаем

$$D(f, df) \leqslant 2D(f) + 1.$$

Построенная схема V для дифференциала является линеаризованной. Поэтому для оценки сложности и глубины схемы для градиента воспользуемся леммой 3, в которой $D(V), H(V) \leqslant 2D(f)+1$, $D'(V) \leqslant D(f)$, $G(V) \leqslant D(f) + 1$ (последнее неравенство справедливо в силу того, что по построению в линейной подсхеме схемы V не может быть двух соединённых дугой скалярных элементов). Теорема 1 доказана.

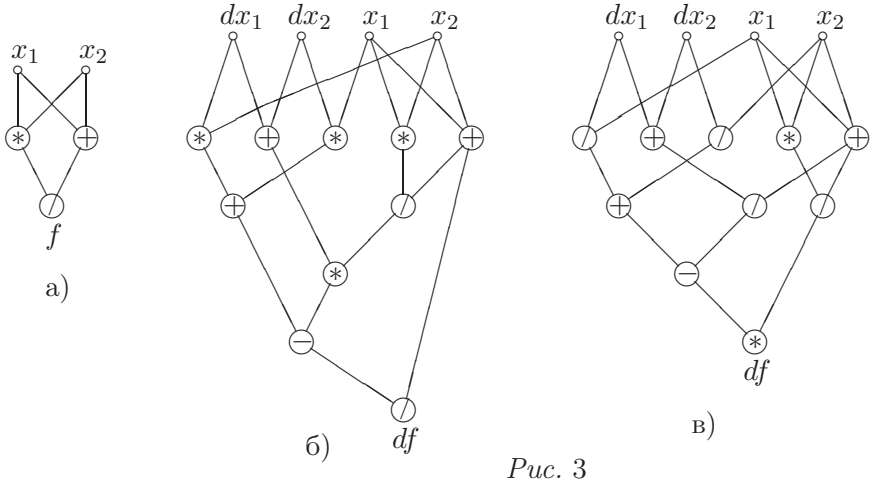


Рис. 3

Отметим, что все дополнительные элементы построенной схемы (т. е. те, которые зависят от дифференциальных входов) реализуют полные дифференциалы.

Также обратим внимание на то, что схема для дифференциала функции, которая является промежуточным звеном при построении схемы для градиента, может иметь самостоятельное значение. Если на дифференциальные входы схемы для df подать вектор констант $\bar{a} = (a_1, \ldots, a_n)$, то полученная схема будет вычислять производную $f'_{\bar{a}}$ функции f по направлению $\bar{a}$. Таким образом, $L(f'_{\bar{a}}) \leqslant L(df)$.

Пример 2. На рис. 3*а* приведён пример схемы, реализующей функцию $f(x_1, x_2) = x_1x_2/(x_1 + x_2)$, а на рис. 3*б* и 3*в* — схемы, реализующие df методом [7, 2] и методом теоремы 1 соответственно (уменьшаемое для элемента вычитания и делимое для элемента деления являются левыми входами).

§ 4. Комбинированный метод

Естественно рассмотреть следующую модификацию предложенного метода, в которой используются как преобразования Баура—Штрассена, так и преобразования теоремы 1.

Неприводимой мультипликативной подсхемой схемы над арифметическим базисом B_A назовём связную подсхему, состоящую только из мультипликативных и скалярных элементов, но не из одних только скалярных элементов, и ведущих в них дуг (т. е. для любых двух элементов подсхемы существует путь, соединяющий их по дугам, принадлежащим подсхеме — ориентация в данном определении игнорируется, и позволяется проходить через входы), входы и все элементы которой не соединяются ни с какими мультипликативными или скалярными элементами вне её. Выходами подсхемы объявляются те её элементы, которые подаются либо на выходы, либо на аддитивные элементы исходной схемы (входы не могут быть выходами, так как реализующие входы элементы по определению не относятся к подсхеме).

Таким образом, реализующую f схему можно представить как состоящую из аддитивных элементов, некоторых скалярных элементов и неприводимых мультипликативных подсхем (с выходами, входами и соединяющими дугами). По неприводимой подсхеме с λ входами, κ выходами, μ мультипликативными и σ скалярными элементами в методе Баура—Штрассена строится подсхема дифференциалов сложности $3\mu+\sigma$, а в методе теоремы 1 — подсхема сложности $\lambda+\mu+\kappa$. Сложность на дополняющих до полной схемы аддитивных и скалярных элементах совпадает в обоих методах.

Из определения неприводимой подсхемы видно, что $\lambda \leqslant \mu + 1$ и $\kappa \leqslant \mu + \sigma$. Неприводимую схему, у которой $\lambda = \mu + 1$ и $\kappa = \mu + \sigma$, назовём *минимальной*. Таким образом, метод Баура—Штрассена выигрыва-

ет только на минимальных подсхемах (выигрыш равен одному функциональному элементу). В комбинированном методе предлагается использовать преобразования схемы, описанные в теореме 1, за исключением минимальных подсхем, которые преобразуются методом Баура—Штрассена.

Пусть K — суммарное число входов и выходов по всем неприводимым мультипликативным подсхемам схемы для f, S' — число не входящих в неприводимые подсхемы скалярных элементов, а P — число минимальных подсхем. Сложность комбинированного метода построения схемы, реализующей f и df, в точности равна $2M + 2A + S + S' + K - P$, что на P меньше, чем в методе теоремы 1, и на $(2M + P - K) + (S - S')$ меньше, чем в методе Баура—Штрассена.

Если учитывать только мультипликативные операции, то метод Баура–Штрассена строит схему сложности $3M$, а комбинированный метод — схему сложности $K - P \leqslant 3M$.

§ 5. Расширение

На практике обычно нельзя считать аддитивные и мультипликативные операции равноценными (для выполнения умножения или деления, как правило, требуется большее число элементарных операций, чем для сложения, вычитания или умножения на скаляр). Поэтому для схемы над арифметическим базисом естественно рассмотреть меру сложности L_ω, в которую аддитивные элементы и скалярные умножения входят как и прежде с весом 1, а мультипликативные элементы — с весом $\omega \geqslant 1$.* Мера сложности L_∞, в которой учитываются только мультипликативные операции, получается как $\lim\limits_{\omega\to\infty}(L_\omega/\omega)$.

Теорема 2. *Пусть f — рациональная функция от n переменных. Тогда*

$$L_\omega(f, df),\ L_\omega(f, \bigtriangledown f) \leqslant 3L_\omega(f) + n.$$

Доказательство. Рассмотрим комбинированный метод построения схемы градиента из § 4. Для этого метода всегда справедливы как оценки сложности метода Баура—Штрассена, так и оценки теоремы 1.

Обозначим $A' = A + S$. Пусть $M \leqslant A' + n$. Тогда из метода Баура—Штрассена следует оценка

$$L_\omega(f, df) \leqslant (3\omega + 1)M + 2A' \leqslant 3(\omega M + A') + n = 3L_\omega(f) + n.$$

*Не является лишённым смысла также рассмотрение модели, в которой аддитивные операции выполняются сложнее мультипликативных. Так может быть, например, при использовании логарифмического представления элементов поля.

В случае $M \geqslant A' + n$ из метода теоремы 1 следует, что

$$L_\omega(f, df) \leqslant (2\omega + 1)M + (\omega + 2)A' + \omega n \leqslant 3(\omega M + A') + n = 3L_\omega(f) + n.$$

Оценка для сложности градиента получается применением леммы 3, и используется тот факт, что число мультипликативных операций не изменяется при транспозиции. Теорема 2 доказана.

В частности, при $\omega \to \infty$ получаем результат работы [7]:

$$L_\infty(f, \bigtriangledown f) \leqslant 3L_\infty(f).$$

Модель, в которой различные операции учитываются с различными весами, мы не рассматриваем в полной общности. Заметим, что на практике деление обычно выполняется сложнее, чем умножение (теоретические оценки сложности деления также, как правило, выше — для стандартного представления конечных полей, например, известные оценки сложности умножения и деления различаются по порядку). Метод теоремы 1 в модели с высокой стоимостью деления выглядит не очень выигрышно, так как использует много дополнительных операций деления. Однако в таком случае можно воспользоваться способом, указанным В. Штрассеном [17], в котором числители и знаменатели рациональных функций вычисляются отдельно, а необходимые деления выполняются только на выходах схемы (см. также [13]).

Рассмотрим одно из следствий соотношения между сложностью функции и её градиента, указанное в [7].

Следствие 2. *Пусть $C = (c_{ij})$ — квадратная невырожденная матрица порядка n. Тогда сложности вычисления обратной матрицы C^{-1} и определителя* $\det C$ *(как функций коэффициентов c_{ij}) связаны соотношением*

$$L_\omega(C^{-1}) \leqslant 3L_\omega(\det C) + n^2 + \omega.$$

Доказательство. Пусть $C^{-1} = (b_{ij})$. Тогда согласно правилу Крамера

$$b_{ij} = \frac{(\det C)'_{c_{ji}}}{\det C} = (\ln \det C)'_{c_{ji}}.$$

Таким образом, $L_\omega(C^{-1}) = L_\omega(\bigtriangledown \ln \det C)$. Из теоремы 2 следует, что

$$L_\omega(d(\ln \det C)) \leqslant 3L_\omega(\det C) + n^2 + \omega$$

(используется схема, реализующая $\det C$ и $d(\det C)$, и один элемент деления), откуда с учетом леммы 3 получаются заявленные оценки.

В частности, при $\omega = 1$ имеем

$$L(C^{-1}) \leqslant 3L(\det C) + n^2 + 1,$$

а для мультипликативной сложности справедлива оценка

$$L_\infty(C^{-1}) \leqslant 3L_\infty(\det C) + 1.$$

§ 6. Замечание о сложности второго дифференциала

Метод теоремы 1 можно распространить на построение схемы для второго дифференциала. Напомним, что второй дифференциал определяется как квадратичная форма дифференциалов переменных, т. е.

$$d^2 f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} f''_{x_i x_j} dx_i dx_j.$$

При этом для рациональных функций всегда имеет место равенство смешанных производных $f''_{x_i x_j} = f''_{x_j x_i}$.

Теорема 3. *Для рациональной m-компонентной вектор-функции f справедливы неравенства*

$$L(f, df, d^2 f) \leqslant 7L(f) + m + n; \qquad D(f, df, d^2 f) \leqslant 2,5D(f) + 1,5.$$

Доказательство. Схему, построенную при доказательстве теоремы 1, дополним подсхемой, вычисляющей вторые дифференциалы по следующим правилам:

$$d^2(u \pm v) = d^2 u \pm d^2 v; \qquad d^2(a \cdot u) = a \cdot d^2 u;$$
$$d^2(uv)/(uv) = d^2 u/u + d^2 v/v + 2(du/u)(dv/v);$$
$$d^2(u/v)/(u/v) = d^2 u/u - d^2 v/v - 2(dv/v)(du/u - dv/v).$$

Сложность такой схемы складывается из:

$A + M + S$ элементов исходной схемы;

не более $2A + n$ элементов, вычисляющих выражения вида du/u и $d^2 u/u$;

не более $2A + 5M$ элементов, участвующих в вычислении дифференциалов для аддитивных операций и выражений вида $d(u \circ v)/(u \circ v)$ и $d^2(u \circ v)/(u \circ v)$ — для мультипликативных операций;

не более $\min\{2M + 2S, 4A + 2m\}$ элементов заключительных умножений для мультипликативных элементов суммарно с элементами умножения дифференциалов на скаляры.

Число элементов в последней группе не превосходит $M+S+2A+m$, откуда получаем

$$L(f, df, d^2 f) \leqslant 7(M+A) + 2S + m + n,$$

что доказывает утверждение теоремы относительно сложности.

Для оценки глубины достаточно рассмотреть цепочку, в которой мультипликативные элементы чередуются с аддитивными. Теорема 3 доказана.

Как следствие, такие же оценки справедливы для схемы, реализующей первую и вторую производные функции f по некоторому направлению.

Отметим, что все дифференциалы $df, d^2 f, \ldots, d^k f$ и, следовательно, все производные функции f вплоть до k-го порядка по некоторому направлению могут быть вычислены схемой сложности $O(M(k)L(f))$ и глубины $O(\log k(D(f) + \log\log k))$, где $M(k)$ — порядок сложности вычисления младших k членов произведения степенных рядов с глубиной $O(\log k)$. Такая схема может быть построена методом из [16, 12], а оценка глубины следует из [16]*.

Автор благодарен научному руководителю С. Б. Гашкову за постановку задачи, полезные замечания и внимание к работе.

ЛИТЕРАТУРА

1. **Гашков С. Б.** Замечание о минимизации глубины булевых схем. // Вестник МГУ. Сер. 1. Математика. Механика. 2007. № 3. С. 7–9.

2. **Гашков С. Б., Гашков И. Б.** О сложности вычисления дифференциалов и градиентов. // Дискретная математика. 2005. Т. 17, вып. 3. С. 45–67.

3. **Гашков С. Б., Кочергин В. В.** Об аддитивных цепочках векторов, вентильных схемах и сложности вычисления степеней. // Методы дискретного анализа в теории графов и сложности. Сб. науч. тр. Вып. 52. Новосибирск: Институт математики, 1992. С. 22–40.

4. **Ким К. В., Нестеров Ю. Е., Черкасский Б. В.** Оценка трудоемкости вычисления градиента // Докл. АН СССР. 1984. Т. 275, № 6. С. 1306–1309.

5. **Ложкин С. А.** О связи между глубиной и сложностью эквивалентных формул и о глубине монотонных функций алгебры логики // Проблемы кибернетики. Вып. 38. М.: Наука, 1981. С. 269–271.

*Для инвертирования степенного ряда можно построить схему сложности $O(M(k))$ и глубины $O(\log k \log\log k)$. Метод из [16] обоснован для рядов с коэффициентами из $\mathbb{R}$ или $\mathbb{C}$, но можно показать, что этот результат справедлив для произвольного поля коэффициентов.

6. Лупанов О. Б. О вентильных и контактно-вентильных схемах // Докл. АН СССР. 1956. Т. 111, № 6. С. 1171–1174.

7. Baur W., Strassen V. The complexity of partial derivatives // Theoret. Comput. Sci. 1983. V. 22. P. 317–330. [Русский перевод: **Баур В., Штрассен Ф.** Сложность частных производных // Кибернетический сборник (Новая серия). Вып. 22. М.: Мир, 1985. С. 3–18.]

8. Bernstein D. J. The transposition principle // *http://cr.yp.to/transposition.html*.

9. Bostan A., Lecerf G., Schost E. Tellegen's principle into practice // ISSAC Conf. Philadelphia, 2003. Philadelphia: ACM Press, P. 37–44.

10. Fiduccia C. M. On the algebraic complexity of matrix multiplication. Ph. D. thesis. Brown Univ., 1973.

11. Hoover H., Klawe M., Pippenger N. Bounding fan-out in logical networks // J. Assoc. Comput. Mach. 1984. V. 31, N 1. P. 13–18.

12. Kaltofen E., Shoup V. Subquadratic-time factoring of polynomials over finite fields // Math. Comput. 1998. V. 67, N 223. P. 1179–1197.

13. Kaltofen E., Singer M. Size efficient parallel algebraic circuits for partial derivatives // IV ICCAPR Conf. Singapore, 1991. P. 133–145.

14. Linnainmaa S. Taylor expansion of the accumulated rounding error // BIT. 1976. V. 16, N 2. P. 146–160.

15. Morgenstern J. How to compute fast a function and all its derivations // SIGACT News. 1985. V. 16, N 4. P. 60–62.

16. Reif J., Tate S. Optimal size integer division circuits // SIAM J. Comput. 1990. V. 19, N 5. P. 912–925.

17. Strassen V. Vermeidung von divisionen // J. für die reine und angewandte Math. 1973. V. 264. P. 184–202.

Адрес автора:

МГУ, мех.-мат. факультет,
Ленинские горы,
119991 Москва,
Россия.

Статья поступила
26 марта 2007 г.

Переработанный вариант —
11 мая 2007 г.