Дискретная математика

том 17 выпуск 3 * 2005

УДК 519.7

О сложности вычисления дифференциалов и градиентов

© 2005 г. С. Б. Гашков, И. Б. Гашков

Получены оценки сложности схемной реализации системы дифференциалов от первого до k-го порядка произвольной элементарной функции через схемную сложность этой функции. Аналогичные оценки получены для сложности реализации матрицы Якоби и матрицы Гессе данной функции. Указаны некоторые приложения к получению оценок сложности многочленов нескольких переменных, линейных преобразований и квадратичных форм.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, гранты 02–01–10142 и 02–01–00985, и программы президента Российской Федерации поддержки ведущих научных школ, гранта НШ-1807.2003.1.

1. Введение

В [1] было показано, что сложность вычисления всех частных производных рациональной функции не больше чем в четыре раза превосходит сложность этой функции, и дано несколько примеров применения этого результата к доказательству нижних оценок сложности полиномов многих переменных. В [2] этот результат получен независимо и применен к построению программ компьютерной алгебры для вычисления градиентов функций. Согласно [3], упр. 4.6.4.71, подобный результат имеется также в [4]. Мы даем здесь другое доказательство (надеемся, более прозрачное) этого результата. Оно основано на лемме о транспонировании линейных схем, реализующих линейные отображения. Лемма была открыта независимо разными авторами (см. [3]), но осталась малоизвестной, и, вероятно, авторы работ [1, 2] о ней не знали.

Мы также указываем верхние оценки сложности схемной реализации системы дифференциалов от первого до k-го порядка для произвольной элементарной функции, выражающиеся через схемную сложность этой функции, и аналогичные оценки для реализации матрицы Якоби и матрицы Гессе данной функции и приводим некоторые приложения к оценкам сложности многочленов нескольких переменных, квадратичных форм и линейных преобразований.

2. Определения и обозначения

Рассматриваются базис

$$B = B_{ar} = \{x + y, x - y, -x - y, xy, x/y\} \cup \{ax : a \in K\},$$

где K — произвольное поле, и базис

$$B = B_{an} = B_{ar} \cup \{\exp x, \ln x, \sin x, \cos x, \arcsin x, \operatorname{arctg} x\},$$

где K — поле действительных или комплексных чисел. Все утверждения справедливы также для конечных базисов

$$B_0 = \{x + y, x - y, -x - y, xy, x/y\},$$
$$B_1 = B_0 \cup \{\exp x, \ln x, \sin x, \cos x, \arcsin x, \operatorname{arctg} x\}.$$

Пусть $F: K^n \to K^m$ есть отображение, реализуемое над базисом B и состоящее из функций $f_1, \ldots, f_m$ аргументов $X = (x_1, \ldots, x_n)$. Обозначим dF его дифференциал, то есть отображение $K^n \times K^n \to K^m$, состоящее из функций df_i с аргументами $dx_1, \ldots, dx_n, x_1, \ldots, x_n$ и линейное относительно формальных переменных $dx_1, \ldots, dx_n$ (дифференциалов переменных $x_1, \ldots, x_n$). Обозначим

$$J(F, X) = \left(\frac{\partial f_i}{\partial x_j}\right)$$

его матрицу Якоби, составленную из частных производных функций f_i в произвольной точке $X = (x_1, \ldots, x_n)$ из области определения отображения F. Отображение $F: K^n \to K^m$ называем также (m, n)-вектор-функцией. Предполагаем, что все ее переменные существенны, так как несущественные переменные можно удалить.

Стандартным образом (так же, как в [1, 5, 6, 7]) определяется понятие схемы над данным базисом B. Как обычно, сложность схемы — это число ее элементов. Глубина схемы — это длина максимальной цепи элементов, идущей от входов к выходам схемы. Сложность схемы S обозначаем $C_B(S)$, а глубину — $D_B(S)$. Сложностью вектор-функции F называем минимальную сложность вычисляющей ее схемы и обозначаем $C_B(F)$. Аналогично определяем глубину $D_B(F)$.

Многочлены и рациональные функции рассматриваем над произвольным полем.

Все реализуемые в базисе B_{an} функции являются элементарными, и обратно. Элементарные функции могут быть не всюду определенными.

Сложность минимальной схемы, реализующей все элементы якобиевой матрицы, обозначаем $C_B(J(F))$. В случае одновременной реализации $J(F)$ и F используем обозначение $C_B(F, J(F))$. Если $m = 1$, то вместо F и $J(F)$ пишем, как обычно, f и $\operatorname{grad} f$.

3. Лемма о сложности линейного отображения

Пусть

$$B_l = \{x + y, x - y, -x - y\} \cup \{ax : a \in F\}$$

— линейный базис над произвольным полем K. Тогда для любого линейного преобразования $X \to AX$, где X — n-мерный вектор-столбец над K, а A — $m \times n$ матрица над этим полем без нулевых строк и столбцов, справедливо следующее утверждение.

Лемма 1. *Справедливо равенство*

$$L_{B_l}(A) + m = L_{B_l}(A^T) + n,$$

где $L_{B_l}(A)$ — сложность реализации линейного преобразования с матрицей A в базисе B_l, а A^T — транспонированная матрица A.

Для доказательства леммы нужны некоторые определения и еще две леммы. Приводимое доказательство взято из [6].

Пусть S — произвольная схема в линейном базисе B_l, реализующая линейное преобразования с матрицей A. Так как некоторые элементы схемы реализуют вычитания, пометим ребра, ведущие в соответствующие входы этих элементов символом минус единица. Назовем транспонированной схемой к S такую схему S^T, которая получается из S сменой ориентации ребер на противоположную и заменой входов на выходы и соответственно выходов на входы (в том числе и у элементов схемы). Тогда аддитивные элементы, имеющие степень ветвления $d > 2$, превращаются в аддитивные элементы с d входами. Так как такие элементы не входят в базис B_l, они заменяются на эквивалентные схемы с d входами, состоящие из $d-1$ аддитивных элементов (некоторые из которых могут быть вычитаниями в соответствии с распределением символов -1 по упомянутым d входным ребрам), и имеющие вид бинарного корневого дерева с $\lceil \log_2 d \rceil$ ярусами. Элементы с $d = 2$ заменяются на аддитивные элементы. Элементы с $d = 1$ заменяются на элементы скалярного умножения на 1, которые потом из схемы удаляются. С элементами скалярного умножения поступаем аналогично, но из схемы их не удаляем, а присоединяем к выходу вставленной перед ним подсхемы (в случае $d = 1$ она отсутствует, то есть в этом случае элемент просто остается без изменения).

Назовем вентильной (m,n)-схемой произвольный ориентированный граф без ориентированных циклов с n входными и m выходными вершинами, в котором ребра ориентированы от входов к выходам и некоторые ребра помечены символами из поля K. Для любого ориентированного пути, соединяющего две вершины схемы, произведение всех меток его ребер назовем проводимостью пути (если меток на нем нет, то проводимость равна 1). Функцией проводимости между упорядоченной парой вершин назовем сумму проводимостей всех ориентированных путей, соединяющих эти вершины. Вентильная (m,n)-схема реализует $m \times n$ матрицу A, если функция проводимости между i-м входом и j выходом равна a_{ji}.

Если в вентильной (m,n)-схеме V, реализующей матрицу A, изменить ориентацию всех ребер на противоположную и поменять местами входы и выходы, то полученная транспонированная схема V^T будет реализовывать транспонированную матрицу A^T.

Любой схеме S с n входами и m выходами в базисе B_l сопоставим вентильную (m,n)-схему V, пометив ребра идущие в вершины, соответствующие скалярным элементам, соответствующими константными множителями. Следующая лемма легко доказывается по индукции.

Лемма 2. *Если схема S реализует линейное преобразование с $m \times n$ матрицей A над полем K, то соответствующая ей вентильная (m,n)-схема V реализует матрицу A.*

Из леммы 2 выводится следующее утверждение.

Лемма 3. *Пусть S — произвольная схема в линейном базисе B_l с n входами и m выходами и S^T — транспонированная к ней схема. Тогда сложности схем S и S^T связаны соотношением*

$$L_{B_l}(S) + m = L_{B_l}(S^T) + n,$$

и множества скалярных элементов в обеих схемах совпадают. Если схема S реализует линейное преобразование с матрицей A, то схема S^T реализует линейное преобразование, определяемое транспонированной матрицей A^T.

Доказательство. Возьмем соответствующую схеме S вентильную схему V. Согласно лемме 2 схема V реализует матрицу A, следовательно, схема V^T реализует матрицу

A^T. Но вентильная схема, соответствующая схеме S^T, реализует ту же матрицу, что и V^T. Применяя лемму 2, получаем, что схема S^T реализует линейное преобразование, определяемое матрицей A^T.

Число u скалярных элементов при транспонировании схемы не меняется. Если S имела l аддитивных элементов, то число r ребер в S равно $2l + u + m$ (с учетом ребер, направленных в выходы). Такое же число ребер будет и в транспонированной схеме до замены многовходовых аддитивных элементов на двувходовые. Число двувходовых элементов, которые получаются из элемента (или входа схемы) со степенью ветвления d, равно $d - 1$, поэтому общее число получившихся двувходовых элементов равно $r - (l + u + n) = l + m - n$, поэтому

$$L_{B_l}(S^T) = u + l + m - n = L_{B_l}(S) + m - n.$$

Замечание 1. Если базис B_l не содержит элементов $-x - y$, то в схеме S^T иногда появляются элементы $-x$, прямо соединенные с выходами схемы. Если характеристика поля $\operatorname{char} K = 2$, то такие элементы не появляются.

4. О сложности вычисления функций одновременно с их дифференциалами и матрицами Якоби

Следующая теорема обобщает соответствующие результаты [1, 2]. Число мультипликативных элементов в схеме S обозначим $M(S)$.

Теорема 1. *Существует схема S в базисе $B = B_{ап}$, вычисляющая (m, n)-вектор-функцию F, такая, что*

$$C(S) = C_B(F),$$
$$C_B(F, J(F)) \leqslant (m + 3)C(S) + (2m - 2)M(S) + m(m - n).$$

Справедливы также соотношения

$$C_B(F, dF) \leqslant 4C(S),$$
$$D_B(F, dF) \leqslant 3D(S).$$

В частности, при $m = 1$

$$C_B(f, df) \leqslant 4C_B(f),$$
$$C_B(f, \operatorname{grad} f) \leqslant 4C_B(f) + 1 - n.$$

Доказательство. Число элементов cos, sin, arcsin в схеме S обозначим $CSA(S)$, число элементов arctg обозначим $A(S)$, а общее число всех этих элементов обозначим $T(S)$.

Линейную (в том смысле, что все ее элементы линейно зависят от дифференциалов $dx_1, \ldots, dx_n$) относительно входов $dx_1, \ldots, dx_n$ схему S_{dF}, реализующую дифференциал dF, строим по индукции параллельно построению схемы S_F. В качестве базы индукции возьмем случай, когда $S_F = S$. Пусть очередной элемент схемы S_F реализует функцию $w = u \circ v$, где u, v — функции, реализуемые предшествующими элементами. Допустим также, что элементы, реализующие дифференциалы du, dv, в схему S_{dF} уже добавлены.

Так как для дифференциалов функций многих переменных справедливы формулы (см. [8, 9])

$$d(u \pm v) = du \pm dv, \quad d(uv) = u\,dv + v\,du, \quad d(u/v) = \frac{1}{v}\left(du - \frac{u}{v}dv\right),$$

для реализации дифференциала dw надо добавить в схему S_{dF} один аддитивный элемент, и в случае, если элемент $\circ$ был мультипликативным, еще два мультипликативных элемента (один из них будет делением в случае, если $\circ$ — деление), на один из входов у каждого из которых будет подаваться выход подходящего элемента схемы S_{dF}.

Если очередной элемент $f \in \{\exp, \ln, \sin, \cos, \arcsin, \operatorname{arctg}\}$ схемы S_F реализует функцию $w = f(u)$, то $dw = f'(u)\,du$, и в частности, при $f \in \{\exp, \ln, \sin, \cos, \arcsin\}$, для реализации w надо добавить в схему $S_d F$ не более чем одно скалярное (в том смысле, что от дифференциалов $dx_1, \ldots, dx_n$ зависит только один из входов реализующего эту операцию элемента) умножение или деление (появление знака минус в тригонометрическом случае dw устраняем, перенося знак минус с входа элемента умножения на его выход или заменяя сложение вычитанием и наоборот), и в нелинейную схему S_F в тригонометрическом случае надо добавить не более одного вспомогательного элемента, параллельно расположенного с элементом u, например, в случае $f = \arcsin$ таким элементом будет $\cos$ согласно формуле

$$dw = f'(u)\,du = \frac{du}{\sqrt{1-u^2}} = \frac{du}{\cos w}.$$

Случай $f = \operatorname{arctg}$ аналогичен рассмотренному, но в нем согласно формуле

$$dw = f'(u)\,du = \frac{du}{1+u^2} = \cos^2 w\,du$$

используем два вспомогательных элемента.

Заметим, что глубина элемента dw удовлетворяет неравенству

$$\operatorname{dep}(dw) \leqslant \max\{\operatorname{dep}(w), \operatorname{dep}(du), \operatorname{dep}(dv)\} + 3,$$

где

$$\operatorname{dep}(w) = \max\{\operatorname{dep}(u), \operatorname{dep}(v)\} + 1.$$

Отсюда по индукции выводим, что

$$\operatorname{dep}(dw) \leqslant 3\operatorname{dep}(w),$$

и построенная схема S_{dF} с входами $dx_1, \ldots, dx_n$ реализует линейное преобразование с матрицей $J(F, X)$ сложностью

$$C(S_{dF}) \leqslant C(S) - T(S) + 2M(S) + T(S) = C(S) + 2M(S)$$

и глубиной

$$D(S_{dF}) \leqslant 3D(S).$$

Схема $S(F, dF)$, которая состоит из схемы S_F сложности $C(S) + \Delta(S)$, где

$$\Delta(S) \leqslant CSA(S) + 2A(S) \leqslant 2T(S),$$

и схемы S_{dF}, вычисляет одновременно функцию F и ее дифференциал dF и удовлетворяет неравенствам теоремы.

Для получения оценки сложности одновременной реализации F и ее якобиевой матрицы $J(F)$ в построенной схеме $S(F, dF)$ заменим подсхему S_{dF} на транспонированную схему S_{dF}^T с дифференциальными входами $dy_1, \ldots, dy_m$ и n выходами. Согласно лемме 1, эта схема реализует совместно со схемой S_F линейное преобразование с матрицей $J(F)^T$ и имеет сложность $L(S) + m - n$, где $L(S)$ есть сложность схемы S_{dF}. Расположив параллельно m экземпляров схемы S_{dF}^T с непересекающимися дифференциальными входами и общими остальными входами, связанными со схемой S_F, и подавая на входы i-й схемы соответствующий единичный вектор $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, получаем схему сложности не выше $m(L(S) + m - n) + C(S) + \Delta(S)$.

4.1. Замечания

Если базис не содержит arctg, то

$$\Delta(S) \leqslant T(S) \leqslant C(S) - M(S), \qquad C_B(F, dF) \leqslant 3C(S) + M(S),$$
$$C_B(F, J(F)) \leqslant (m + 2)C(S) + (2m - 1)M(S) + m(m - n).$$

Если

$$B = B_{sr} = B_e \cup \{\sin, \arcsin, \operatorname{arctg}, \sqrt{\ }\},$$

то

$$\Delta(S) \leqslant 3T(S) \leqslant 3(C(S) - M(S)), \qquad C_B(F, dF) \leqslant 5C(S),$$
$$C_B(F, J(F)) \leqslant (m + 4)C(S) + (2m - 3)M(S) + m(m - n).$$

Если

$$B = B_e = B_{ar} \cup \{\exp, \ln\},$$

то $\Delta(S) = 0$ и

$$C_B(F, dF) \leqslant 2C(S) + 2M(S),$$
$$C_B(F, J(F)) \leqslant m(m - n) + (3m + 1)C(S).$$

Если базис B не содержит деления, то глубина $D_B(F) \leqslant 2D(S)$. Для базиса с делением можно получить ту же оценку, если применять формулу

$$d(u/v) = \frac{du}{v} - \frac{(u/v)}{v} dv,$$

но тогда оценка дополнительной сложности заменится на оценку

$$\Delta(S) \leqslant 2T(S) + D(S) \leqslant 2T(S) + M(S),$$

где $D(S)$ — число элементов деления.

Оценка для $C_B(F, J(F))$ нетривиальна только для $m < n$. Для $m = n$ она совпадает с вытекающей из первого неравенства теоремы оценкой $nL(S) + C(S) + \Delta(S)$. Из частного случая $m = 1$ можно вывести оценку и для общего случая, но чуть менее точную.

5. О сложности второго дифференциала и матрицы Гессе

Второй дифференциал $d^2 f$ — это квадратичная форма

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} dx_i dx_j$$

от дифференциалов переменных.

Следствие 1. *Существует схема S в базисе $B = B_{an}$, вычисляющая f и такая, что*

$$C(S) = C_B(f),$$
$$C_B(f, \operatorname{grad} f, df, d^2 f) \leqslant 12C(S) + 2M(S) + 1.$$

Доказательство. Заметим, что

$$d^2 f = \sum_{i=1}^{n} d\,(\partial f/\partial x_i)\, dx_i.$$

Поэтому к любой схеме, реализующей одновременно f и $d(\operatorname{grad} f)$, можно прибавить n элементов умножения и $n - 1$ элемент сложения (увеличив глубину схемы не более чем на $1 + \log_2 n$).

Из доказательства теоремы 1 видно, что схему S, реализующую f в базисе B, можно преобразовать в схему S_F, реализующую одновременно f и $\operatorname{grad} f$ со сложностью

$$C(S_F) \leqslant L(S) + C(S) + \Delta(S) + 1 - n,$$

где

$$L(S) = C(S) + 2M(S),$$
$$\Delta(S) \leqslant CSA(S) + 2A(S) \leqslant 2T(S).$$

Ее мультипликативная сложность равна

$$M(S_F) = 3M(S) + U(S) + A(S),$$

где $U(S)$ есть число всех унарных функциональных элементов в S. Действительно, каждый из этих элементов порождает один элемент скалярного умножения или деления, элемент arctg порождает также один элемент возведения в квадрат, а каждый из $M(S)$ элементов умножения или деления порождает два таких же скалярных элемента в ее линейной части. Для построения схемы S_F линейная часть схемы S транспонируется, но число скалярных элементов и число $\Delta(S)$ не меняются. Ясно, что

$$U(S_F) \leqslant U(S) + T(S),$$
$$T(S_F) \leqslant T(S) + CSA(S),$$

так как только элементы sin, cos, arcsin порождают новые тригонометрические элементы при дифференцировании.

Применим теорему 1 к построенной схеме S_F, реализующей $(n+1, n)$-вектор-функцию $F = (f, \operatorname{grad} f)$. Согласно этой теореме, существует схема $S_{F,dF}$ в базисе B, вычисляющая одновременно F и $dF = (df, d(\operatorname{grad} f))$, со сложностью

$$C(S_{F,dF}) \leqslant L(S_{dF}) + C(S_F) + \Delta(S_F),$$

где

$$\begin{aligned} C(S_F) &= L(S) + C(S) + \Delta(S) + 1 - n \\ &= 2C(S) + 2M(S) + CSA(S) + 2A(S) + 1 - n, \\ L(S_{dF}) &= C(S_g) + 2M(S_g) \\ &= L(S) + C(S) + \Delta(S) + 1 - n + 6M(S) + 2U(S) + 2A(S) \\ &= 2C(S) + 8M(S) + CSA(S) + 2U(S) + 4A(S) + 1 - n. \end{aligned}$$

Формально

$$\Delta(S_F) \leqslant 2T(S_g) \leqslant 2T(S) + 2CSA(S),$$

но фактически новые унарные элементы появляются только при дифференцировании унарных элементов, добавленных при преобразовании схемы S в схему S_g, а так как число тригонометрических элементов равно $CSA(S)$, справедлива оценка

$$\Delta(S_F) \leqslant CSA(S).$$

Окончательно получаем, что

$$\begin{aligned} C(S_{F,dF}) &\leqslant 4C(S) + 10M(S) + CSA(S) + 4A(S) + 2U(S) + 2A(S) + CSA(S) + 2 - 2n \\ &\leqslant 4C(S) + 10M(S) + 6T(S) + 2U(S) + 2 - 2n \\ &\leqslant 12C(S) + 2M(S) + 2 - 2n, \end{aligned}$$

откуда

$$C_B(f, \operatorname{grad} f, df, d^2 f) \leqslant 12C_B(f) + 2M(S) + 1.$$

Матрица Гессе функции f от n переменных есть $n \times n$ матрица

$$H(f) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right),$$

составленная из частных производных второго порядка.

Обозначим $C_B(f, H(f))$ сложность минимальной схемы в базисе B, реализующей все различные элементы этой матрицы одновременно с самой функцией f.

Следствие 2. *Для базиса* $B = B_{an}$

$$C_B(f, \operatorname{grad} f, H(f)) \leqslant (10n + 4)C_B(f) + 1 - n^2.$$

Доказательство. Преобразуем построенную выше схему, оставив без изменения нелинейную часть, расположив параллельно n экземпляров ее линейной части и подставив вместо ее дифференциальных входов все возможные единичные векторы. Полученная схема вместо дифференциала градиента будет вычислять матрицу его коэффициентов, то есть матрицу Гессе. Кроме того, она по-прежнему будет вычислять функцию f и ее

градиент. Подстановка констант 0 и 1 не увеличивает сложность. Поэтому сложность построенной схемы оценивается сверху величиной

$$\begin{aligned} C(S_F) + \Delta(S_F) + nL(S_{dF}) &\leqslant 2C(S) + 2M(S) + 2CSA(S) + 2A(S) + 1 - n \\ &\quad + n(2C(S) + 8M(S) + 4T(S) + 2U(S) + 1 - n) \\ &\leqslant 2C(S) + 2M(S) + 2T(S) + 1 - n \\ &\quad + n(2C(S) + 8M(S) + 4T(S) + 2U(S) + 1 - n) \\ &\leqslant 2(n+1)C(S) + 2(4n+1)M(S) + 2(2n+1)T(S) \\ &\quad + 2nU(S) + 1 - n^2 \\ &\leqslant 2(n+1)C(S) + 2(4n+1)M(S) \\ &\quad + 2(3n+1)U(S) + 1 - n^2 \\ &\leqslant (10n+4)C_B(f) + 1 - n^2. \end{aligned}$$

Следствие доказано.

Заметим, что точность по порядку доказанной оценки показывает пример с функцией $f = x_1 \ldots x_n$.

6. О сложности вычисления разложения элементарной функции в ряд Тейлора

Разложение гладкой функции n переменных в ряд Тейлора до членов k-го порядка в данной точке $(x_1, \ldots, x_n)$ выражается через кратные дифференциалы в этой точке следующим образом (без остаточного члена):

$$f(x_1 + dx_1, \ldots, x_n + dx_n) = \sum_{i=0}^{k} \frac{d^i f}{i!}, \qquad \frac{d^0 f}{0!} = f$$

(см., например [8, 9]).

Сверткой двух n-мерных векторов a, b называется n-мерный вектор $c = a * b$ с координатами

$$c_k = \sum_{j=1}^{k} a_j b_{k-j+1}, \qquad 1 \leqslant k \leqslant n.$$

Сложность k-мерной свертки обозначим $C(k)$. Очевидно, что $C(k) = O(k^2)$. Младшие коэффициенты произведения двух многочленов с векторами коэффициентов a и b вычисляются в точности по формулам свертки. Отсюда следует, что справедлива оценка $C(n+1) \leqslant M(n)$, где $M(n)$ — сложность умножения многочленов степени n над тем же полем, над которым выполняется сверка (известно, что для любого конечного поля $M(n) = O(n \log n \log\log n)$, а для поля действительных или комплексных чисел $M(n) = O(n \log n)$, см. [11, 12]).

Теорема 2. *Для любой элементарной функции f от n переменных и любой вычисляющей ее схемы S сложности $C(S)$ можно построить схему S_k, вычисляющую одновременно $f, df, \ldots, d^k f$ со сложностью*

$$C_B(S_k) = O(M(k)C_B(S)).$$

Доказательство. Для кратных дифференциалов функций многих переменных справедливы соотношение линейности и формула Лейбница (см., например [8, 9])

$$d^k(u \pm v) = d^k u \pm d^k v,$$
$$\frac{d^k(uv)}{k!} = \sum_{i=0}^{k} \frac{d^i u}{i!} \frac{d^{k-i} v}{(k-i)!}.$$

Поэтому вектор $(w, dw, \ldots, d^k w/k!)$, где $w = u \pm v$ или $w = uv$, вычисляется по векторам $(u, du, \ldots, d^k u/k!)$, $(v, dv, \ldots, d^k v/k!)$ либо с помощью операции сложения–вычитания векторов, либо их свертки.

В случае деления $w = u/v$ известны детерминантные формулы для вычисления $d^k w$ (см. [13]), но из них непосредственно не удается получить хорошей оценки сложности этого вычисления. В этом случае проще применить к равенству $u = wv$ правило Лейбница, тогда

$$\frac{d^k u}{k!} = \sum_{i=0}^{k} \frac{d^i w}{i!} \frac{d^{k-i} v}{(k-i)!},$$

откуда

$$\frac{d^k w}{k!} = \frac{1}{v}\left(\frac{d^k u}{k!} - \sum_{i=0}^{k-1} \frac{d^i w}{i!} \frac{d^{k-i} v}{(k-i)!}\right),$$

и применяя рекурсивно эту формулу, вычисляем $(w, dw, \ldots, d^k w/k!)$ по данным $(u, du, \ldots, d^k u/k!)$, $(v, dv, \ldots, d^k v/k!)$ с помощью $(k+1)^2$ операций сложения, умножения и деления.

Эту оценку можно улучшить.

Лемма 4. *Пусть даны* $(u, du, \ldots, d^k u/k!), (v, dv, \ldots, d^k v/k!)$. *Тогда сложность вычисления* $(w, dw, \ldots, d^k w/k!)$, *где* $w = u \circ v$ *есть результат применения арифметической операции к элементарным функциям* u, v, *есть* $O(M(k))$.

Доказательство. Если операция $\circ$ есть умножение, то лемма уже доказана. Так как $w = u/v = u(1/v)$, нужно доказать, что если дан вектор $(v, dv, \ldots, d^k v/k!)$, то сложность вычисления $(w, dw, \ldots, d^k w/k!)$, где $w = 1/v$, есть $O(M(k))$.

Применяя к равенству $1 = wv$ правило Лейбница, находим, что

$$\delta_k = \sum_{i=0}^{k} \frac{d^i w}{i!} \frac{d^{k-i} v}{(k-i)!},$$

где $\delta_k = 1$ при $k = 0$ и $\delta_k = 0$ при $k > 0$. Из сделанного перед леммой замечания следует, что свертка векторов $W = (w, dw, \ldots, d^k w/k!)$ и $V = (v, dv, \ldots, d^k v/k!)$ равна вектору $(\delta_0, \ldots, \delta_k.)$ Поэтому произведение многочленов

$$W(x) = W_0 x^k + \ldots + W_k, \quad W_i = \frac{d^i w}{i!}, \quad i = 0, 1, \ldots, k,$$
$$V(x) = V_0 x^k + \ldots + V_k, \quad V_i = \frac{d^i v}{i!}, \quad i = 0, 1, \ldots, k,$$

имеет вид $x^{2k} + R(x)$, где степень многочлена $R(x)$ меньше k. Отсюда следует, что при делении x^{2k} на $V(x)$ получается частное $W(x)$ и остаток $-R(x)$. Но сложность деления x^{2k} на многочлен $V(x)$ степени k без вычисления остатка равна $3M(k) + O(k)$ (см. [10, 11]). Поэтому, если дан вектор V, то сложность вычисления вектора W не превосходит $3M(k) + O(k)$.

Лемма доказана.

Если $w(x_1, \ldots, x_m) = f(u)$, где f есть гладкая функция одной переменной, а $u = u(x_1, \ldots, x_m)$ есть гладкая функция m переменных, то дифференциал

$$d^n f(u(x)) = d^n w(x)$$

вычисляется по формуле Арбогаста–Бруно:

$$d^n w! = \sum_{k=1}^{n} A_{n,k}(du, \ldots, d^n u) f_k,$$

где $f_k = f^{(k)}(u)$ есть производная порядка k, вычисленная в точке u, а полиномы $A_{n,k}(g_1, \ldots, g_n)$ есть полиномы

$$\sum_{\substack{k_1+\ldots+k_n=k \\ k_1+\ldots+nk_n=n}} n! \prod_{i=1}^{n} \frac{(g_i/i!)^{k_i}}{k_i!}$$

с натуральными коэффициентами от формальных переменных

$$(g_1, \ldots, g_n) = (du, \ldots, d^n u)$$

(см. [3], [13].) Эти полиномы называются однородными полиномами Белла, а полином

$$A_n(f, g) = \sum_{k=1}^{n} A_{n,k}(g_1, \ldots, g_n) f_k$$

от формальных переменных $f_k, g_k, k = 1, \ldots, n$, называется полиномом Белла. Сам Белл [15] имел дело с полиномами

$$Y_n(y; g) = A_n(y; g) = \sum_{k=1}^{n} y^k A_{n,k}(g),$$

которые тоже естественно называть полиномами Белла.

Но формула Бруно непосредственно не дает быстрого алгоритма для вычисления $d^n w$, так как общее число мономов в полиноме Белла равно числу $p(n)$ разбиений n на слагаемые, которое растет экспоненциально (см. [13]). Однако справедлива следующая лемма.

Лемма 5. *Сложность системы полиномов Белла*

$$A_k(f_1, \ldots, f_k; g_1, \ldots, g_k), \qquad k = 1, \ldots, n,$$

есть $O(M(n)(n \log n)^{1/2})$, *поскольку она совпадает с точностью до слагаемого* $O(n)$ *со сложностью вычисления первых* n *коэффициентов суперпозиции степенных рядов.*

Сложность системы полиномов Белла

$$Y_k(y; g_1, \dots, g_k), \qquad k = 1, \dots, n,$$

есть $O(M(n))$.

Доказательство. Для произвольных последовательностей $f_1, f_2, \dots, g_1, g_2, \dots$ формальных переменных определим экспоненциальные производящие функции

$$A(f; g; x) = \sum_{n=1}^{\infty} A_n(f_1, \dots, f_n; g_1, \dots, g_n)\frac{x^n}{n!},$$

$$A_k(g; x) = \sum_{n=k}^{\infty} A_{n,k}(g_1, \dots, g_n)\frac{x^n}{n!},$$

$$G(x) = \sum_{n=1}^{\infty} g_n \frac{x^n}{n!},$$

$$Y(y; g; x) = \sum_{n=0}^{\infty} Y_n(y; g_1, \dots, g_n; x)\frac{x^n}{n!}.$$

Согласно [14] (см. п. 5.3) справедливо равенство

$$A_k(g; x) = \frac{G(x)^k}{k!}.$$

Это равенство легко получить, применяя к его правой части полиномиальную формулу, группируя вместе слагаемые с одинаковым множителем x^n и пользуясь данным выше определением однородных полиномов Белла $A_{n,k}(g_1, \dots, g_n)$. Умножая обе части доказанного равенства на f_k и складывая, после перестановки порядка суммирования

$$\sum_{k=1}^{\infty} f_k \frac{G(x)^k}{k!} = \sum_{k=1}^{\infty} f_k A_k(g; x) = \sum_{k=1}^{\infty} f_k \sum_{n=k}^{\infty} A_{n,k}(g)\frac{x^n}{n!}$$

$$= \sum_{n=1}^{\infty}\sum_{k=1}^{n} f_k A_{n,k}(g)\frac{x^n}{n!} = \sum_{n=1}^{\infty} A_n(f; g)\frac{x^n}{n!} = A(f; g; x)$$

получаем формулу

$$A(f; g; x) = \sum_{n=1}^{\infty} A_n(f; g)\frac{x^n}{n!} = \sum_{n=1}^{\infty} f_n \frac{G(x)^n}{n!},$$

в существенном совпадающую с формулой (5) п. 5.2 в [14]. Полагая в ней $f_k = y^k$, $k = 1, 2 \dots$, получаем формулу

$$Y(y; g; x) = \sum_{n=0}^{\infty} Y_n(y; g_1, \dots, g_n; x)\frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{(xG(y))^n}{n!} = \exp xG(y),$$

имеющуюся в п. 8 главы 2 в [13]. Поэтому для вычисления системы многочленов

$$A_k(f_1, \dots, f_k; g_1, \dots, g_k)/k!, \qquad k = 1, \dots, n$$

(коэффициентов степенного ряда $A(f;g;x)$) достаточно вычислить суперпозицию степенных рядов

$$\sum_{k=1}^{\infty} f_k \frac{y^k}{k!}, \qquad G(x),$$

или, что равносильно, суперпозицию $F_n(G_n(x))$ многочленов n-й степени

$$F_n(x) = \sum_{k=1}^{n} f_k \frac{x^k}{k!}, \qquad G_n(x) = \sum_{k=1}^{n} g_k \frac{x^k}{k!}$$

по модулю x^{n+1}. Так как система чисел $k!$, $k = 1, \ldots, n$, вычисляется в базисе B_{ar} со сложностью $O(n)$, для вычисления системы многочленов

$$A_k(f_1, \ldots, f_k; g_1, \ldots, g_k), \qquad k = 1, \ldots, n,$$

можно построить схему в базисе B_{ar} с входами f_k, g_k, $k = 1, \ldots, n$, сложности $S(n) + O(n)$, где $S(n)$ — сложность вычисления первых n коэффициентов суперпозиции степенных рядов. Из тождества

$$\sum_{n=1}^{\infty} A_n(f_1, 2!\, f_2, \ldots, n!\, f_n; g_1, 2!\, g_2 \ldots n!\, g_n) x^n = \sum_{n=1}^{\infty} f_n(g(x))^n,$$

где

$$g(x) = \sum_{n=1}^{\infty} g_n x^n,$$

следует, что

$$S(n) \leqslant L_{B_{ar}}(A_k(f_1, \ldots, f_k; g_1, \ldots, g_k), k = 1, \ldots, n) + O(n).$$

Поэтому

$$C_{B_{ar}}(A_k(f_1, \ldots, f_k; g_1, \ldots, g_k), k = 1, \ldots, n) = S(n) + O(n).$$

В [16] (см. также [11]) показано, что

$$S(n) = O(M(n)(n \log n)^{1/2}).$$

В [17] показано, что сложность вычисления суперпозиции степенных рядов в случае, когда внешний ряд есть экспонента $\exp(x)$ или логарифм, равна $O(M(n))$. Поэтому из тождества

$$\sum_{n=0}^{\infty} Y_n(y; g_1, \ldots, g_n; x) \frac{x^n}{n!} = \exp xG(y)$$

аналогичным образом выводится оценка

$$C_{B_{ar}}(Y_k(y; g_1, \ldots, g_k), k = 1, \ldots, n) = O(M(n)).$$

Замечание 2. Полагая $y = 1$, из предыдущего равенства выводим, что

$$C_{B_{ar}}\left(\sum_{k=1}^{m} A_{m,k}(y; g_1, \ldots, g_m), m = 1, \ldots, n\right) = O(M(n)).$$

Можно показать, что сложность системы однородных полиномов Белла

$$A_{m,k}(g_1,\dots,g_m),\quad m=1,\dots n,\quad k=1,\dots,m,$$

не превосходит

$$C(n-1)+C(n-2)+\dots+C(1)+n(n-1)=O(nM(n)).$$

Из леммы 5 следует, что при данных $(g_1,\dots,g_n)$, $(f_1,\dots,f_n)$, где

$$g_k=d^k u,\quad f_k=f^{(k)}(u),\quad k=1,\dots,n,$$

сложность вычисления вектора $(w,dw,\dots,d^n w)$, где

$$w(x)=f(u(x)),$$

есть $O(M(n)(n\log n)^{1/2})$. Доказательство состоит в применении формулы Бруно

$$d^n w!=A_n(f;g).$$

При вычислении $(w,dw,\dots,d^n w/n!)$ по данным $(u,du,\dots,d^n u/n!)$ оценка сложности возрастет очевидно только на $O(n)$.

Но в нужных далее частных случаях приведенную выше оценку можно улучшить.

Лемма 6. *Пусть даны* $(u,du,\dots,d^n u/n!)$. *Тогда сложность вычисления* $(w,dw,\dots,d^n w/n!)$, *где* $w=f(u)$, $f\in\{\ln,\exp,\sin,\cos,\operatorname{arctg},\arcsin\}$ *есть* $O(M(n))$.

Доказательство. Если $f=\ln$ или $f=\operatorname{arctg}$, то $df(u)=du/u$ или $df(u)=du/(1+u^2)$. Так как

$$\frac{d^n f(u)}{n!}=\frac{1}{n}\frac{d^{n-1}(df(u))}{(n-1)!},$$

сложность вычисления $(f(u),df(u),\dots,d^n f(u)/n!)$ есть $O(M(n))$ согласно лемме 4.

Замечание о сумме однородных полиномов Белла и равенство

$$\frac{d^n\exp(u)}{n!}=\frac{1}{n}\frac{1}{(n-1)!}\sum_{k=1}^{n}A_{n,k}\exp(u),$$

доказывает лемму в случае $f=\exp$.

В случае $f=\sin$ из равенства

$$d^n\sin u=\sum_{k=1}^{n}A_{n,k}\sin(u+k\pi/2),$$

следует, что если вычислены четыре последовательности сумм

$$B_{n,i}=\sum_{i=k\mod 4}A_{n,k},\qquad i=0,1,2,3,$$

то $(w,dw,\dots,d^n w/n!)$, где $w=\sin u$, вычисляется со сложностью $O(n)$ по формулам

$$\frac{d^k w}{k!}=\frac{1}{k}\frac{1}{(k-1)!}((B_{k,0}-B_{k,2})\sin u+(B_{k,1}-B_{k,3})\cos u).$$

Суммы $B_{n,i}$ можно вычислить со сложностью $O(n)$, если вычислены $Y_n(\varepsilon^i)$, $i = 0, 1, 2, 3$, где

$$Y_n(1) = \sum_{i=0}^{3} B_{n,i}, \qquad Y_n(\varepsilon) = B_{n,0} - B_{n,2} + \varepsilon(B_{n,1} - B_{n,3}),$$

$$Y_n(\varepsilon^2) = B_{n,0} + B_{n,2} - B_{n,1} - B_{n,3}, \qquad Y_n(\varepsilon^3) = B_{n,0} - B_{n,2} + \varepsilon(B_{n,3} - B_{n,1}),$$

и ε есть примитивный корень степени 4 из 1. Но при любом a последовательность $Y_m(a)$, $m = 1, \ldots, n$, можно вычислить со сложностью $O(M(n))$, как показано в лемме 5, поэтому утверждение доказано.

Аналогично поступаем в случае $f = \cos$.

В случае $f = \arcsin$ очевидно, что $df(u) = du/\sqrt{1-u^2}$, $\sqrt{x} = \exp((1/2)\ln x)$. Пусть даны $(u, du, \ldots, d^n u/n!)$. Применяя лемму 4 и полученные выше оценки сложности вычисления $(w, dw, \ldots, d^n w/n!)$, где $w = g(u)$, $g \in \{\exp, \ln\}$, получаем нужную оценку сложности $(w, dw, \ldots, d^n w/n!)$, где $w = f(u)$. Лемма доказана.

Доказательство теоремы 2 подобно доказательству теоремы 1. Возьмем схему S, реализующую функцию f, со сложностью $C(S) = C_B(f)$. Схему S_k, вычисляющую одновременно $df, \ldots, d^k f/k!$, строим по индукции параллельно построению схемы S.

Пусть очередной арифметический элемент схемы S реализует функцию $w = u \circ v$, где u, v — функции, вычисляемые предшествующими элементами, и элементы, вычисляющие дифференциалы $du, dv, \ldots, d^k u/k!, d^k v/k!$ в схему S_k уже добавлены. Для вычисления дифференциалов $dw, \ldots, d^k w/k!$ добавим в схему S_k еще k аддитивных элементов и в случае, если элемент $\circ$ есть умножение или деление, добавим $O(M(k))$ мультипликативных и аддитивных элементов согласно лемме 4. Если очередной функциональный элемент g схемы S унарный и вычисляет функцию $w = g(u)$, то, согласно лемме 6, для вычисления дифференциалов $dw, \ldots, d^k w/k!$ надо добавить $O(M(k))$ элементов. Поэтому, применяя индукцию, получаем, что сложность построенной схемы $C_B(S_k) = O(M(k))C_B(S)$.

Замечание 3. Известен алгоритм [12] одновременного вычисления многочлена одной переменной и всех его производных в данной точке со сложностью $O(n \log n)$, но это другая задача, так как в ней многочлен представлен вектором коэффициентов, а не схемой.

6.1. Производные Хассе–Тейхмюллера

Как известно, k-я производная Хассе–Тейхмюллера (см. [11]) определяется как

$$f^{[k]} = \frac{d^k f}{k! d^k x}$$

и аналогичным образом определяются частные производные

$$\frac{\partial^{[k]} f}{\partial^k x} = \frac{\partial^k f}{k! \partial^k x}.$$

Тогда дифференциал k-го порядка определяется более краткой формулой

$$d^{[k]} u(x_1, \ldots, x_n) = \sum_{\substack{k_1 \geq 0, \ldots, k_n \geq 0 \\ k_1 + \ldots + k_n = n}} \frac{\partial^{[k]} f}{\partial^{k_1} x_1 \ldots \partial^{k_n} x_n} d^{k_1} x_1 \ldots d^{k_n} x_n.$$

Правило Лейбница для кратного дифференциала функций нескольких переменных записывается также более кратко:

$$d^{[k]}(f_1 \dots f_n) = \sum_{\substack{k_1 \geqslant 0, \dots, k_n \geqslant 0 \\ k_1 + \dots + k_n = n}} d^{[k_1]} f_1 \dots d^{[k_n]} f_n.$$

Формула Тейлора для нескольких переменных тоже приобретает более простой вид:

$$f(x_1 + dx_1, \dots, x_n + dx_n) = \sum_{i=0}^{k} d^{[i]} f, \qquad d^{[0]} f = f.$$

Формула Арбогаста–Бруно для $g(x_1, \dots, x_m) = f(u(x_1, \dots, x_m))$ приобретает вид

$$d^{[n]} g = \sum_{k=1}^{n} A'_{n,k} f^{[k]}(u),$$

$$A'_{n,k} = A'_{n,k}(du^{[1]}, \dots, d^{[n]} u) = \sum_{\substack{k_1 + \dots + k_n = k \\ k_1 + \dots + n k_n = n}} \frac{k!}{k_1! \dots k_n!} \prod_{i=1}^{n} (d^{[i]} u)^{k_i}.$$

Так как при $k_i \leqslant m_i, i = 1, \dots, n$,

$$\frac{\partial^{[k]} \left(x_1^{m_1} \dots x_n^{m_n} \right)}{\partial^{k_1} x_1 \dots \partial^{k_n} x_n} = \binom{m_1}{k_1} \dots \binom{m_n}{k_n} x_1^{m_1 - k_1} \dots x_n^{m_n - k_n},$$

можно определить частные производные Хассе–Тейхмюллера для любого многочлена так, чтобы это определение было корректным для полей любой характеристики.

Поэтому теорема 2 верна для многочленов над полями любой характеристики. Она также верна и для рациональных функций над полями любой характеристики. Действительно, частные производные Хассе–Тейхмюллера очевидно определяются для любого линейного многочлена. Реализуя рациональную функцию f в виде схемы над базисом B_{ar} и повторяя доказательство теоремы 2 с заменой обычных производных и дифференциалов на производные и дифференциалы Хассе–Тейхмюллера, можно вычислить производные и дифференциалы Хассе–Тейхмюллера функции f, применяя только арифметические операции к константам и переменным.

7. Некоторые замечания и приложения

7.1. О сложности преобразований базисов в алгебре симметрических полиномов и сложности цикловых индикаторов

Известно, что любой симметрический полином n переменных можно представить в виде полинома от n элементарных симметрических полиномов. Под элементарными полиномами можно при этом понимать полиномы σ_k, $k = 1, \dots, n$, первой степени по каждой переменной и полиномы s_k, $k = 1, \dots, n$, состоящие из одночленов, содержащих лишь одну переменную (степенные суммы). Связь между этими системами элементарных полиномов σ_k, $k = 1, \dots, n$, и s_k, $k = 1, \dots, n$, дается формулами Варинга. Как известно

[13], эти формулы можно выразить через полиномы Белла следующим образом:

$$\sigma_k = F_k(s_1,\dots,s_k) = (-1)^k \frac{Y_k(1;-s_1,-s_2,-2s_3,\dots,-(k-1)!\,s_k)}{k!},$$
$$s_k = F_k^{-1}(\sigma_1,\dots,\sigma_k) = -((k-1)!)^{-1} A_k(f;g),$$
$$f_j = (-1)^{j-1}(j-1)!, \quad g_j = (-1)^j j!\sigma_j, \quad j = 1,\dots,k, \quad k = 1,\dots,n.$$

Из леммы 5 следует, что сложность полиномиального преобразования $F_k, k = 1,\dots,n$, равна $O(M(n))$. Преобразование (F_k^{-1}) от системы σ_k, $k = 1,\dots,n$, к системе s_k, $k = 1,2,\dots$, согласно известным рекуррентным формулам Ньютона сводится к делению степенных рядов (фактически к делению многочленов) и выполняется, как известно (см. [11]), со сложностью $O(M(n))$.

В теории симметрических полиномов используется также система сумм однородных произведений h_k, $k = 1,2,\dots$, определяемая через равенство степенных рядов

$$(1-x_1x)\dots(1-x_nx) = 1-\sigma_1x+\sigma_2x^2-\dots+(-1)^n\sigma_nx^n$$
$$= (1+h_1x+h_2x^2+\dots)^{-1}.$$

В [13] и [14] приведены формулы, связывающие эту систему с каждой из двух выше упомянутых систем, выраженные через полиномы Белла $A_n(f;g)$ и в некоторых случаях через $Y_n(1;g)$.

Преобразование от системы h_k, $k = 1,\dots,n$, к системе σ_k, $k = 1,\dots,n$, и обратное к нему сводится к возведению степенного ряда в минус первую степень и может быть, как известно (см. [11]), выполнено со сложностью $O(M(n))$. Поэтому преобразование от системы $h_k, k = 1,\dots,n$, к системе $s_k, k = 1,\dots n$, может быть выполнено со сложностью $O(M(n))$, а обратное преобразование может быть выполнено со сложностью по порядку равной сложности системы полиномов $Y_k(g_1,\dots,g_k)$, $k = 1,\dots,n$, то есть также со сложностью $O(M(n))$.

Из леммы 5 следует, что сложность каждого из упомянутых выше полиномиальных преобразований равна $O(M(n))$.

В теории групп подстановок используется так называемый цикловой индикатор симметрической группы

$$C_n(x_1,\dots,x_n) = \sum \frac{n!}{k_1!\dots k_n!}\left(\frac{x_1}{1}\right)^{k_1}\dots\left(\frac{x_n}{n}\right)^{k_n}.$$

Как известно [13], он выражается через полином Белла

$$C_n(x_1,\dots,x_n) = Y_n(1;x_1,x_2,2!\,x_3,\dots,(n-1)!\,x_n).$$

Очевидно, что и полином Белла Y_n аналогично выражается через цикловой индикатор. Поэтому сложность циклового индикатора равна сложности полинома Y_n с точностью до аддитивного слагаемого $O(n)$.

В [13], гл. 4, п. 5, и [14] гл. 5, упр. 4, 8, показано, что

$$Y_n(x;1,\dots,1) = a_n(x) = \sum_{k=0}^{n} S(n,k)x^k,$$

где $S(n,k)$ — числа Стирлинга второго рода, $a_n(1) = B_n$ — числа Белла,

$$A_n(f_1,\dots,f_n,1,\dots,1) = \sum_{k=1}^{n} S(n,k)f_k.$$

Поэтому

$$C_{B_{ar}}(a_1(x),\dots,a_n(x)) = O(M(n)),$$
$$C_{B_{ar}}(B_1,\dots,B_n) = O(M(n)),$$
$$C_{B_{ar}}\left(\sum_{k=1}^{n} S(n,k)x_k, k = 1,\dots,n\right) = O(M(n)(n\log n)^{1/2}).$$

В [18] (см. [3], упр. 4.7.16) введено понятие степеноида произвольного степенного ряда

$$V(z) = v_1 z + \dots + v_n z^n + \dots$$

Так называется полином

$$V_n(x) = \sum_{k=0}^{n} v_{nk} x^k,$$

где

$$v_{nk} = \frac{n!}{k!}[z^n]V(z)^k.$$

Очевидно, он выражается через полином Белла

$$V_n(x) = Y_n(x; v_1, 2!\, v_2, \dots, n!\, v_n),$$

что неявно отмечено в [3], упр. 4.7.19.

7.2. О сложности проверки на локальный экстремум

Якобианом называется определитель $\det J(F)$ квадратной якобиевой матрицы $J(F)$. Как известно, если $\det J(F, X_0) \neq 0$ то в окрестности точки X_0 отображение F можно обратить, то есть найти такое отображение F^{-1}, определенное в некоторой окрестности точки $Y_0 = F(X_0)$, что в соответствующих окрестностях точек X, Y будут справедливы тождества

$$F^{-1}(F(X)) = X, \qquad F(F^{-1}(Y) = Y,$$
$$J(F^{-1}, F(X)) = J^{-1}(F, X), \qquad d(F^{-1})(F(X)) = (dF(X))^{-1}.$$

Известно [1], что сложность вычисления определителя и обращения невырожденной матрицы порядка n равна $O(n^\mu)$, где $\mu < 2{,}376$ — экспонента матричного умножения. Поэтому из теоремы 1 следует, что

$$C_B(F, J^{-1}(F, X)) \leqslant O(nC_B(F)) + O(n^\mu),$$
$$C_B(F, \det J(F, X)) \leqslant O(nC_B(F)) + O(n^\mu).$$

Поэтому сложность классических алгоритмов анализа для тестирования точки из области определения элементарной функции одной переменной $f(x)$ на стационарность, невырожденность и локальную экстремальность по порядку равна сложности вычисления этой функции в базисе B_{an}, а в случае n переменных она по порядку равна $nC_B(f)+n^{\mu+1}$.

7.3. О сложности билинейных и квадратичных форм

Квадратной матрице A можно сопоставить билинейную и квадратичную формы

$$A(X,Y) = \sum_{i,j=1}^{n} a_{i,j} x_i y_j,$$

$$Q(X) = \sum_{i,j=1}^{n} a_{i,j} x_i x_j$$

и линейный оператор $A(X)$,

$$Y = A(X), \quad y_i = \sum_{j=1}^{n} a_{i,j} x_j, \quad i = 1, \ldots, n.$$

Обычно квадратичные формы рассматривают для симметрических матриц A, то есть таких, что $A = A^T$. Пусть $B = B_{ar}$ или $B = \{x + y, x - y, xy\} \cup \{ax : a \in K\}$. Очевидно, что

$$C_B(Q(X)) \leqslant C_B(A(X,Y)) \leqslant C_B(A(X,Y), A(X)) \leqslant 2n - 1 + C_{B_l}(A(X)).$$

Для симметрической матрицы A над полем характеристики, не равной 2,

$$A(X,Y) = \frac{1}{4}(Q(X+Y) - Q(X-Y)),$$

поэтому

$$C_B(A(X,Y)) \leqslant 2C_B(Q(X)) + 2n + 2.$$

Справедливы следующие утверждения.

Для произвольной $n \times n$ матрицы A

$$C_B(A(X,Y), A(Y), A^T(X)) \leqslant 4C_B(A(X,Y)) + 1 - n,$$
$$C_B(A(Y), A^T(X)) \leqslant 4C_B(A(X,Y)) - n.$$

Если $A = A^T$, то

$$C_B(Q(X), A(X)) \leqslant 4C_B(Q(X)) + 1,$$
$$C_B(A(X)) \leqslant 4C_B(Q(X)),$$

следовательно, $C_B(A(X))$, $C_B(Q(X))$, $C_B(A(X,Y))$ по порядку равны.

Для доказательства первого неравенства применим неравенство

$$C_B(f, \operatorname{grad} f) \leqslant 4C_B(f) + 1 - n$$

из теоремы 1 к случаю $f(X,Y) = A(X,Y)$. Это можно сделать, потому что

$$\frac{\partial A(X,Y)}{\partial x_i} = \sum_{j=1}^{n} a_{i,j} y_j,$$

есть i-я компонента оператора $A(Y)$,

$$\frac{\partial A(X,Y)}{\partial y_j} = \sum_{i=1}^{n} a_{i,j} x_i$$

есть j-я компонента оператора $A^T(X)$.

Второе неравенство следует из первого.

Для доказательства третьего неравенства применим неравенство

$$C_B(f, \operatorname{grad} f) \leqslant 4C_B(f) + 1 - n$$

к случаю $f(X) = Q(X)$. Это можно сделать, потому что

$$\frac{\partial Q(X)}{\partial x_i} = 2\sum_{j=1}^{n} a_{i,j}x_j$$

есть удвоенная i-я компонента оператора $A(X)$. Остается добавить к оценке n скалярных умножений на $1/2$.

Последнее неравенство очевидно следует из предыдущего.

Очевидно, что

$$C_B(A(X,Y)) \leqslant 3n^2 - 1,$$
$$C_B(Q(X)) \leqslant \frac{3(n^2+n)}{2} - 1.$$

Покажем, что существуют билинейные формы с алгебраическими комплексными коэффициентами и с квадратичной сложностью. Для этого сопоставим произвольной билинейной форме $A(X,Y)$ многочлен

$$f_A(x) = A(U,W) = \sum_{k=0}^{n^2-1} b_k x^k,$$

где

$$b_{i+jn} = a_{i,j}, \qquad i,j = 0,1,\ldots,n-1,$$
$$U = (1, x, x^2, \ldots, x^{n-1}),$$
$$W = (1, x^n, x^{2n}, \ldots, x^{(n-1)n}),$$

и заметим, что его сложность

$$C_B(f_A(x)) \leqslant C_B(A(X,Y) + 2n - 3.$$

Возьмем, например, многочлен $f_A(x)$ степени $n^2 - 1$ с коэффициентами $\exp(2\pi i/p_k)$, где i есть мнимая единица, а $p_1, p_2, \ldots$ есть последовательность простых чисел. Согласно [7], его сложность равна $O(n^2)$. Следовательно, у матрицы A с элементами $a_{kj} = \exp(2\pi i/p_{j+n(k-1)})$ сложность соответствующей билинейной формы и линейного оператора по порядку не меньше n^2. Аналогично можно построить пример $n \times m$ матрицы, у которой соответствующая билинейная форма имеет сложность $O(nm)$.

В данном примере матрица несимметрична, поэтому его нельзя использовать для построения квадратичной формы сложности $O(n^2)$. Пример такой квадратичной формы можно получить, взяв любую билинейную форму $A(Y,Z)$ сложности по порядку n^2 с матрицей размера $\lfloor n/2 \rfloor \times \lceil n/2 \rceil$ и любые две квадратичные формы $F(Y)$, $G(Z)$ и рассмотрев квадратичную форму

$$Q(X) = F(Y) + G(Z) + A(Y,Z)$$

от n переменных $X = (Y, Z)$. Из этой формы можно получить формы F, G подстановками нулей вместо Z и Y соответственно, а билинейную форму $A(Y, Z)$ — вычитанием из Q форм B и C. Отсюда следует, что

$$C_B(A) \leqslant 3C_B(Q) + 2.$$

В [7] приведен другой пример линейного оператора с почти квадратичной сложностью, при этом дан только набросок доказательства, существенно более сложного, чем приведенное выше, и с немного более слабой оценкой.

7.4. О сложности вычисления второго дифференциала

При малых k мультипликативные константы в теореме 2, разумеется, легко указать более точно. Например, если число элементов exp, ln в схеме S обозначить $EL(S)$, число элементов cos, sin, arctg, arcsin обозначить $T(S)$, а число всех мультипликативных элементов обозначить $MUL(S)$, то для любой элементарной функции f от n переменных и любой вычисляющей ее схемы S сложности $C(S)$ можно построить схему S_2, вычисляющую одновременно f, df, $d^2 f$ со сложностью

$$C(S_2) = 3C(S) + 6MUL(S) + 3EL(S) + 5T(S) \leqslant 9C_B(S).$$

7.5. Верхние оценки сложности некоторых полиномов

Пусть $f = x_1 \dots x_n$, тогда $(f, \operatorname{grad} f, H(f))$ можно считать $(1 + n(n+1)/2, n)$-вектор-функцией, состоящей из многочленов

$$f = x_1 \dots x_n, \qquad \frac{\partial f}{\partial x_i} = x_1 \dots \hat{x}_i \dots x_n,$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = x_1 \dots \hat{x}_i \dots \hat{x}_j \dots x_n, \qquad 1 \leqslant i < j \leqslant n.$$

Применяя к $f = x_1 \dots x_n$ теорему 1, можно получить схему сложности $3n - 6$ и глубины $2\lceil \log_2 n \rceil - 2$.

Пусть

$$g = \sum_{i=1}^{n} x_1 \dots x_{i-1} y_i x_{i+1} \dots x_n.$$

Применяя к тому же одночлену f следствие из теоремы 1, получаем схему, одновременно реализующую многочлены f, g, $\operatorname{grad} f$ и

$$h = \sum_{1 \leqslant i < j \leqslant n} x_1 \dots x_{i-1} y_i x_{i+1} \dots x_{j-1} y_j x_{j+1} \dots x_n$$

со сложностью $14n - 20$, построенную только из умножений и сложений. Глубина схемы для f, g, h равна

$$2(2\lceil \log_2 n \rceil - 2) + \lceil \log_2 n \rceil + 2 = 5\lceil \log_2 n \rceil - 2.$$

В базисе B_{ar} можно получить лучшую оценку, если взять для f, $\operatorname{grad} f$ схему сложности $2n - 2$ и глубины $\lceil \log_2 n \rceil + 1$ и применить к ней теорему 1. Можно получить более

точные оценки, если не пользоваться теоремой 1, а подсчитать число элементов в схеме непосредственно. Одночлен f вместе с df вычисляется со сложностью $4n-4$ и глубиной $2\lceil\log_2 n\rceil$. Для вычисления $d(f/x_i)$ применяется формула

$$\frac{1}{x_i}\left(df - \frac{f}{x_i}dx_i\right),$$

тогда глубина возрастает на 2, а сложность возрастает на $4(n-1)$. При вычислении $d^2 f$ сложность увеличивается на $2n$, а глубина увеличивается на $\lceil\log_2 n\rceil + 2$. Оценка для сложности f, g, $\operatorname{grad} f$, h равна $10n-8$ и для глубины равна $3\lceil\log_2 n\rceil + 4$.

Прямое применение следствия из теоремы 1 для оценки сложности системы многочленов

$$f, g, x_1 \dots \hat{x}_i \dots x_n, x_1 \dots \hat{x}_i \dots \hat{x}_j \dots x_n, \qquad 1 \leqslant i < j \leqslant n,$$

дает оценку $9n^2$. Если построить схему согласно этому следствию и учесть элементы, исчезающие при подстановке констант, то мультипликативная сложность данной системы многочленов оказывается асимптотически не большей n^2.

Если применить теорему 2 к одночлену $f = x_1 \dots x_n$, то, так как

$$\frac{d^k f}{k!} = \sum_{1\leqslant i_1<\dots<i_k\leqslant n} x_1 \dots \hat{x}_{i_1} \dots \hat{x}_{i_k} \dots x_n dx_{i_1} \dots dx_{i_k},$$

справедливо следующее утверждение.

Сложность совместного вычисления в базисе $\{x \pm y, xy\}$ многочленов $f, f_1, \dots f_k$, где

$$f_k = \sum_{1\leqslant i_1<\dots<i_k\leqslant n} x_1 \dots \hat{x}_{i_1} \dots \hat{x}_{i_k} \dots x_n y_{i_1} \dots y_{i_k},$$

оценивается как $O(M(k)n)$.

Приведенные утверждения верны для любого конечного поля и базиса из сложения и умножения, в частности, для булевых функций и базиса из конъюнкции и сложения по модулю два.

Список литературы

1. Baur W., Strassen V., The complexity of partial derivatives. *Theoret. Computer Sci.* (1983) **22**, 317–330.
2. Ким К. В., Нестеров Ю. Е., Черкасский Б. В., Оценка трудоемкости вычисления градиента. *Докл. АН СССР* (1984) **275**, №6, 1306–1309.
3. Кнут Д., *Искусство программирования. Основные алгоритмы*, т. 1. Вильямс, Москва, 2000.
4. Linnainmaa S., Taylor expansion of the accumulated rounding error. *BIT, Nord. Tidskr. Inf.-Behandl.* (1976) **16**, 146–160.
5. Григорьев Д. Ю., Нижние оценки в алгебраической сложности вычислений. *Теория сложности вычислений. Записки научных семинаров ЛОМИ* (1982) **118**, 25–82.
6. Gashkov S., Kochergin V., On addition chains of vectors, gate circuits, and the complexity of computation of power. *Syberian Adv. Math.* (1994) **4**, №4, 1–16.
7. Heintz J., Sievekieng M., Lower bounds for polynomials with algebraic coefficients. *Theoret. Computer Sci.* (1980) **11**, №3, 321–330.

8. Архипов Г.И., Садовничий В. А., Чубариков В. Н., *Лекции по математическому анализу.* Высшая школа, Москва, 1999.

9. Гурса Э., *Курс математического анализа.* ОНТИ СССР, Москва, 1936.

10. Гашков С.Б., Чубариков В. Н., *Арифметика. Алгоритмы. Сложность вычислений.* Высшая школа, Москва, 2000.

11. von zur Gathen J., Gerhard J., *Modern computer algebra.* Cambridge Univ. Press, Cambridge, 1999.

12. Ахо Ф., Хопкрофт Дж., Ульман Дж., *Построение и анализ вычислительных алгоритмов.* Мир, Москва, 1979.

13. Риордан Дж., *Введение в комбинаторный анализ.* ИЛ, Москва, 1963.

14. Риордан Дж., *Комбинаторные тождества.* Наука, Москва, 1982.

15. Bell E. T., Exponential polynomials. *Ann. Math.* (1934) **35**, 258–277.

16. Brent R. P., Kung H. T., Fast algorithms for manipulating formal power series. *J. Assoc. Comput. Mach.* (1978) **25** (**4**), 581–595.

17. Brent R. P., Multiple-precision zero-finding methods and the complexity of elementary function evaluation. In: *Analytic Computational Complexity, Proc. Symp. Carnegie–Mellon Univ., Pittsburgh 1975.* Academic Press, New York, 1976, pp. 151–176.

18. Steffensen J. F., The poweroid, an extension of the mathematical notion of power. *Acta Math.* (1941) **73**, 333–366.

Статья поступила 21.09.2004.