

УДК 519.61

ПРИКЛАДНАЯ МАТЕМАТИКА

Д. З. Геворкян

О сложности линейных вычислений на многопроцессорных системах

(Представлено чл.-корр. АН Армянской ССР Р. Р. Варшамовым 1/II 1986 г.)

При решении различных прикладных задач часто требуется вычислять произведение $\bar{y} = (y_1, y_2, \dots, y_N)^T$ фиксированной $(N \times N)$ -матрицы F на произвольный вектор переменных $\bar{x} = (x_1, x_2, \dots, x_N)^T$:

$$\bar{y} = F \cdot \bar{x} \quad (1)$$

При этом большинство решаемых задач а) имеют массовый характер; б) должны решаться в ограниченном времени; в) требуют обработки больших массивов информации. Это приводит к необходимости разработки вычисляющих произведение (1) параллельных алгоритмов, предназначенных для реализации на многопроцессорных вычислительных системах, представляющих собой множество (из K) процессорных элементов (ПЭ), объединенных с помощью специальной системы коммутации и выполняющих любую из l заданных операций (^{1,2}). Если K —ограничено, то говорят, что алгоритм обладает ограниченным параллелизмом, в противном случае—неограниченным параллелизмом.

В зависимости от параллелизма выполнения операций и передачи данных многопроцессорные системы делятся на системы типа SIMD (одна команда—много данных), MISD (много команд—одно данное) и MIMD (много команд—много данных) (²).

В работе (³) были построены реализуемые на системах типа SIMD параллельные алгоритмы с ограниченным параллелизмом, вычисляющие дискретные ортогональные преобразования Фурье, Уолша, ВКФ.

В данной работе исследуется сложность вычисления произведения (1) на системах типа SIMD и MISD. С этой целью введены модели линейных SIMD- и MISD-вычислений и найдены нижние оценки для числа параллельных шагов, необходимых для вычисления (1) посредством этих вычислений. Построены асимптотически оптимальные—с числом шагов, по порядку не превышающим найденную нижнюю оценку,—линейные SIMD-вычисления с ограниченным параллелизмом, вычисляющие широкий класс дискретных ортогональных преобразований (Фурье, Хаара, Уолша и др.). Предложенные алгоритмы обладают тем свойством, что число K используемых процессорных элементов может быть задано заранее, что важно с точки зрения практической реализации и надежности.

Пусть $X = \{x_1, x_2, \dots, x_N\}$ —множество входных переменных над

полем комплексных чисел C . и пусть $z_j^s = x_{i_k}$; $-m \leq j \leq 0$; $s \in S_j \subset \{1, 2, \dots, K\}$ $x_{i_k} \in X$ (соответствие взаимнооднозначно).

Определение 1. Линейным K -параллельным SIMD-вычислением назовем последовательность групп операций такую, что j -я ($j=1, 2, \dots, L$) группа есть:

- либо $z_j^s \leftarrow z_p^r$ ($r \neq s$) для всех $s \in S_j \subset \{1, 2, \dots, K\}$ (обмены);
- либо $z_j^s \leftarrow \lambda_j^s \cdot z_q^s$ ($\lambda_j^s \in C$) для всех $s \in S_j \subset \{1, 2, \dots, K\}$ (умножения);
- либо $z_j^s \leftarrow z_t^s + z_l^s$ для всех $s \in S_j \subset \{1, 2, \dots, K\}$ (сложения).

Индексы p, q, t, l, r зависят от j и s ; $-m \leq p, q, t, l < j$; $r \in S_p$.

Каждой рабочей переменной z_j^s линейного SIMD-вычисления естественным образом соответствует некоторая линейная функция от $x_1, x_2, \dots, x_N$, называемая значением z_j^s . Скажем, что линейное K -параллельное SIMD-вычисление вычисляет произведение $\bar{y} = (y_1, y_2, \dots, y_N)^T$ фиксированной $(N \times N)$ -матрицы F на произвольный вектор $\bar{x} = (x_1, x_2, \dots, x_N)^T$, если при любых значениях x_i значения всех y_m встречаются среди значений z_j^s , и обозначим через $SMK(F) = \{\beta\}$ класс всех таких вычислений, и пусть $SMK(F, \lambda) = \{\beta \in SMK(F) / (1/|J|) \sum_{j \in J} (1/|S_j|) \sum_{s \in S_j} |\lambda_j^s| \leq \lambda\}$, где J — множество всех индексов j , соответствующих умножениям в вычислении β .

Обозначим через $L_\beta^0, L_\beta^+, L_\beta^x$ число параллельных шагов (число групп) в вычислении β , на которых выполняются обмены, сложения и умножения на комплексные константы соответственно; $\mu \geq 1$ — отношение среднего времени умножения к среднему времени сложения на ПЭ.

Теорема 1. Пусть $\beta \in SMK(F)$ и $c = \max_{j \in J} \{2, |\lambda_j^s|^{1/\mu}\}$. Тогда

$$L_\beta^+ + \mu \cdot L_\beta^x \geq \log |\Delta(F)| / (K \cdot \log c) \quad (2)$$

где $\Delta(F)$ — наибольший по модулю минор матрицы F .

Теорема 2. Для всякого $\beta \in SMK(F, \lambda)$ справедливо неравенство

$$L_\beta^0 + (\log_2 3) \cdot L_\beta^+ + (\log_2 (\lambda + 1)) L_\beta^x \geq 2 \cdot \log_2 |\det F| / K \quad (3)$$

Определение 2. Линейным MISD-вычислением γ длины L_γ с длиной конвейера K назовем последовательность групп операций такую, что j -я ($j=1, 2, \dots, L_\gamma$) группа есть

$$\bar{v}_j^s \leftarrow A_j^s \cdot \bar{u}_j^s,$$

где $r_1 = \min\{j, K\} \geq s \geq r_2 = \begin{cases} 1, & \text{при } j \leq L-K \\ j+K-L, & \text{при } j > L-K \end{cases}$; $\bar{v}_j^s = (v_j^s(1), v_j^s(2))^T$;

A_j^s — (2×2) -матрицы с элементами $a_j^s(t, l) \in C$; $\bar{u}_j^s \in X^2$ при $s=1$ и $L-K \leq j \leq L$ и $\bar{u}_j^s = \bar{v}_{j-1}^{s-1}$ в других случаях.

Класс линейных MISD-вычислений γ с длиной конвейера K , вычисляющих произведение (1) (при любых значениях x_i значения y_m встречаются среди значений $\bar{v}_j^s(t)$, $t=1, 2$) обозначим через $MSK(F)$ и пусть

$$\text{MSK}(F, \alpha) = \{\gamma \in \text{MSK}(F) \mid \|\alpha_j^s(t, l)\| \leq \alpha; j = 1, 2, \dots, L_\gamma; s = r_1, r_1 + 1, \dots, r_1\}$$

Теорема 3. Для всякого $\gamma \in \text{MSK}(F)$ справедливо неравенство

$$L_\gamma \geq \log |\det F| / \log b, \quad (4)$$

$$\text{где } b = (1/L_\gamma) \sum_{j=1}^{L_\gamma} \left[\prod_{s=r_1}^{r_1} (1 + (1/2) \|A_j^s\|)^2 \right].$$

Теорема 4. Для всякого $\gamma \in \text{MSK}(F, \alpha)$ справедливо неравенство

$$L_\gamma \geq \log |\Delta(F)| / \log(2\alpha) + K \quad (5)$$

Покажем асимптотическую достигаемость оценок (2), (3) для достаточно широкого класса дискретных ортогональных преобразований (ДОП), т. е. для вычисления произведения (1), когда $F = H_N$ есть $(N \times N)$ -матрица ($N = q^m$), строки которой соответствуют дискретным функциям в некотором ортогональном базисе и которую можно представить в виде произведения слабозаполненных матриц (4):

$$H_N = G_{m-1} \cdot G_{m-2} \cdot \dots \cdot G_0 \quad (6)$$

Большинство элементов в матрицах $G_r (r = 0, m-1)$ равны нулю, а ненулевые элементы независимо от базисной системы функций расположены в определенном порядке (4). На представлении (6) основаны алгоритмы быстрых дискретных ортогональных преобразований (БДОП), вычисляющие ДОП с матрицей H_N по соотношениям:

$$\bar{z}_0 = x; \bar{z}_{r+1} = G_r \cdot \bar{z}_r, r = 0, 1, \dots, m-1; \bar{y} = \bar{z}_m$$

Для вычисления $\bar{z}_{r+1}$ по $\bar{z}_r$ (r -й этап БДОП) требуется q^{m-1} умножений $(q \times q)$ -матриц $V_r^s (s = 1, q^{m-1})$, составленных из элементов G_r , на $(q \times 1)$ -подвекторы вектора $\bar{z}_r$. В зависимости от элементов V_r^s можно получить класс $\mathcal{H}$ матриц H_N , соответствующих ДОП с алгоритмами БДОП, в частности, дискретным преобразованиям Фурье (ДПФ), Уолша (ДПУ), Виленкина-Крестенсена (ВКФ), Хаара (ДПХ) (4).

Лемма. Пусть $H_N \in \mathcal{H}$. Тогда ее можно представить в виде:

$$H_N = (G'_{m-1} \cdot P_q) \cdot (G'_{m-2} \cdot P_q) \cdot \dots \cdot (G'_0 \cdot P_q), \quad (7)$$

где $G'_r (r = 0, m-1)$ — квазидиагональные матрицы с блоками V_r^s на диагонали; P_q — матрица перестановки:

$$P_q \cdot (\alpha^0, \alpha^1, \dots, \alpha^{N-1})^T =$$

$$= (\alpha^0, \alpha^{N/q}, \dots, \alpha^{(q-1)N/q}, \alpha^1, \alpha^{N/q+1}, \dots, \alpha^{(q-1)N/q+1}, \dots, \dots, \alpha^{N-1})^T$$

Теперь приведем вычисляющие ДОП с матрицами $H_N \in \mathcal{H}$ линейные SIMD-вычисления, которые схематически можно описать следующим образом.

Алгоритм I.

Вход. Вектор $\bar{x} = (x_0, x_1, \dots, x_{N-1})^T$, $N = q^m$.

Число ПЭ. $q^{m'}$ ($m' \leq m-1$).

Выход. Вектор $\bar{y} = H_N \cdot \bar{x}$, где $H_N \in \mathcal{H}$.

Вычисление. Вычисление состоит из m этапов, определяемых соотношениями:

$$\bar{z}_0 = \bar{x}; \bar{z}_{r+1} = G'_r \cdot (P_q \cdot \bar{z}_r), r = 0, 1, \dots, m-1; \bar{y} = \bar{z}_m$$

На каждом этапе сначала осуществляется перестановка компонент исходного вектора $\bar{z}_r = (\sigma_r^0, \dots, \sigma_r^{N-1})^T$ согласно оператору P_q . Затем выполняются $q^{m-m'-1}$ операций, p -я из которых ($p=0, 1, \dots, q^{m-m'-1}-1$) состоит из следующих шагов:

шаг I. В каждую s -ю ветвь вычисления (в s -й ПЭ) засылается под-вектор $\bar{z}_r^l = (\sigma_r^l, \sigma_r^{lq+l}, \dots, \sigma_r^{(q-1)N/q+l})^T$ вектора $P_q \cdot \bar{z}_r$, где $l = p \cdot q^{m'} + s$ ($s = \overline{1, q^m}$).

шаг II. Параллельно в каждой ветви s ($s = \overline{1, q^m}$) вычисляются $V_r^l \cdot \bar{z}_r^l$.

шаг III. Полученные на втором шаге результаты посылаются обратно.

Перейдем теперь к случаю, когда число ПЭ— K задано заранее. *Алгоритм II* (вычисление ДОП потока векторов).

Вход. Поток векторов $\chi = \{\bar{x}(1), \bar{x}(2), \dots, \bar{x}(i), \dots\}$.

Число ПЭ. Произвольное K .

Выход. Поток векторов $Y = \{\bar{y}(1), \bar{y}(2), \dots, \bar{y}(i), \dots\}$, где $\bar{y}_i = H_N \cdot \bar{x}(i)$, $H_N \in \mathcal{H}$; $i = 1, 2, \dots$

Вычисление. Представим число K в виде:

$$K = \sum_{h=0}^{t-1} q^{m_h},$$

где $0 \leq m_h \leq m-1$; $m_{h+1} \geq m_h$ и сгруппируем ветви вычисления (соответственно ПЭ системы), включив в h -ю группу q^{m_h} ветвей. Из очередных K векторов потока χ образуем t массивов, включив в h -й массив $X(h)$ q^{m_h} векторов. В каждой h -й группе ветвей параллельно вычисляются ДОП векторов из h -го массива, последовательно для каждого вектора $\bar{x}(i) \in X(h)$ применяя алгоритм 1 при $m' = m_h$. После этого вычисляются ДОП следующих K векторов потока χ и т. д.

Утверждение. При любом K построенные линейные K -параллельные SIMD-вычисления асимптотически оптимальны (т. е. асимптотически достигаемы оценки (2) и (3)) в классах $SMK(H_N, \lambda)$, где λ определяются элементами матриц G_r в представлении (6) для $H_N \in \mathcal{H}$. В частности для ДПФ, ДПУ, ВКФ $\lambda = 1$.

Замечание. Полученные результаты можно обобщить для задачи вычисления ДОП функций, определенных на абелевой группе G порядка $N = q^m$ по системе характеров G .

Автор выражает искреннюю благодарность С. С. Агаяну за постановку задач и руководство работой.

Вычислительный центр Академии наук Армянской ССР
и Ереванского государственного университета

Դ. Զ. ԳԵՎՈՐԳՅԱՆ

Բազմապարզանքային համակարգերում գծային հաշվումների բարդության վերաբերյալ

Աշխատանքում հետազոտվում է SIMD և MISD տեսակի համակարգերի (2) միջոցով $(N \times N)$ -մատրիցի և $(N \times 1)$ -վեկտորի արտադրյալը հաշվելու բարդությունը: Այդ նպատակով ներմուծված են գծային SIMD-և MISD-հաշ-

վումների հասկացությունները և գտնված են դրանց վրա սահմանված բարդության ստորին գնահատականներ: Կառուցված են դիսկրետ օրթոգոնալ ձևափոխություններ հաշվող ասիմպտոտիկ օպտիմալ (ներմուծված բարդության նկատմամբ) գծային SIMD-հաշվումներ, ըստ որում պահանջվող պրոցեսորային տարրերի թիվը կամայական է, որը շատ կարևոր է գործնական տեսանկյունով:

ЛИТЕРАТУРА — ԳՐԱԿԱՆՈՒԹՅՈՒՆ

¹ Элементы параллельного программирования, Радио и связь, М., 1983. ² Алгоритмы, математическое обеспечение и архитектура многопроцессорных вычислительных систем, Наука, М., 1982. ³ S. S. Agaian, D. Z. Gevorgian, Colloquia Mathematica Societatis János Bolyai, 44, Theory of algorithms, Péch (Hungary), p. 15—25 (1984). ⁴ А. И. Солодовников, И. И. Канатов, А. М. Спиваковский, Вопросы теории систем автоматического управления. Вып. 2, Изд-во ЛГУ, с. 99—112 (1976).