The Non-continuous Direction Vector I Test

Minyi Guo[1], Weng-Long Chang[2] and Jiannong Cao[3]
[1]Department of Computer Software,
The University of Aizu, Aizu-Wakamatsu City, Fukushima 965-8580, Japan
[2]Department of Management Information
Southern Taiwan University of Technology, Tainan County, Taiwan 710, R.O.C.
[3]Department of Computing
Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
E-mail: [1]minyi@u-aizu.ac.jp, [2]changwl@csie.ncku.edu.tw, [3]csjcao@comp.polyu.edu.hk

Abstract

In this paper, we offer the non-continuous direction vector I test, an extension of the direction vector I test, to make sure whether there are integer-valued solutions for one-dimensional arrays with constant bounds and non-one-increment.

Index Terms – Parallelizing Compilers, Data Dependence Analysis.

1. Introduction

The data dependence problem in general case can be reduced to that of checking whether a system of one linear equation with *m* unknown variables has a simultaneous integer solution, which satisfies the constraints for each variable in the system. Assume that a linear equations in a system is written as (1–1): $a_1X_1 + a_2X_2 + \cdots + a_{m-1}X_{m-1} + a_mX_m = a_0$, where each a_j is an integer for $0 \le j \le m$ and each X_k is a scalar integer variable for $1 \le k \le m.$ Suppose that the constraints to each variable in (1–1) are represented as (1–2): $M_k \le X_k \le N_k$, $X_k = M_k + (m\text{-}1) * INC_k$ and $1 \le m \le P$, where M_k, N_k and INC_k are integers for $1 \le k \le m$ and M_k, N_k and INC_k are, respectively, lower bound, upper bound and increment of a general loop and *P* is the number of loop iteration in the general loop and $P = {(N_k - M_k)}/{INC_k} + 1.$ Famous data dependence methods include [1-9].

2. Background

2.1. The Direction Vector I Test

Definitions 2–1, cited from [2, 9], defines an interval equation.

Definition 2–1: Let $a_1, \cdots, a_{m-1}, a_m$, *L* and *U* be integers. A linear equation (2–1), $a_1X_1 + a_2X_2 + \cdots + a_{m-1}X_{m-1} + a_mX_m =$ [*L, U*], is referred to as an interval equation.

In light of [9], the direction vector I test considers a pair of same index variables to justify the movement of the two variables to the right. A pair of same index variables in the equation (2–1) can be moved to the right if the coefficients of the two variables have small enough values to justify the movement of the two variables to the right.

Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04)
1087-4089/04 $ 20.00 © 2004 IEEE

3. The Non-continuous Direction Vector I Test

3.1. Non-continuous Interval-Equation

Definition 3–0: Let $[M, N, INC, \frac{N-M}{INC}+1]$ represent the *non-continuous* integer intervals from M to N, i.e., the set of all of the *non-continuous* integers, $\left\{M+(P-1)\times INC \mid 1 \le P \le \frac{N-M}{INC}+1\right\}$. ■

Definition 3-1: Let $a_0, a_1, \cdots, a_{m-1}, a_m$ be integers. For each k, $1\le k \le m$, let each of M_k and N_k be integer, where $M_k \le N_k$. If $m > 0$, then the equation, $a_1X_1 + \cdots a_mX_m = a_0$, is said to be $([M_1, N_1, INC_1, \frac{N_1-M_1}{INC_1}+1]; \ldots; [M_m, N_m, INC_m, \frac{N_m-M_m}{INC_m}+1])$-integer solvable if there exist integers $j_1, j_2, \ldots, j_m$, such that

· $a_1j_1 + a_2j_2 + \cdots + a_{m-1}j_{m-1} + a_mj_m = a_0$.

· For each k, $1\le k \le m$: $j_k = M_k+(p-1)\times INC_k$, where p is an integer and $1 \le p \le \frac{N_k-M_k}{INC_k}+1$. ■

Definition 3-2: Let $a_0, a_1, \cdots, a_{m-1}, a_m$ L and U be integers. A *non-continuous interval equation* is an equation in the form of $a_1X_1 + \cdots + a_mX_m = [L, U, INC, \frac{U-L}{INC}+1]$, which denotes the set of normal equations consisting of: $a_1X_1 + \cdots + a_mX_m = L$, …, $a_1X_1 + \cdots + a_mX_m = L + (\frac{U-L}{INC}) \times INC = U$. ■

Definition 3-3: Let $a_0, a_1, \cdots, a_{m-1}, a_m$, L and U be integers. For each k, $1\le k \le m$, let each of M_k and N_k be either an integer, where $M_k \le N_k$. If $m > 0$, then the non-continuous interval equation $a_1X_1 + \cdots + a_mX_m = [L, U, INC, \frac{U-L}{INC}+1]$ is said to be $([M_1, N_1, INC_1, \frac{N_1-M_1}{INC_1}+1]; \ldots; [M_m, N_m, INC_m, \frac{N_m-M_m}{INC_m}+1])$-integer solvable if one or more of the equations in the set which it denotes is $([M_1, N_1, INC_1, \frac{N_1-M_1}{INC_1}+1]; \ldots; [M_m, N_m, INC_m, \frac{N_m-M_m}{INC_m}+1])$-integer solvable. ■

3.2. Mathematical Preliminaries

Definition 3-4: Let S and S' be sets of *non-continuous* integers. We define an addition and a substitution operation on sets of non-continuous integer as follows: $S + S' = \{s + s' \mid s \in S \text{ and } s' \in S'\}$ and $S - S' = \{s - s' \mid s \in S \text{ and } s' \in S'\}$. Note that if S is the non-continuous integer interval $[L, U, INC, \frac{U-L}{INC}+1]$ and $S' = \{s_1, s_2, \cdots, s_n\}$, it follows that

$$[L, U, INC, \frac{U-L}{INC}+1] + S' = \bigcup_{i=1}^{n}[L+s_i, U+s_i, INC, \frac{U-L}{INC}+1]$$

and

$$[L, U, INC, \frac{U-L}{INC}+1] - S' = \bigcup_{i=1}^{n}[L-s_i, U-s_i, INC, \frac{U-L}{INC}+1].$$ ■

Lemma 3-1: Let $[L, U, INC, \frac{U-L}{INC}+1]$ be a *non-continuous* integer interval. Let $[M,N,DIF,\frac{N-M}{DIF}+1]$ be also a *non-continuous* integer interval, where $M+DIF<N$. Let S = {$b*y + c*z$| y and z are, respectively, one element in $[M,N,DIF,\frac{N-M}{DIF}+1]$ and $y < z$}. Let

$$t=\begin{cases}\max(|b*DIF|,|c*DIF|) & \text{if } b*c>0\\ \max(\min(|b*DIF|,|c*DIF|),|(b+c)*DIF|) & \text{if } b*c<0.\end{cases}$$

(Part I):

$$[L,U,INC,\frac{U-L}{INC}+1]+S=[L-(b^- -c)^+ *(N-M-DIF)+(b+c)*M+c*DIF,\ U+(b^+ +c)^+ *(N-M-DIF)+(b+c)*M+c*DIF,\ INC,\ (U-L)+\frac{(N-M-DIF)*((b^+ +c)^+ +(b^- -c)^+}{INC}+1]$$

iff $t\le U-L+INC, 0\le t\le U-L+INC$ $t\le U-L+INC$, $0\le t\le U-L+INC$ and t is a multiple of INC

(Part II):

$$[L,U,INC,\frac{U-L}{INC}+1]-S=[L-(b^+ +c)^+ *(N-M-DIF)-(b+c)*M-c*DIF,U+(b^- -c)^+ *(N-M-DIF)-(b+c)*M-c*DIF,INC,\frac{(U-L)+(N-M-DIF)*((b^- -c)^+ +(b^+ +c)^+)}{INC}+1]$$

iff $t\le U-L+INC, 0\le t\le U-L+INC$ and t is a multiple of INC. ■

Proof: Omitted due to space limit. ■

3.3. Non-continuous Interval-Equation Transformation

First, if two variables are related by a direction vector constraint of "=," they may be replaced by a single variable. Second, terms with zero coefficients may be omitted. Finally, a ">" constraint from one variable to another may be replaced by a constraint in the reverse direction. Taking all of those points into account, we propose Lemma 3-2, which is extended from Theorem 3 in [9].

Lemma 3-2: Let E=[(3-1), (3-2)], where (3-1) is equal to $\sum_{q=1}^{n}a_qX_q+\sum_{q=n+1}^{m}(b_qY_q+c_qZ_q)$ $[L,U,INC,\frac{U-L}{INC}+1]$, and (3-2) is equal to $X_q\in[M_q,N_q,INC_q,\frac{N_q-M_q}{INC_q}+1]$ for $1\le q\le n$ and Y_q and $Z_q\in[M_q,N_q,INC_q,\frac{N_q-M_q}{INC_q}+1]$ and $Y_q<Z_q$. for $n+1\le q\le m$. Let E'=[(3-3), (3-4)], where (3-3) is equal to $\sum_{q=1}^{n}a_qX_q+\sum_{q=n+1}^{m-1}(b_qY_q+c_qZ_q$

$$=[L-(b_m^+ +c_m)^+(N_m-M_m-INC_m)-(b_m+c_m)*M_m-c_m*INC_m,$$

$$U+(b_m^- -c_m)^+(N_m-M_m-INC_m)-(b_m+c_m)*M_m$$

Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04)
1087-4089/04 $ 20.00 © 2004 IEEE

$-c_m * INC_m, INC,$

$$(U-L)+(N_m-M_m-INC_m)*\frac{((b_m^- - c_m)^+ + (b_m^+ + c_m)^+)}{INC}+1],$$

and (3-4) is equal to

$$X_q \in [M_q, N_q, INC_q, \frac{N_q - M_q}{INC_q}+1]$$

for $1 \le q \le n$, Y_q and $Z_q \in$

$$[M_q, N_q, INC_q, \frac{N_q - M_q}{INC_q}+1]$$

and $Y_q < Z_q$ for $n+1 \le q \le m-1$.

Let

$$t_m = \begin{cases} \text{if } b_m * c_m > 0 \\ \quad \max(|b_m * INC_m|, |c_m * INC_m|) \\ \text{if } b_m * c_m < 0. \\ \quad \max(\min(|b_m * INC_m|, |c_m * INC_m|), \\ \qquad |(b_m + c_m) * INC_m|) \end{cases}$$

If $t_m \le U - L + INC, 0 \le t_m \le U - L + INC$, and t_m is a multiple of INC, then E is integer solvable iff E' is integer solvable.

Proof: Omitted due to space limit. ■

We take an example to show the power of Lemmas 3-1 and 3-2. Consider the normal linear equation (Ex1): $X_1 - X_2 = 0$, subject to the constraints X_1 and X_2: [1, 9, 2, 5] and $X_1 < X_2$. First, the *non-continuous direction vector I test* transforms the equation (Ex1) into the following *non-continuous interval equation* (Ex1-1): $X_1 - X_2$ = [0, 0, 2, 1]. In light of Lemmas 3-1 and 3-2, because the coefficients of X_1 and X_2 are, respectively, 1 and -1, t_1 is equal to 2. Since $t_1 \le 2$, $0 \le t_1 \le 2$ and t_1 is a multiple of 2, the condition of the movement for the pair of the same index variable, X_1 and X_2 is satisfied according to Lemma 3-2. Therefore, X_1 and X_2 are selected to move to the right-hand-side of (Ex1-1). Due to Lemma 3-2, a new non-continuous interval equation is obtained (Ex1-2): 0 = [2, 8, 2, 4]. Because $2 \le 0$ is false, 0 is not one element in the non-continuous integer interval [2, 8, 2, 4]. Thus, the non-continuous direction vector I test concludes that there is no integer-valued solution.

3.4. Interval-Equation Transformation Using the GCD Test

If all coefficients for variables in the non-continuous interval equation have no sufficiently small values to justify the movements of variables to the right, then Lemmas 3-1 and 3–2 can not be applied to result in the immediate movement. While every variable in a non-continuous interval equation cannot be moved to the right, Theorem 3–1 and Lemma 3-3 describe a transformation using the GCD test that enables additional variables to be moved.

Theorem 3-1: Let E=[(3-1), (3-2)], and let $g = gcd(a_1, \dots, a_n, b_{n+1}, \dots, b_m, c_{n+1}, \dots, c_m)$. E is integer solvable iff $g*\lceil L/g \rceil$ is one element of the integer set $\{L+(m-1)\times INC \mid 1 \le m \le \frac{U-L}{INC}+1\}$.

Proof: Omitted due to space limit. ■

Lemma 3-3: Let E=[(3-1), (3-2)], and let $g=gcd(a_1, \cdots, a_n, b_{n+1}, \cdots, b_m, c_{n+1}, \cdots, c_m)$. Let E'=[(3-5), (3-6)], where (3-5) is equal to

$$\sum_{q=1}^{n} \frac{a_q}{g} X_q + \sum_{q=n+1}^{m} (\frac{b_q}{g} Y_q + \frac{c_q}{g} Z_q) = [\frac{L}{g}, \frac{U}{g},$$

$\frac{INC}{g}, \frac{U-L}{INC}+1]$, and (3-6) is equal to

$$\forall X_q \in [M_q, N_q, INC_q, \frac{N_q - M_q}{INC_q}+1]$$

for $1 \le q \le n$ and

$$\forall Y_q \text{ and } Z_q \in [M_q, N_q, INC_q, \frac{N_q - M_q}{INC_q}+1] \text{ and}$$

$Y_q < Z_q$ for $n+1 \le q \le m$. If *L*, *U* and *INC* are, respectively, a multiple of *g* then E is integer solvable iff E' is integer solvable.

Proof: **Omitted due to space limit.** ■

3.5. Time Complexity

A pair of same index variables with small enough coefficients is easily found according to Lemmas 3–1 and 3–2. In light of Lemmas 3–1 and 3–2, it is obvious that the worst-case time complexity to finding a pair of coefficients enough is O(*m*), where *m* is the number of variables in a non-continuous interval equation. The number of looking for all pairs of small enough coefficients in a non-continuous interval equation is at most $m/2$ times because the number of pairs moved in the non-continuous interval equation is at most $m/2$ pairs. Thus, the worst-case time complexity to move all pairs is $O(m^2)$.

To calculate the new non-continuous integer interval on the right-hand side of a non-continuous interval equation due to the movement of the qualified pairs actually is equivalent to apply a single *Banerjee-Wolfe inequality*. Applying a single Banerjee-Wolfe inequality to calculate the lower bound and the upper bound of the new non-continuous integer interval needs a constant time O(*y*), where *y* is a constant. Thus, for calculating all new non-continuous integer interval, the worst-case time complexity is O(*m*) because there are at most $m/2$ moves.

If all coefficients in a non-continuous interval equation have no absolute values of 1, then Lemma 3–3 employs the GCD test to reduce all coefficients to obtain small enough coefficients to justify the movement of a pair of same index variables to the right. In the worst cases, the non-continuous direction vector I test contains *m* GCD tests. That study [2] shows that a large percentage of all coefficients have absolute values of 1 in one-dimensional array references with linear subscripts in real programs. Therefore, the GCD test is not always applied to reduce all coefficients in the equations inferred from one-dimensional array references with linear subscripts in real programs because all coefficients in the equations have at least an absolute value of 1. The worst-case time complexity to the non-continuous direction vector I test to testing those one-dimensional array references with linear subscripts in real programs is immediately derived to be $O(m^2)$. The worst-case time complexity of the direction vector I test is also $O(m^2)$ [9]. Therefore, it is inferred that the non-continuous direction vector I test still remains the efficiency of the direction vector I test.

4. Experimental Results

We have tested our method and performed experiments on the codes abstracted from two numerical packages: Vector Loop and Livermore [10, 11]. 603 pairs of tested one-dimensional array references consisting of the same pair of array references with different direction vectors were observed under *constant* bounds and *non-one-increment*. The proposed method is only applied to test those one-dimensional arrays with subscripts under constant bounds and non-one-increment. It is very clear from Table 1 that the proposed method could properly solve whether there are *definitive results* for one-dimensional arrays with subscripts under constant bounds and non-one-increment.

Benchmark	The number of *definitive results*
Vector Loop	522
Livermore	81

Table 1. The result is to solve whether there are integer-valued solutions for one-dimensional arrays with subscripts under constant bounds and non-one-increment.

5. Conclusions

According to the time complexity analysis, the proposed method remains the efficiency of the direction vector I test. Therefore, assume that depending on the application domains and environments, the proposed method can be applied independently or together with other famous methods to analyze data dependence for linear-subscript array references.

References

[1] Kleanthis Psarris, David Klappholz, and Xiangyun Kong. "On the Accuracy of the Banerjee Test," Journal of Parallel and Distributed Computing, 12(2), June 1991, pp. 152-158.

[2] Xiangyun Kong, David Klappholz and Kleanthis Psarris. "The I Test," IEEE Transaction on Parallel and Distributed Systems," Vol. 2, No. 3, July 1991, pp. 342–359.

[3] Weng-Long Chang, Chih-Ping Chu and J. Wu, "A Multi-dimensional Version of the I Test," *Parallel Computing*, Vol. 27-13, Sept. 2001, pp. 1783-1799.

[4] Weng-Long Chang, Chih-Ping Chu and J. Wu, "A Precise Dependence Analysis for Multi-dimensional Arrays Under Specific Dependence Direction," Journal of System and Software. (Accepted, 2001).

[5] Weng-Long Chang and Chih-Ping Chu. "The Generalized Direction Vector I Test," *Parallel Computing*, Vol. 27-8, July 2001, pp. 1117-1144.

[6] Weng-Long Chang and Chih-Ping Chu. "The Infinity Lambda Test: A Multi-dimensional Version of Banerjee Infinity Test," *Parallel Computing*, Vol. 26-10, Aug. 2000, pp. 1275-1295.

[7] Weng-Long Chang, Chih-Ping Chu, and Jesse Wu. "The Generalized Lambda Test: A Multi-dimensional Version of Banerjee's Algorithm," International Journal of Parallel and Distributed Systems and Networks, Vol. 2, Issue 2, 1999, pp. 69-78.

[8] Weng-Long Chang and Chih-Ping Chu. "The Extension of the I Test," Parallel Computing, Vol. 24, Number 14, Nov. 1998, pp. 2101-2127.

[9] Kleanthis Psarris, Xiangyun Kong, David Klappholz. "The Direction Vector I test," IEEE Transaction on Parallel and Distributed Systems, Vol. 4, No. 11, 1993, pp. 1280–1290.

[10] David Levine, David Callahan and Jack Dongarra, "A comparative study of automatic vectorizing compilers," Parallel Computing 17(1991), pp.1223-1244.

[11] W. Blume and R. Eigenmann. "Performance analysis of parallelizing compilers on the perfect benchmark S® programs," IEEE Transaction on Parallel and Distributed Systems, Vol. 3, No. 6 (November 1992), pp. 643–656.

Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04)
1087-4089/04 $ 20.00 © 2004 IEEE