The Generalized Lambda Test

Weng-Long Chang Chih-Ping Chu
Institute of Information Engineering, National Cheng Kung University
Tainan, Taiwan 701, R.O.C.
Jesse Wu
Department of Mathematics, National Cheng Kung University
Tainan, Taiwan 701, R.O.C.
E-mail: chucp@server2.iie.ncku.edu.tw

Abstract

In this paper, we generalize the λ test. The generalized λ test can be applied towards determining whether there exist data dependences for coupled arrays with both constant and variable limits under any given direction vectors, improving the applicable range of the λ test. Experimental data reflecting the effect of the generalized λ test are also presented.

1. Introduction

The study in [9] shows that about 46 percent and 2 percent of examined two-dimensional and three-dimensional array references, respectively, have coupled subscripts. The problem of data dependence analysis for *m*-dimensional array references, each with *n* index variables, can be reduced to that of checking whether a system of *m* linear equations with *n* unknown variables has a simultaneous integer solution, which satisfies the constraints for each variable in the system. It is assumed that *m* linear equations in a system are written as (1–1): $a_{1,1}X_1 + \cdots + a_{1,n}X_n = 0, \cdots, a_{m,1}X_1 + \cdots + a_{m,n}X_n = 0,$ where each $a_{i,j}$ is an integer for $1 \le i \le m$ and $1 \le j \le n$. It is postulated that the constraints to each variable in (1–1) are represented as (1–2): $P_{r,0} + \sum_{s=1}^{r-1} P_{r,s}X_s \le X_r \le Q_{r,0} + \sum_{s=1}^{r-1} Q_{r,s}X_s,$ where $P_{r,0}$, $Q_{r,0}$, $P_{r,s}$ and $Q_{r,s}$ are integers for $1 \le r \le n$. If each of $P_{r,s}$ and $Q_{r,s}$ is zero in the limits of (1–2), then (1–2) will be reduced to (1–3): $P_{r,0} \le X_r \le Q_{r,0}$, where $1 \le r \le n$. That is, the bounds for each variable X_r are constants.

The *Banerjee Test* (*Banerjee algorithm* and *Banerjee inequalities*) only handles one linear equation with the bounds of (1–2) or (1–3) under given direction vectors [2]. The I test and the Direction Vector I test figure out integer solutions for one linear equation with constant bounds and given direction vectors [3, 4]. The λ test extends the Banerjee inequalities to allow *m* linear equations (1–1) under the constraints of (1–3) and given direction vectors to be tested simultaneously [1]. The Power and Omega tests gain more accurate outcomes, but they have exponential worst-case time complexity [6, 7, 8]. For array references with nonlinear subscripts, the range test can be applied to test data dependency [5]. In the rest of this paper the organization is depicted as follows. In Section 2, the problem of data dependence is reviewed. The summary accounts of the λ test are presented. In Section 3, the theoretical aspects and the worst-case time complexity of the generalized λ test are described. Experimental results showing the advantages of the generalized λ test are given in Section 4. Finally, brief conclusions are given in Section 5.

2. Background

It is assumed that S_1 and S_2 are two statements within a general loop. The general loop is presumed to contain *d* common loops. Statements S_1 and S_2 are postulated to be embedded in *d*+*p* loops and *d*+*q* loops, respectively. An array *A* is supposed to appear simultaneously within statements S_1 and S_2. If statement S_2 uses the element of the array *A* defined first by another statement S_1, then S_2 is true-dependent on S_1. If the statement S_2 defines the element of the array *A* used first by another statement S_1, then S_2 is anti-dependent on S_1. If the statement S_2 redefines the element of the array *A* defined first by another statement S_1, then S_2 is output-dependent on S_1.

Coupled references are groups of reference positions sharing one or more index variables [1, 6]. The Banerjee inequalities are first applied to test each equation in (1–1). If every equation intersects *V*, the convex set defined by the constraints of each variable in (1–1), then the λ test is employed to simultaneously check every equation. The λ test forms linear combinations of coupled references that eliminate one or more instances of index variables when direction vectors are not considered. While direction vectors are considered, the λ test generates new linear combinations that use a pair of relative index variables. Simultaneous constrained real-valued solutions exist if and only if the Banerjee inequalities find solutions in all the linear combinations generated [1].

3. The Generalized λ Test

Suppose that f_i and g_i are the lower bound function and the upper bound function for the i-th variable in (1–1) with n unknown variables under the constraints of (1–2), where $2 \le i \le n$. The original constraints (1–2) for each variable in (1–1) are rewritten as $P_{1,0} \le X_1 \le Q_{1,0}$ and $f_i(X_1, \cdots, X_{i-1}) \le X_i \le g_i(X_1, \cdots, X_{i-1})$,

where $f_i(X_1, \cdots, X_{i-1}) = P_{i,0} + \sum_{s=1}^{i-1} P_{i,s} X_s$ and

$$g_i(X_1, \cdots, X_{i-1}) = Q_{i,0} + \sum_{s=1}^{i-1} Q_{i,s} X_s \text{ for } 2 \le i \le n. \qquad (3\text{–}1)$$

The theories of the proposed test method are started to state from the case of m=2, for the convenience of presentation.

3.1. Two Dimensional Array References

In the case of two dimensional array references, two equations in (1–1) are $F_1 = 0$ and $F_2 = 0$, where $F_i = a_{i,0} + a_{i,1} X_1 + \cdots + a_{i,n} X_n$ for $1 \le i \le 2$. A linear equation for convenience is directly referred as a hyperplane in R^n. By [1], an arbitrary linear combination of the two equations can be written as $\lambda_1 F_1 + \lambda_2 F_2 = 0$. The domain of (λ_1, λ_2) is the whole R^2 space. Let $F_{\lambda_1, \lambda_2} = \lambda_1 F_1 + \lambda_2 F_2$; that is $F_{\lambda_1, \lambda_2} = (\lambda_1 a_{1,1} + \lambda_2 a_{2,1}) X_1 + \cdots + (\lambda_1 a_{1,n} + \lambda_2 a_{2,n}) X_n$. F_{λ_1, λ_2} is viewed in two ways. With (λ_1, λ_2) fixed, F_{λ_1, λ_2} is a linear function of $(X_1, \cdots, X_n)$ in R^n. With $(X_1, \cdots, X_n)$ fixed, it is a linear function of (λ_1, λ_2) in R^2. Furthermore, the coefficient of each variable in F_{λ_1, λ_2} is a linear function of (λ_1, λ_2) in R^2, i.e., $\Psi^{(i)} = \lambda_1 a_{1,i} + \lambda_2 a_{2,i}$ for $1 \le i \le n$. The equation $\Psi^{(i)} = 0$, $1 \le i \le n$, is called a Ψ equation. Each Ψ equation corresponds to a line in R^2, which is called a Ψ line. Each Ψ line separates the whole space into two closed halfspaces $\Psi_i^+ = \{(\lambda_1, \lambda_2) \mid \Psi^{(i)} \ge 0\}$ and $\Psi_i^- = \{(\lambda_1, \lambda_2) \mid \Psi^{(i)} \le 0\}$ that intersect at the Ψ line. A nonempty set $C \subset R^m$ is a cone if $\varepsilon \vec{\lambda} \in C$ for each $\vec{\lambda} \in C$ and $\varepsilon \ge 0$. It is obvious that each cone contains the zero vector. Moreover, a cone that includes at least one nonzero vector $\vec{\lambda}$ must consists of the "ray" of $\vec{\lambda}$, namely $\{\varepsilon \vec{\lambda} | \varepsilon \ge 0\}$ Such cones can clearly be viewed as the union of rays. There are at most n Ψ lines which together divide R^2 into at most $2n$ regions. Each region contains the zero vector. Any one nonzero element $\vec{\lambda}$ and the zero vector in the region forms the ray of $\vec{\lambda}$, namely $\{\varepsilon \vec{\lambda} | \varepsilon \ge 0\}$ Therefore, each region can be viewed as the union of the rays. It is very obvious from the definition of a cone that each region is a cone [1].

In the following, Lemmas 3–1 to 3–3 are an extension of Lemmas 1 to 3 in [1], respectively; Definitions 3–1 to 3–3 are cited from [1, 2] directly.

Lemma 3–1: Suppose that a bounded convex set V is defined simply by the limits of (3–1). If $F_{\lambda_1, \lambda_2} = 0$ intersects V for every (λ_1, λ_2) in every Ψ line, then $F_{\lambda_1, \lambda_2} = 0$ also intersects V for every (λ_1, λ_2) in R^2.

If the constraints of (3–1) plus dependence directions define V, we have a similar lemma. The following definition cited from [2] will first define the new limits for each pair of relative variables with a given dependence direction.

Definition 3–1: Given m linear equations (1–1) beneath the constraints of (3–1) and a specific direction vector $\vec{\theta} = (\theta_1, \cdots, \theta_d)$, where d refers to the number of common loops, if $\theta_k \in \{<, >\}$, $1 \le k \le d$, then the bounds of (3–1) for each pair of relative variables will be redefined, assuming $X_{2k-1} \theta_k X_{2k}$ and X_{2k-1} and X_{2k} refer to the same loop indexed variable. The new constraints for X_{2k-1} and X_{2k} are defined as follows: if $\theta_k \in \{<\}$ then (3–2): $f_{2k-1}(X_1, \cdots, X_{2k-2}) \le X_{2k-1} \le g_{2k-1}(X_1, \cdots, X_{2k-2}) - 1$ and $f_{2k}(X_1, \cdots, X_{2k-1}) = 1 + X_{2k-1} \le X_{2k} \le g_{2k}(X_1, \cdots, X_{2k-1})$; if $\theta_k \in \{>\}$, then (3–3): $f_{2k-1}(X_1, \cdots, X_{2k-2}) \le X_{2k-1} \le g_{2k-1}(X_1, \cdots, X_{2k-2})$ and $f_{2k}(X_1, \cdots, X_{2k-1}) \le X_{2k} \le X_{2k-1} - 1 = g_{2k}(X_1, \cdots, X_{2k-1})$.

Then we should discuss the rule to select the minimum and maximum limits for each pair of relative variables when their corresponding dependence directions are given. Each dependence direction is known to relate a unique pair of loop indices, which are associated with one of the common loops. Obviously, we should choose the new constraints (3–2) or (3–3) for each pair of relative variables and the original bounds (3–1) for other variables not to be constrained by dependence directions such that F_{λ_1, λ_2} has the minimum value and the maximum value. Let $\Phi_{(2k-1, 2k)}$ be the sum of the coefficients of X_{2k-1} and X_{2k} in F_{λ_1,λ_2} where X_{2k-1} and X_{2k} are related by a dependence direction, i.e., $\Phi_{(2k-1,2k)} = \lambda_1(a_{1,2k-1} + a_{1,2k}) + \lambda_2(a_{2,2k-1} + a_{2,2k})$ [1]. The minimum point and maximum point of F_{λ_1,λ_2} in *V*, in the presence of dependence directions, depend not only on the sign of the coefficient of each variable but also on the sign of $\Phi_{(2k-1,2k)}$, as clearly undertaken from statements above. From [1], the equation $\Phi_{(2k-1,2k)} = 0$ is called a Φ equation. Each Φ equation corresponds to a Φ line in R^2. There are at most *n/2* Φ lines. All Φ lines and Ψ lines divide R^2 space into at most *3n* regions. Each region is still a cone.

Lemma 3–2: Suppose that a bounded convex set *V* is denoted by the limits of (3–1) as well as dependence directions. If $F_{\lambda_1,\lambda_2} = 0$ intersects *V* for each (λ_1, λ_2) in every Φ line and every Ψ line, then $F_{\lambda_1,\lambda_2} = 0$ also intersects *V* for every (λ_1, λ_2) in R^2.

As a matter of fact, it suffices to test a single point in each Φ line or each Ψ line for determining whether F_{λ_1,λ_2} intersects *V* for every (λ_1, λ_2) in those lines.

Lemma 3–3: Suppose that a bounded convex set *V* is denoted by the limit of (3–1) and dependence directions. Given a line in R^2 corresponding to an equation $a\lambda_1 + b\lambda_2 = 0$, if $F_{\lambda_1,\lambda_2} = 0$ intersects *V* in R^n for any fixed point $(\lambda_1^0, \lambda_2^0) \neq (0, 0)$ in the line, then for every (λ_1, λ_2) in the line, $F_{\lambda_1,\lambda_2} = 0$ also intersects *V*.

Definition 3–2: Given an equation of the form $a\lambda_1 + b\lambda_2 = 0$ where a, b are not zero simultaneously, a canonical solution of the equation is defined as follows: $(\lambda_1, \lambda_2) = (1, 0)$, if $a = 0$; $(\lambda_1, \lambda_2) = (0, 1)$, if $b = 0$; $(\lambda_1, \lambda_2) = (b, -a)$, if neither a, b is zero.

Definition 3–3: The Λ set is denoted to be the set of all canonical solutions to Φ equations and Ψ equations. The hyperplane in R^n corresponding to $\lambda_1 F_1 + \lambda_2 F_2 = 0$, where (λ_1, λ_2) is a canonical solution in the Λ set, is called a λ plane.

There are at most *n* Ψ equations if *V* is denoted by the bounds of (3–1) only. There are at most *n* Ψ and *n*/2 Φ equations if *V* is defined by the limits of (3–1) and dependence directions. Each of the equations generates a canonical solution according to Definition 3–2. Each canonical solution forms a λ plane in light of Definition 3–3. Obviously, λ planes tested are at most *n* if *V* is defined by the constraints of (3–1) only, and are at most 3*n*/2 if *V* is denoted by the bounds of (3–1) as well as dependence directions. If *V* is denoted by the constraints of (1–3) only, then there are no more than *n* hyperplanes in the set [1]. If the bounds of (1–3) as well as dependence directions define *V*, then there are no more than 3*n*/2 hyperplanes in the set [1]. It is at once concluded that the number of λ planes tested by the generalized version of the λ test is the same as that of λ planes checked by the λ test.

We use the following example to explain the enhanced power of the generalized λ test over the λ test, when it is used to test data dependence for variable bounds under given dependence directions.

Consider the following equations

$$X_1 - X_4 = 0 \qquad \text{(ex3)}$$

$$-X_2 + X_3 = 0 \qquad \text{(ex4)}$$

subject to the bounds $1 \le X_1 \le 10, 1 \le X_2 \le 10, 1 - 2X_1 \le X_3 \le 10 + X_1$ and $1 - 2X_2 \le X_4 \le 10 + X_2$, and the limits of a direction vector $X_1 < X_2$ and $X_3 < X_4$.

According to Definition 3–1, the constraints for X_1, X_2, X_3 and X_4 will be redefined by $1 \le X_1 \le 9$, $1 + X_1 \le X_2 \le 10$, $1 - 2X_1 \le X_3 \le 9 + X_1$ and $1 + X_3 \le X_4 \le 10 + X_2$. The Banerjee algorithm is first used to resolve the problem. The extreme values computed by the Banerjee algorithm to (ex3) are –19 and 25, respectively. The Banerjee algorithm assumes that there are real-valued solutions because $-19 \le 0 \le 25$. Similarly, the extreme values for (ex4) are –27 and 8, respectively. The Banerjee algorithm also assumes that there are real-valued solutions because $-27 \le 0 \le 8$. Therefore, the Banerjee algorithm concludes that there may be real-valued solutions. If the generalized λ test is then used to resolve the same problem, the Ψ and Φ equations are generated. The Ψ equations are $0 * \lambda_1 + 1 * \lambda_2 = 0$ and $1 * \lambda_1 + 0 * \lambda_2 = 0$. The Ψ equations have two canonical solutions (1, 0) and (0, 1). Each canonical solution yields one λ plane corresponding to an original hyperplane inferred from (ex3) or (ex4). The two λ planes are not necessarily tested by the generalized λ test because the Banerjee algorithm has tested them. The Φ equations are $\lambda_1 - \lambda_2 = 0$ and $-\lambda_1 + \lambda_2 = 0$. The Λ set from the Φ equations is easily determined: $\Lambda = \{(1, 1)\}$. The canonical solution (1, 1) in the Λ set gives the λ plane: $X_1 - X_2 + X_3 - X_4 = 0$. Based on the Banerjee algorithm, the maximum value for the λ plane is immediately inferred to be -2. Similarly, the minimum value to the λ plane can also be derived to be -38. Since the maximum value for the λ plane is less than zero, the generalized λ test in light of Lemmas 3–1 to 3–3 infers that there is no solution.

3.2. Multidimensional Array References

We take account of m linear equations in (1–1) with $m > 2$ for generalizing the generalized λ test. All m linear equations are assumed to be connected; otherwise they can be partitioned into smaller systems. As stated before, we can hypothesize that there are no redundant equations. By [1], an arbitrary linear combination of m linear equations can be written as $\sum_{i=1}^{m} \lambda_i F_i = 0,$ where $F_i = \sum_{j=1}^{n} a_{i,j} X_{i,j}$. Let $F_{\lambda_1, \cdots, \lambda_m} = \sum_{i=1}^{m} \lambda_i F_i,$ and then $F_{\lambda_1, \cdots, \lambda_m} = (\sum_{j=1}^{m} \lambda_j a_{j,1}) X_1 + \cdots + (\sum_{j=1}^{m} \lambda_j a_{j,n}) X_n.$ It is to be determined whether $F_{\lambda_1, \cdots, \lambda_m} = 0$ intersects V in R^n space for arbitrary $(\lambda_1, \cdots, \lambda_m).$ The coefficient of each variable in $F_{\lambda_1, \cdots, \lambda_m}$ is a linear function of $(\lambda_1, \cdots, \lambda_m)$ in $R^m,$ which is $\Psi^{(i)} = \sum_{j=1}^{m} \lambda_j a_{j,i}$ for $1 \le i \le n.$ The Ψ equation $\Psi^{(i)} = 0,$ $1 \le i \le n,$ is called a Ψ equation. A Ψ equation corresponds to a hyperplane in $R^m,$ called a Ψ plane. Each Ψ plane divides the whole space into two closed halfspaces $\Omega_i^+ = \{(\lambda_1, \cdots, \lambda_m) | \Psi^{(i)} \ge 0\}$ and $\Omega_i^- = \{(\lambda_1, \cdots, \lambda_m) | \Psi^{(i)} \le 0\}.$ Let $\Phi_{(2k-1,2k)}$ be the sum of the coefficients of X_{2k-1} and X_{2k} in $F_{\lambda_1, \cdots, \lambda_m}$ where X_{2k-1} and X_{2k} are related by a dependence direction, i.e., $\Phi_{(2k-1,2k)} = \sum_{i=1}^{m} \lambda_i (a_{i,2k-1} + a_{i,2k}).$ The equation $\Phi_{(2k-1,2k)} = 0,$ is called a Φ equation. A Φ equation corresponds to a hyperplane in R^m , which is called a Φ plane. Each Φ plane separates the whole space into two closed halfspaces $\delta^+_{(2k-1,2k)} = \{(\lambda_1, \cdots, \lambda_m) | \Phi_{(2k-1,2k)} \ge 0\}$ and $\delta^-_{(2k-1,2k)} = \{(\lambda_1, \cdots, \lambda_m) | \Phi_{(2k-1,2k)} \le 0\}.$ If V is defined by the constraints of (3–1) only, then a nonempty set $\bigcap_{i=1}^{n} \Omega_i,$ where $\Omega_i \in \{\Omega_i^+, \Omega_i^-\}$ is called a λ region. If the bounds of (3–1) as well as dependence directions denote V, then a nonempty set $(\bigcap_{i=1}^{n} \Omega_i) \cap (\bigcap_{k=1}^{d} \delta_{(2k-1,2k)}),$ where $\Omega_i \in \{\Omega_i^+, \Omega_i^-\}$ and $\delta_{(2k-1,2k)} \in \{\delta^+_{(2k-1,2k)}, \delta^-_{(2k-1,2k)}\}$ is called a λ region. The intersection of $\delta_{(2k-1,2k)}$ is taken for all pairs of index variables, which are related by a dependence direction. Every λ region is a cone in R^m space. The λ regions in R^m space have several lines as the frame of their boundaries. Each line (called a λ line) is the intersection of some Ψ and Φ planes.

In the following, Lemmas 3–4 and 3–5 are an extension of Lemmas 5 and 6 in [1], respectively.

Lemma 3–4: If $F_{\lambda_1, \cdots, \lambda_m} = 0$ intersects V for every $(\lambda_1, \cdots, \lambda_m)$ in every λ line, then $F_{\lambda_1, \cdots, \lambda_m} = 0$ also intersects V for every $(\lambda_1, \cdots, \lambda_m)$ in R^m space.

Lemma 3–5: Given a line in R^m which crosses the origin of the coordinates, if $F_{\lambda_1, \cdots, \lambda_m} = 0$ intersects V in R^n space for any fixed point $(\lambda_1^0, \cdots, \lambda_m^0) \ne (0, \cdots, 0)$ in the line, then for every $(\lambda_1, \cdots, \lambda_m)$ in the line, $F_{\lambda_1, \cdots, \lambda_m} = 0$ also intersects V.

There is a finite set of hyperplanes in R^m space such that S intersects V if and only if every hyperplane in the set intersects V. If V is denoted by the constraints of (1–3) only, then there are no more than $\binom{n}{m-1}$ hyperplanes in the set [1]. If V is defined by the limits of (3–1) only, then the number of hyperplanes in the set is also at most $\binom{n}{m-1}.$ If V is denoted by the bounds of (1–3) as well as dependence directions, then there are no more than $\binom{3n/2}{m-1}$ hyperplanes in the set [1]. If V is defined by the limits of (3–1) and dependence directions, then the number of hyperplanes is also at most $\binom{3n/2}{m-1}$ in the set. It is right away derived that the number of λ planes tested in the generalized λ test is the same as that of λ planes checked in the λ test. The detail of the generalized λ test in the general case is not considered since the discussion is similar to the case of $m = 2$.

3.3 Time Complexity

The common phases for the λ test and the generalized λ test include (1) calculating λ values and (2) examining each λ plane. λ values are easily determined according to Φ equations, Ψ equations and Definition 3–2. It is clear that the time complexity to computing a λ value is $O(y)$ from Definition 3–2, where y is a constant. Each λ value corresponds to a λ plane. Each λ plane is tested to see if it intersects V, by checking its minimum and maximum values. The extreme values can be calculated from the Banerjee inequalities and also computed from the Banerjee algorithm. The time complexity to the Banerjee inequalities and the Banerjee algorithm are $O(z)$ and $O(z^2)$, respectively, where z is the number of variables in the λ plane. Hence, the time complexity of the λ test and the generalized λ test for examining a λ

plane is at once derived to be $O(z)$ and $O(z^2)$, respectively. The number of λ planes checked to the λ test and the generalized λ test is the same, and is at most $\binom{3n/2}{m-1}$, where *m* is the number of coupled dimensions and *n* is the number of variables in coupled references, in light of statements in Section 3.3 [1]. Therefore, the worst-case time complexity for the λ test and the generalized λ test is immediately inferred to be $O\left(\begin{bmatrix}3n/2\\m-1\end{bmatrix}*(z+y)\right)$ and $O\left(\begin{bmatrix}3n/2\\m-1\end{bmatrix}*\left(z^2+y\right)\right)$, respectively. Two-dimensional arrays with coupled subscripts appear quite frequently in real programs, as clearly indicated from statements in Section 1. The number of examining λ planes to each two-dimensional array tested is at most $3n/2$ according to statements in Section 3.2 [1]. If the λ test and the generalized λ test are applied to deal with the array, then their worst-case time complexity is $O(3n/2*(z+y))$ and $O\left(3n/2*\left(z^2+y\right)\right)$, respectively. The number of checking λ planes is almost 1 due to the regularity of coefficients in coupled subscripts in real two-dimensional arrays tested. Hence, the worst-case time complexity of the λ test and the generalized λ test for testing those real two-dimensional arrays is nearly equal to $O(z+y)$ and $O(z^2+y)$, respectively. The generalized λ test slightly decreases the efficiency of the λ test because the number of variables, *z*, in the λ plane tested is generally very small.

4. Experimental Results

We tested the generalized λ test and performed experiments on Personal Computer Intel 80486 through the codes cited from five numerical packages EISPACK, LINPACK, Parallel loops, Livermore loops and Vector loops [11, 12]. The codes include more than 37000 lines of statements, and 17433 pairs of array references consisting of the same pair of array references with different direction vectors were found to have coupled subscripts. The λ test detected that there were no data dependences for 6722 pairs of coupled arrays under constant bounds and direction vectors. The generalized λ test checked that there were no data dependences for 3826 pairs of coupled arrays beneath variable limits and direction vectors as well as 6722 pairs of coupled arrays subject to constant constraints and direction vectors.

The generalized λ test in our experiments is only applied to test those arrays with coupled subscripts. The generalized λ test found 10548 cases that had no data dependence. The improvement rate can be affected by two factors. First, the frequency of coupled subscripts. Second, the "success rate" of the generalized λ test, by which we mean how often a generalized λ test detects a case where there is no data dependence. Let *b* be the number of the coupled subscripts found in our experiments, and let *c* be the number that is detected to have no data dependence from the coupled subscripts. Thus the success rate is denoted to be equal to c/b. In our experiments, 17433 pairs of array references were found to have coupled subscripts, and 10548 of them were found to have no data dependence. So the success rate in our experiments was about equal to 60.5 percent. The generalized λ test increases the success rate of the λ test. The increasing success rate was about 21.9 percent.

In our experiments, λ planes always subsumed the hyperplane from each dimension of an array reference. Note that the Banerjee algorithm needs first to test these hyperplanes. Comparing with the Banerjee algorithm, the generalized λ test examined a total number of 17473 additional λ planes in our experiments. That is, almost every generalized λ test had examined only one additional λ plane. In light of this fact, the additional time needed by the generalized λ test is very small.

Timing results are shown in Table 1. Each row shows how much additional time (compared with the Banerjee algorithm) was needed in the generalized λ test. For instance, the first row shows that there were 66 subroutines in which the generalized λ test consumed no more than 40 percent additional time. For most of the subroutines (294 out of 310), the generalized λ tests never need more than 100 percent additional time. Additional time was spent mostly on 1) calculating λ values, and 2) examining each λ plane.

Test time increase	Number of subroutines
0%–40%	66
41%–70%	105
71%–90%	83
91%–100%	40
101%–150%	11
200%–250%	5

Table 1. Timing results of the generalized λ test.

The Power test and the Omega test were also employed to resolve 10548 pairs of arrays with coupled references, respectively. These two tests were found to acquire the same accurate results as the generalized λ test. Suppose that k_1, k_2 and k_3 are the execution time to treat data dependence problem of a coupled-subscript array for the generalized λ test, Power test and Omega test, subsequently. The speed-up in Table 2 is defined to be the set of k_2/k_1 and k_3/k_1. Each row in table 2 shows how many times the execution time of the Power and Omega tests took longer than the execution time of the generalized λ test. For example, the first row shows that there were 132 subroutines in which the execution time of the two tests took from 16.4 to 22.7 times longer than that of the generalized λ test. For all of the subroutines in our experiments, the execution time of the two tests was indicated from Table 2 to take from 7.1 to 22.7 times longer than the execution time of the generalized λ test. This indicates that for multidimensional arrays with coupled subscripts the efficiency of the generalized λ test is much better than that of the Power test and the Omega test.

Speed-up	Total number of subroutines
16.4–22.7	132
14.7–15.9	210
13.1–14.5	166
12.5–12.9	80
10.3–12.3	22
7.1–8.3	10

Table 2. The speed-up of the generalized λ test when compared with the Power and Omega tests.

5. Conclusions

The generalized λ test enhances significantly data dependence analysis of λ test when there are coupled subscripts in multidimensional array references. The generalized λ test only ascertains whether real-valued solutions exist because, like the λ test, it is based on equality consistency checking. The generalized λ test is exactly equivalent to a multidimensional version of the Banerjee algorithm because it can determine simultaneous constrained real-valued solutions. Li [1] found that the λ test for coupled array references under constant bounds usually increases the cost of the *Banerjee inequalities* by a factor of two or less. The generalized λ test for coupled array references beneath variable limits usually also increases the cost of the *Banerjee algorithm* by a factor of two or less, as shown from our experimental results.

The Power test is a combination of Fourier-Motzkin variable elimination with an extension of Euclid's GCD algorithm [6, 7]. The Omega test combines new methods for eliminating equality constraints with an extension of Fourier-Motzkin variable elimination [8]. The two tests currently have the highest precision and the widest applicable range in the field of data dependence analysis for testing array references with linear subscripts. However, the cost of the two tests is very expensive because the worst-case of Fourier-Motzkin variable elimination is exponential in the number of free variables [6, 7, 8]. Triolet [10] found that using Fourier-Motzkin variable elimination for dependence testing takes from 22 to 28 times longer than the Banerjee inequalities. The Range test is now the highest precision and the widest applicable range in the field of data dependence analysis for testing array references with nonlinear subscripts. In our experiments, the generalized λ test, Power test and Omega test share the same accurate results for 10548 pairs of arrays with coupled references. However, the efficiency of the generalized λ test is much better than that of the Power test and Omega test.

The generalized λ test extends the applicable range of the λ test and, according to the time complexity analysis, only slightly decreases the efficiency of the λ test. Therefore, the generalized λ test seems to be a practical scheme to analyze data dependence for coupled-subscript array references.

References

[1] Z. Li, P.-C. Yew and C.-Q. Zhu. An Efficient Data Dependence Analysis for Parallelizing Compilers. *IEEE Transaction on Parallel and Distributed Systems*, Vol. 1, No. 1, pp. 26−34, January 1990.

[2] U. Banerjee. *Dependence Analysis for Supercomputing*, Kluwer Academic Publishers, 1988, ISBN 0−89838−289−0.

[3] X. Kong, D. Klappholz and K. Psarris. The I Test. *IEEE Transaction on Parallel and Distributed Systems*, Vol. 2, No. 3, pp. 342−359, July 1991.

[4] X. Kong, D. Klappholz and K. Psarris. The Direction Vector I Test. *IEEE Transaction on Parallel and Distributed Systems*, Vol. 4, No. 11, pp. 1280−1290, November 1993.

[5] W. Blume and R. Eigenmann. The Range Test: A Dependence Test for Symbolic, Non-linear Expressions. *IEEE Supercomputing*, Washington, D.C., pp. 528−537, November, 1994.

[6] M. Wolfe and C.-W. Tseng. The Power Test for Data Dependence. *IEEE Transaction on Parallel and Distributed Systems*, Vol. 3, No. 5, pp. 591−601, September 1992.

[7] M. Wolfe. *High Performance Compilers for Parallel Computing*, Addison-Wesley Publishing Company, 1996, ISBN 0−8053−2730−4.

[8] W. Pugh. A Practical Algorithm for Exact Array Dependence Analysis. *Communication of the ACM*, 35(8):102−114, August 1992.

[9] Z. Shen, Z. Li and P.-C. Yew. An Empirical Study of Fortran Programs for Parallelizing Compilers. *IEEE Transaction on Parallel and Distributed Systems*, Vol. 1, No. 3, pp. 356−364, July 1992.

[10] R. Triolet, F. Irigoin and P. Feautrier. Direct Parallelization of Call Statements. In Proc. SIGPLAN Symp. Compiler Construction, Palo Alto, CA, pp. 176−185, June 1986.

[11] B. J. Smith et al. *Matrix Eigensystem Routines-Eispack Guide*, Heidelberg: Springer, 1976.

[12] W. Blume and R. Eigenmann. Performance Analysis of Parallelizing Compilers on the Perfect Benchmarks® Programs. *IEEE Transaction on Parallel and Distributed Systems*, Vol. 3, No. 6, pp. 643−656, November 1992.