А. С. Антонов, Вл. В. Воеводин, В. М. Репин

Алгоритм канонизации описания структуры программ в рамках линейной модели*

1 Введение

Доказано, что для большого класса программ можно построить параметрическое описание их *графа алгоритма* (информационной истории) средствами только статического анализа в следующем виде [1,2]. Для каждого входа каждого оператора анализируемой программы указывается конечное число троек (F, Δ_I, Δ_N). Δ_N — линейный многогранник в пространстве Ω_N внешних переменных программы, значения которых на момент построения графа неизвестны. Δ_I — линейный многогранник в пространстве итераций Ω_I, размеры и положение которого могут зависеть от точек многогранника Δ_N. Функция F линейная и имеет вид $F = Jx + \Phi N$, где N — это вектор внешних переменных программы, а J, Φ — целочисленные матрицы. Областью определения функции F является многогранник Δ_I. Значениями функций являются точки пространства итераций, причем точка x задает конечную вершину дуги, а выражение $Jx + \Phi N$ — начальную.

Многогранники Δ_I и Δ_N задаются системами линейных уравнений и неравенств с целочисленными коэффициентами, и никаких дополнительных ограничений на них, вообще говоря, не накладывается. Вместе с тем, часто возникает необходимость привести данное описание к *канонической форме*, которая подразумевает выполнение следующих условий:

*Работа поддержана Российским Фондом Фундаментальных Исследований, грант 96–01–01433

• никакие из многогранников Δ_I, Δ_N не являются целочисленно пустыми, т.е. все они содержат хотя бы одну точку с целочисленными координатами;

• ни одно из ограничений многогранников Δ_I, Δ_N не является избыточным;

• ни одно из ребер многогранников Δ_I не может стягиваться в точку ни при каких значениях внешних переменных из соответствующего многогранника Δ_N.

Необходимость приведения к канонической форме возникает по многим причинам. Во-первых, в процессе построения самого графа алгоритма желательно сразу выкидывать пустые многогранники, так как в противном случае время построения может значительно увеличиться (здесь и далее под пустотой будем понимать отсутствие в многограннике точки с целочисленными координатами). Во-вторых, многие свойства программ, определяемые по виду функций F, могут искажаться именно на пустых областях, что недопустимо, и поэтому все тройки с пустыми областями должны быть удалены. В-третьих, для генерации эффективных программ в описании многогранников не должно быть избыточных ограничений, так как это приведет к появлению лишних проверок в условных операторах новой программы. И, наконец, целый класс методов анализа структуры и свойств программ требует выполнения последнего условия канонического описания, которое гарантирует сохранность "формы" всех многогранников Δ_I.

Данная методика канонизации сразу ориентирована на практическое применение. Сначала излагаются простые, но эффективные методы, основанные на достаточных условиях и работающие в большинстве реальных ситуаций, и лишь в конце идет полная проверка.

Итак, пусть дана область в пространстве итераций $\Delta_I = \{AI \leq BN+C\}$, где $I \in Z^{n\times1}$ — вектор параметров циклов, $N \in Z^{k\times1}$ — вектор внешних переменных, $A \in Z^{m\times n}$, $B \in Z^{m\times k}$ и $C \in Z^{m\times1}$ — заданные матрицы и вектор. Кроме того, заданы ограничения на внешние переменные $\Delta_N = \{FN \leq D\}$, где $F \in Z^{l\times k}$ и $D \in Z^{l\times1}$ — заданные матрица и вектор. В некоторых случаях в описании областей Δ_N и Δ_I

вместо неравенств могут быть равенства. В дальнейшем, и равенства, и неравенства будем называть *ограничениями.*

Описываемый далее алгоритм канонизации опирается на работу [3]. Все действия алгоритма производятся над ограничениями с целочисленными коэффициентами. В дальнейшем существенно, что ни одно действие алгоритма не нарушает этого условия.

2 Описание алгоритма канонизации

2.1 Нормализация

Сначала производится процесс, называемый в [3] *нормализацией.* Он заключается в приведении НОД коэффициентов при переменных в описании областей Δ_I и Δ_N к 1. Кроме того, что нормализация необходима для проведения Омега-теста, описываемого далее, она полезна и по другим соображениям. Во-первых, уменьшаются абсолютные значения коэффициентов при переменных. Во-вторых, в некоторых случаях после нормализации область сужается, или упрощается дальнейший анализ. Например, при поиске параллельных гиперплоскостей до нормализации нужно искать ограничения с пропорциональными коэффициентами, а после нормализации — лишь с одинаковыми по модулю коэффициентами.

Для того, чтобы нормализовать ограничение $a_1 i_1 + \ldots + a_n i_n \leq b_1 n_1 + \ldots + b_k n_k + c$, вычисляется наибольший общий делитель g коэффициентов $a_1, \ldots, a_n, b_1, \ldots, b_k$, и все коэффициенты, включая c, делятся на g. Если c не делится нацело на g, то в качестве свободного члена берётся $\lfloor c/g \rfloor$, при этом гиперплоскость в пространстве $\Omega_I \times \Omega_N$, соответствующая ограничению, сдвигается до ближайшей целочисленной точки, а анализируемая область сужается. Если нормализуется ограничение-равенство $a_1 i_1 + \ldots + a_n i_n = b_1 n_1 + \ldots + b_k n_k + c$, и c не делится нацело на g, то данное уравнение не имеет целочисленных решений, и, следовательно, вся область будет целочисленно пустой. Таким образом нормализуются ограничения как из Δ_I, так и из Δ_N.

2.2 Исключение переменных из равенств

Для исключения переменной из равенства недостаточно действовать обычным методом — домножением коэффициентов ограничения на коэффициент из другого ограничения при исключаемой переменной, так как в этом случае можно решить только задачу определения рациональной непустоты области, после чего остается нерешенным вопрос, принадлежит ли области хотя бы одна целочисленная точка.

Исключение переменных из равенств, изложенное в [3], позволяет уменьшить размерность рассматриваемой области и упростить все последующие алгоритмы. При этом коэффициент при исключаемой переменной всегда приводится к единице, что позволяет решить не только задачу определения рациональной непустоты, но и задачу определения целочисленной непустоты.

2.2.1 Исключение переменных из равенств, входящих в описание области из пространства внешних переменных

Чтобы исключить переменную из равенства $f_1 n_1 + \ldots + f_k n_k = d$, входящего в описание области из пространства внешних переменных, сначала проверим, нет ли такого j, что $|f_j| = 1$. Если такое j есть, то просто выражаем переменную n_j через остальные переменные и свободный член, после чего выражение для n_j подставляем во все остальные ограничения.

Если ограничение не содержит коэффициентов, по модулю равных 1, то обозначим через p коэффициент с наименьшим абсолютным значением, и пусть $s = |f_p| + 1$. Определим $\widehat{mod}$ следующим образом:

$$a \ \widehat{mod} \ b = a - b\lfloor a/b + 1/2 \rfloor.$$

Определим новую переменную σ такую, что

$$s\sigma = \sum_j (f_j \ \widehat{mod} \ s) n_j + d \ \widehat{mod} \ s.$$

Заметим, что $f_p \; \widehat{mod} \; s = -sign(f_p)$. Поэтому выразим из данного равенства переменную n_p:

$$n_p = -sign(f_p)s\sigma + \sum_{j \neq p} sign(f_p)(f_j \; \widehat{mod} \; s)n_j + sign(f_p)d \; \widehat{mod} \; s$$

и подставим во все ограничения. Для исходного равенства после подстановок и нормализации получим

$$-|f_p|\sigma + \sum_{j \neq p} \left(\lfloor f_j/s + 1/2 \rfloor + (f_j \; \widehat{mod} \; s) \right) n_j + \lfloor d/s + 1/2 \rfloor + d \; \widehat{mod} \; s = 0.$$

Модуль коэффициента при переменной σ получается равным модулю коэффициента при исходной переменной n_p, а все остальные коэффициенты уменьшаются по крайней мере в полтора раза. Поэтому после нескольких повторений этого алгоритма при какой-нибудь переменной получится коэффициент, по модулю равный 1, что позволит нам исключить эту переменную.

Таким образом можно исключить столько переменных, сколько имеется равенств в описании области Δ_N. Соответственно на столько же уменьшится размерность пространств Ω_N и $\Omega_I \times \Omega_N$.

2.2.2 Исключение переменных из равенств, входящих в описание области из пространства итераций

Исключаться из равенств, входящих в описание области пространства итераций, должны только переменные пространства итераций, либо, если после некоторого количества исключений в равенстве останутся только внешние переменные, то исключается одна из них с запоминанием выполняемых подстановок. Подстановки при исключении и при введении новых переменных должны запоминаться для последующего добавления к описаниям целочисленно непустых областей.

Рассмотрим одно из ограничений-равенств, входящих в описание области пространства итераций: $\Gamma = a_1 i_1 + \ldots + a_n i_n = b_1 n_1 + \ldots + b_k n_k + c$. После нормализации наибольший общий делитель коэффициентов при всех переменных равен 1, однако для коэффициентов при переменных пространства итераций это может быть не так. Пусть g — наибольший общий делитель коэффициентов при переменных пространства итераций. Тогда

- если $g = 0$, то в ограничение входят только внешние переменные; в этом случае описанным ранее образом исключается одна из них с запоминанием производимых подстановок;
- если $g = 1$, то исключаем одну из переменных пространства итераций аналогично тому, что было описано для пространства внешних переменных;
- если $g > 1$, то вводим новую переменную σ и добавляем ограничение

  $$g\sigma = \sum_{j=1}^{n}(a_j \widehat{mod}\ g)i_j - \sum_{j=1}^{k}(b_j \widehat{mod}\ g)n_j - (c \widehat{mod}\ g).$$

  Переменную σ в дальнейшем используем как внешнюю переменную. После исключения переменной из полученного равенства и нормализации получаем, что в исходном равенстве наибольший общий делитель коэффициентов при переменных пространства итераций равен 1.

Таким образом можно исключить столько переменных, сколько имеется равенств в описании области Δ_I. Соответственно на столько же уменьшится размерность пространств Ω_I и $\Omega_I \times \Omega_N$.

2.3 Эвристические методы

Применяемые здесь эвристики не дают полного и точного решения в общем случае, но, как показывает практика, позволяют покрыть достаточно большой класс реальных фрагментов программ. При этом важно, что описываемые эвристики достаточно простые и проверяются быстро. Сначала ищем параллельные гиперплоскости
$\Gamma_i = a_{i1}i_1 + \ldots + a_{in}i_n \leq b_{i1}n_1 + \ldots + b_{ik}n_k + c_i$,

$\Gamma_j = a_{j1}i_1 + \ldots + a_{jn}i_n \le b_{j1}n_1 + \ldots + b_{jk}n_k + c_j$, $1 \le i \ne j \le m$,
такие, что $\frac{a_{i1}}{a_{j1}} = \ldots = \frac{a_{ik}}{a_{jk}} = p$.

Умножая ограничение Γ_i на $|a_{j1}|$ и ограничение Γ_j на $|a_{i1}|$, получим
$\Gamma'_i = a'_{i1}i_1 + \ldots + a'_{in}i_n \le b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i$,
$\Gamma'_j = a'_{j1}i_1 + \ldots + a'_{jn}i_n \le b'_{j1}n_1 + \ldots + b'_{jk}n_k + c'_j$, $1 \le i \ne j \le m$,
где $\frac{a'_{i1}}{a'_{j1}} = \ldots = \frac{a'_{ik}}{a'_{jk}} = p'$, где $p' = 1$ или $p' = -1$.

В первом случае ($p' = 1$) ищем и отбрасываем избыточные неравенства: если $b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i \le b'_{j1}n_1 + \ldots + b'_{jk}n_k + c'_j$, то ограничение Γ'_j является избыточным, и его можно отбросить, иначе можно отбросить избыточное ограничение Γ'_i. При этом вся область Ω_N пространства внешних переменных может разбиться на две подобласти, в одной из которых избыточно ограничение Γ'_i, а в другой — Γ'_j. В дальнейшем для обеих областей строится каноническое представление.

Во втором случае ($p' = -1$) получаем слой, в котором могут располагаться целочисленные точки:
$-(b'_{j1}n_1 + \ldots + b'_{jk}n_k + c'_j) \le a'_{i1}n_1 + \ldots + a'_{ik}n_k \le b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i$.
Для него производим следующие проверки:

1. Проверка на противоречивость (рациональная пустота): если для какой-либо пары параллельных гиперплоскостей $-(b'_{j1}n_1 + \ldots + b'_{jk}n_k + c'_j) > b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i$, то область является рационально (а значит и целочисленно) пустой, поэтому противоположное условие добавляется к ограничениям пространства внешних переменных.

2. Поиск противоположных неравенств, замена их на равенства и исключение переменных из этих равенств: если $-(b'_{j1}n_1 + \ldots + b'_{jk}n_k + c'_j) = b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i$, то заменяем эти два неравенства на равенство $a'_{i1}n_1 + \ldots + a'_{ik}n_k = b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i$ и производим исключение переменной из данного равенства.

3. Необходимое условие целочисленной непустоты области, в описание которой входят пары параллельных гиперплоскостей — между такими гиперплоскостями должна быть по крайней мере одна

целочисленная точка: должно $\exists s \in Z$:

$$-(b'_{j1}n_1 + \ldots + b'_{jk}n_k + c'_j) \le s \le b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i$$

и $a'_{i1}n_1 + \ldots + a'_{ik}n_k = s$. Второе верно $\Leftrightarrow s = rq$, где $r \in Z$, а $q = НОД(a'_{i1}, \ldots, a'_{ik})$. Поэтому должно $\exists r \in Z : \frac{-(b'_{j1}n_1+\ldots+b'_{jk}n_k+c'_j)}{q} \le r \le \frac{b'_{i1}n_1+\ldots+b'_{ik}n_k+c'_i}{q}$. Этим добавляется дополнительное ограничение на внешние переменные. Для параллельных гиперплоскостей в пространстве внешних переменных проверка такого условия заключается только в сравнении свободных членов.

4. Если область задана парами гиперплоскостей, параллельных осям координат, то предыдущее условие становится необходимым и достаточным: поскольку любые ограничения имеют вид $-(b'_{j1}n_1 + \ldots + b'_{jk}n_k + c'_j) \le a'_{it}i_t \le b'_{i1}n_1 + \ldots + b'_{ik}n_k + c'_i$, то если из каждого такого двойного неравенства определим переменную i_t, то получим целочисленную точку пространства итераций, удовлетворяющую всем ограничениям. Для параллельных гиперплоскостей в пространстве внешних переменных, как и в предыдущем пункте, достаточно произвести сравнение свободных членов.

2.4 Исключения Фурье-Моцкина

Прямой ход метода Фурье-Моцкина[4] позволяет определить рациональную пустоту или непустоту исходной области. Однако, если для последней исключаемой при помощи этого метода переменной целочисленной точки нет, то можно утверждать, что и вся область целочисленно пуста, то есть в таком случае задача будет решена до конца.

Один шаг исключений Фурье-Моцкина состоит в следующем. Сначала выбирается переменная z, которая будет исключаться на этом шаге. Если все ограничения на z имеют вид либо $\beta \le bz$, либо $az \le \alpha$, где a и b — положительные целые числа, то при выборе достаточно большого или, соответственно, достаточно маленького значения переменной z можно удовлетворить все такие неравенства. Поэтому все неравенства, содержащие эту переменную, можно удалить и дальше не рассматривать. Пусть теперь существуют ограничения как того, так и другого типа. Тогда для каждой пары ограничений $\beta \leq bz$ и $az \leq \alpha$ составим двойное неравенство $a\beta \leq abz \leq b\alpha$. Для того, чтобы существовала рациональная точка, удовлетворяющая этому неравенству, необходимо, чтобы выполнялось условие $a\beta \leq b\alpha$. Выполнив такое действие для всех пар ограничений, получим описание области в пространстве на единицу меньшей размерности, которая называется *real shadow*.

Если на каком-либо шаге исключений получим, что одно из ограничений сверху на некоторую переменную меньше какого-либо ограничения снизу на ту же переменную, то можно утверждать, что область рационально пуста. Если для последней исключаемой переменной не получим противоречия, то можно утверждать, что область является рационально непустой. Если между максимальным из ограничений снизу и минимальным из ограничений сверху на последнюю исключаемую переменную нет ни одной целочисленной точки, то можно утверждать, что исходная область целочисленно пуста. Обратное, вообще говоря, не верно.

После исключения всех переменных пространства итераций получается дополнительный набор ограничений на внешние переменные. Это приводит к сужению области пространства внешних переменных. В дальнейшем целочисленную точку имеет смысл искать только в полученной области.

2.5 Исключения Омега-теста

Омега-тест позволяет определить целочисленную непустоту для большого числа областей.

Один шаг исключений Омега-теста состоит в следующем. Сначала выбирается переменная z, которая будет исключаться на этом шаге. Если все ограничения на z имеют вид либо $\beta \leq bz$, либо $az \leq \alpha$, где a и b — положительные целые числа, то при выборе достаточно большого или, соответственно, достаточно маленького значения переменной z можно удовлетворить все такие неравенства. Поэтому все неравенства, содержащие эту переменную, можно удалить и дальше не рассматривать. Пусть теперь существуют ограничения как того, так и другого типа. Тогда для каждой пары ограничений $\beta \le bz$ и $az \le \alpha$ составим двойное неравенство $a\beta \le abz \le b\alpha$.

Определим условие, при котором существует целочисленное решение неравенства $a\beta \le b\alpha$, но не существует целочисленного решения $a\beta \le abz \le b\alpha$, то есть между $a\beta$ и $b\alpha$ нет целых чисел, делящихся на ab. Пусть $i = \lfloor \beta/b \rfloor$. Тогда

$$ab\, i < a\beta \le b\alpha < ab(i+1).$$

Так как $ab(i+1) - b\alpha > 0$ и все коэффициенты целочисленные $\Rightarrow$

$$ab(i+1) - b\alpha \ge 1 \Rightarrow a(i+1) - \alpha \ge 1/b \Rightarrow$$

$$a(i+1) - \alpha \ge 1 \Rightarrow ab(i+1) - b\alpha \ge b.$$

Аналогично $a\beta - abi \ge a$. Отсюда $b\alpha - a\beta \le ab - a - b$. Поэтому если выполняется обратное неравенство $b\alpha - a\beta > ab - a - b + 1 = (a-1)(b-1)$, то существует целочисленное решение исходного двойного неравенства. Область в пространстве на 1 меньшей размерности, определяемая условиями

$$b\alpha - a\beta > (a-1)(b-1)$$

для каждой пары неравенств $\beta \le bz$ и $az \le \alpha$, называется *dark shadow.*

Если для всех ограничений снизу $b = 1$ или для всех ограничений сверху $a = 1$, то dark shadow совпадает с real shadow, получение которой описано в предыдущем пункте. В этом случае рациональная непустота области влечёт за собой целочисленную непустоту. Чем меньше коэффициенты a и b, тем ближе граница dark shadow к границе real shadow, а значит, тем большая часть области будет исчерпана при проверке Омега-теста. Поэтому для исключения выбирается в первую очередь переменная с возможно меньшими коэффициентами.

Если для последней исключаемой переменной будет найдено целочисленное значение, удовлетворяющее всем ограничениям, то можно утверждать, что исходная область является целочисленно непустой, а значит задача будет решена до конца.

Исключив таким образом все переменные пространства итераций, получим дополнительный набор ограничений на внешние переменные. Область в пространстве внешних переменных при этом разбивается на две, в одной из которых (где удовлетворяются условия, полученные в результате работы Омега-теста) определено, что целочисленная точка есть и, следовательно, остается только получить неизбыточное описание, а в другой (если она не пуста) нужно продолжать поиск.

2.6 Перебор гиперплоскостей

Если исключениями Омега-теста не получили целочисленной точки, то остаётся проверить только точки между границами dark shadow и real shadow. Выше показано, что если существует целочисленная точка, то должно выполняться

$$ab - a - b + a\beta \geq b\alpha \geq abz \geq a\beta.$$

Определим максимальный из коэффициентов a в ограничениях сверху на z, и для каждого ограничения снизу $bz \geq \beta$ проверяем, существует ли целочисленная точка на гиперплоскости $bz = \beta + i$ для всех i таких, что $(ab - a - b)/a \geq i \geq 0$.

Перебор гиперплоскостей, находящихся в обычно достаточно узкой области между границами real shadow и dark shadow, позволяет решить задачу определения целочисленной пустоты или непустоты до конца. При этом перебор, производимый описанным выше способом, позволяет минимизировать количество перебираемых гиперплоскостей. Этот процесс может быть весьма долгим, но, к счастью, он редко необходим на практике, поэтому он производится в последнюю очередь. Решение задачи для каждой такой гиперплоскости может привести к новому разбиению области в пространстве внешних переменных. Перебор заканчивается, если оказывается исчерпанной вся область Δ_N или если будут перебраны все гиперплоскости.

2.7 Получение неизбыточного описания

На этапе проверки эвристик мы избавились от избыточных гиперплоскостей, параллельных другим гиперплоскостям. Теперь для полученных описаний целочисленно непустых областей реализуем основной алгоритм получения неизбыточного описания. Для каждой гиперплоскости $\Gamma = a_1 i_1 + \ldots + a_n i_n \leq b_1 n_1 + \ldots + b_k n_k + c$ рассматриваем гиперплоскость $\Gamma' = a_1 i_1 + \ldots + a_n i_n \geq b_1 n_1 + \ldots + b_k n_k + c + 1$, и если область $(\Omega_I \setminus \Gamma) \cap \Gamma'$ целочисленно пуста для любых значений внешних переменных, то ограничение Γ является избыточным, и мы можем его отбросить. Если же эта область целочисленно пуста лишь для некоторых значений внешних переменных, то соответствующая область в пространстве внешних переменных разбивается на две подобласти, в одной из которых ограничение Γ избыточно, а в другой существенно.

2.8 Гарантия сохранности формы многогранников

Для завершения канонизации описания полученных многогранников нужно гарантировать, что в каждой из полученных целочисленно непустых областей с неизбыточным описанием ни одно ребро не стягивается в точку. Это означает, что нужно произвести новое подразбиение области пространства внешних переменных таким образом, чтобы описание каждой полученной подобласти было каноническим.

Условие нестягивания ребер многогранника в точку эквивалентно тому, что никакие две вершины не должны совпасть при изменении значений внешних переменных. Это можно проверить следующим образом. Для поиска вершин в пространстве итераций перебираем все наборы из n ограничений. Если система, составленная из этих ограничений, совместна, то для принадлежности вершины многограннику необходимо, чтобы она удовлетворяла всем остальным ограничениям. Если вершина удовлетворяет этим ограничениям только для некоторой подобласти в области Δ_N пространства внешних переменных, то это приводит к дополнительному подразбиению области Δ_N.

3 Заключение

На основании описываемых методов в рамках проекта V-Ray разработан комплекс программ, который для заданной области пространства итераций выдает список целочисленно непустых подобластей и, если требуется, для полученных областей строится неизбыточное описание. Важно отметить, что описываемый и реализованный алгоритм позволяет решить поставленную задачу до конца для любой исходной области. Однако, как показывает практика анализа реальных программ, почти все случаи разрешаются до конца уже на этапе проверки эвристик, и крайне редко приходится производить перебор гиперплоскостей.

Литература

[1] Воеводин. В. В. *Математические основы параллельных вычислений*// М.: Изд. Моск. ун-та, 1991. 345с.

[2] Воеводин. Вл. В. *Теория и практика исследования параллелизма последовательных программ*// Программирование. 1992. № 3. С. 38-53.

[3] Pugh. W. *A Practical Algorithm for Exact Array Dependence Analysis*// Communications of the ACM. August 1992. Vol. 35, № 8. P. 102-114.

[4] Danzig G., Eaves. B. C. *Fourier-motzkin elimination and its dual*// Journal of Combinatorial Theory. 1973. A(14). P. 288-297.