

SAND93-0074

Неограниченное издание

Январь 1993

МНОГОМЕРНАЯ СПЕКТРАЛЬНАЯ БАЛАНСИРОВКА ЗАГРУЗКИ

Брюс Хендриксон, Роберт Лиланд.

Sandia National Laboratories Albuquerque, 1993

Резюме

Мы описываем алгоритм для балансирования загрузки научных вычислений, который обобщен и улучшен использованием спектрального деления пополам. Благодаря использованию различных собственных векторов, наш новый спектральный алгоритм может делить вычисление на 4 или 8 частей сразу. Эти многомерные спектральные алгоритмы разбиения генерируют сбалансированные разделения, которые имеют более низкие коммуникационные накладные расходы и менее дороги для вычислений, чем произведенные спектральным делением пополам. Кроме того, они автоматически уменьшают число сообщений на архитектуре сети или гиперкуба. Эти спектральные разделение далее улучшены многомерным обобщением алгоритма разбиения Керниган-Лин (Kernigan-Lin). Преводятся результаты на нескольких вычислительных сетках, сравниваются с другими популярными методами.

1. **Введение.** Эффективное использование распределенной памяти для параллельных вычислений требует, чтобы вычислительная нагрузка была сбалансирована на

процессорах так, чтобы минимизировать связь между процессорами. Это требование отображения можно отразить на проблеме из теории графов, в которой узлы представляют вычисление, грани представляют связь, и цель состоит в том, чтобы назначить равное число вершин на каждый процессор таким способом, который минимизирует число граней, пересекающихся между процессорами. Практический опыт показал, что качество этого отображения может иметь существенное влияние на производительность, следовательно имеется значительный интерес в эффективных алгоритмах отображения.

Новая эвристика для разделения графа в контексте отображения параллельных вычислений была недавно предложена Саймоном (Simon) [15]. Этот спектральный метод рекурсивно делит пополам граф, рассматривая собственный вектор ассоциированной матрицы, чтобы выяснить основные свойства графа. Метод получил много внимания, т.к. предлагает хорошее соотношение между универсальностью, качеством и эффективностью. Идея относительно использования собственных векторов для разделения графов, возникла в начале 70-ых у Доната и Хоффмана (Donath, Hoffman) [3,4], и Файдера (Fieder) [6,7]. Недавно эти идеи были восстановлены и далее разработаны несколькими авторами [2,12,13,14].

С одним исключением, вся эта предыдущая работа базировалась на приложении рекурсивного деления пополам. Рендел (Rendl) и Волкович (Wolkowicz) описывают алгоритм для разделения графа на произвольное число наборов без рекурсии [14]. Однако, их алгоритм требует, чтобы k собственных векторов матрицы, представляющей граф были определены, чтобы делить граф на k частей. Это делает алгоритм непрактичным для деления больших графов на сотни или тысячи наборов. Главная особенность работы, описанной в этой бумаге

- то, что предлагается практическое обобщение спектрального деления пополам, чтобы было возможным деление на более чем два набора сразу. Это особенно актуально в контексте современных параллельных супервычислений.

Мы описываем два новых алгоритма, спектральное деление на четыре (quadrisection) и спектральное деление на восемь (octasection), которые делят на четыре набора, используя два собственных вектора и на восемь наборов, используя три собственных вектора соответственно. Оценки эффективности разделения сразу на несколько наборов показали, что новые алгоритмы производят лучшее разделение чем спектральное деление пополам (Манхэттан показатель – показатель пересылки на гиперкубе). Эмпирическое изучение показало, что этот показатель является подходящей мерой для моделирования производительности машин с архитектурой гиперкуба, уменьшение этого показателя соответствует уменьшению скопления связей в пределах вычислительной сети [9]. Показатель пересылки также подходит для двух- или трех- мерных сетевых архитектур (с или без циклических соединений). Большинство параллельных машин с распределенной памятью имеют или сетевую или гиперкубическую архитектуру, следовательно эти новые алгоритмы имеют широкое применение.

В дополнение к созданию лучших разбиений, эти многомерные спектральные алгоритмы имеют два других существенных преимущества. Во первых, логарифмические отношения между числом требуемых собственных векторов и числа разделений означают, что на практике новые методы фактически требуют меньше сетевых вычислений чем спектральное деление пополам для определения того же числа разделений. Во вторых, используя доступную избыточность в области решений, многомерные спектральные методы способны

правильно делить сильно симметричные графы, которые спектральная бисекция делит с трудом.

Многие из данных результатов взяты из недавно законченного доклада [10]. Цель этого документа состоит в том, чтобы представить эти результаты в менее формальном и более понятном виде. Кроме того, содержимое § 6 и § 7 еще не издавалось. В § 2 этого документа, мы описываем спектральное деление пополам и приводим математическую и нотационную базу для дальнейших разделов. В § 3 - § 5 мы описываем наши спектральные алгоритмы деления на четыре и восемь (quadrisection, octasection).

Физический объект, чье поведение имитирует алгоритмы разбиения, описан в § 6, новый алгоритм, частично совершенствующий спектральное разделение представлен в § 7. Результаты некоторых типовых вычислений представлены в § 8, далее следуют заключения в § 9.

2. Спектральное деление пополам. Различная формулировка спектрального деления пополам может быть найдена в публикациях нескольких авторов [2,12,15]. Мы определяем задачу проблему следующим образом.

Граф G представлен : набор вершин V и набор граней E . V_i соответствует вершине с индексом i , $E_{i,j}$ означает грань между V_i и V_j , \sum_{V_i} и $\sum_{E_{ij}}$ означают суммы по вершинам и граням соответственно. Теперь назначим переменную x_i каждому V_i так что $x_i = \pm 1$ и $\sum_{V_i} x_i = 0$.

Первое условие подразумевает разбиение на два различных набора. Второе требует чтобы наборы были равного размера, учитывая четность исходного количества. Назовем этот вектор вектор-индикатор, т.к. он показывает принадлежность каждой вершины к набору.

Следующий шаг должен обратить внимание, что функция $f(\mathbf{x}) = \frac{1}{4} \sum_{E_{ij}} (\mathbf{x}_i - \mathbf{x}_j)^2$ определяет число граней, пересекающихся между наборами, $(\mathbf{x}_i - \mathbf{x}_j)^2$ ничего не вносит в сумму, если X_i и X_j имеют тот же самый знак, и вносит 4, если они имеют противоположный знак.

Теперь, когда мы имеем объективную функцию, чтобы минимизировать, мы преобразовываем $f(\mathbf{x})$ к матричной форме, так как это будет делать решение более очевидным. Вначале, определим матрицу смежности:

$$A_{ij} = \begin{cases} 1, & \text{если } (V_i, V_j) \in E \\ 0, & \text{иначе} \end{cases} \quad (1)$$

Мы также определим матрицу степени $D = \text{diag}(d_i)$, где d_i — это степень V_i , т.е. число граней, смежных вершине V_i .

Продолжим преобразования следующим образом:

$$\sum_{E_{ij}} (\mathbf{x}_i - \mathbf{x}_j)^2 = \sum_{E_{ij}} (\mathbf{x}_i^2 + \mathbf{x}_j^2) - \sum_{E_{ij}} 2\mathbf{x}_i \mathbf{x}_j \quad (2)$$

$$\sum_{E_{ij}} 2\mathbf{x}_i \mathbf{x}_j = \sum_{V_i} \sum_{V_j} \mathbf{x}_i A_{ij} \mathbf{x}_j = \sum_{V_i} \mathbf{x}_i \sum_{V_j} A_{ij} \mathbf{x}_j = \mathbf{x}^T \mathbf{A} \mathbf{x} \quad (3)$$

$$\sum_{E_{ij}} (\mathbf{x}_i^2 + \mathbf{x}_j^2) = \sum_{E_{ij}} 2 = 2|E| = \sum_{V_i} \mathbf{x}_i^2 d_i = \mathbf{x}^T \mathbf{D} \mathbf{x} \quad (4)$$

Здесь \mathbf{x}^T — транспонированный вектор \mathbf{x} .

Получаем:

$$\sum_{E_{ij}} (\mathbf{x}_i - \mathbf{x}_j)^2 = \mathbf{x}^T \mathbf{D} \mathbf{x} - \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T (\mathbf{D} - \mathbf{A}) \mathbf{x} \quad (5)$$

Теперь определим матрицу Лапласа для графа G:

$$L_{ij} = \begin{cases} -1, & \text{если } (V_i, V_j) \in E, i \neq j \\ d_i, & \text{если } i = j \\ 0, & \text{иначе} \end{cases} \quad (6)$$

Отсюда, $L = D - A$

$$f(x) = \frac{1}{4} x^T L x \quad (7)$$

Соединяя это с ограничениями на вектор x , мы определяем дискретную проблему деления пополам:

$$\text{Минимизировать:} \quad \frac{1}{4} x^T L x \quad (8)$$

$$\text{При условии:} \quad x^T I = 0, \quad x_i = \pm 1$$

Где I – это n -мерный вектор $(1, 1, \dots, 1)^T$.

Так как разбиение графа – NP-трудная задача, необходимо ослабить ограничения дискретности на x и сформулировать новую непрерывную задачу:

$$\text{Минимизировать:} \quad \frac{1}{4} x^T L x \quad (9)$$

$$\text{При условии:} \quad x^T I = 0, \quad x^T x = n$$

В которой элементы x могут брать любые значения, удовлетворяющие ограничениям. Эта непрерывная проблема – только приближение к дискретной проблеме, и значения, определяющие ее решение должны быть отображены назад к

± 1 в соответствии с некоторой соответствующей схемой. Идеально, когда решение близко к ± 1 .

Мы теперь начинаем решать (9), отмечая, что матрица Лапласа имеет несколько важных свойств.

Если $u^1, u^2 \dots$ – нормализованные собственные векторы L с соответствующими собственными значениями $\lambda_1 \leq \lambda_2 \leq \lambda_3 \dots$, следующая теорема доказана в [10].

Теорема 2.1. Матрица L имеет следующие свойства:

(I) L – симметрична;

(II) u^i попарно ортогональны;

(III) $u^1 = n^{-0.5} \mathbf{1}$, $\lambda_1 = 0$;

(IV) Если граф замкнутый, то только λ_1 принимает нулевое значение.

Затем выразим X в терминах собственных векторов L : $x = \sum a_i u^i$, где a_i – вещественные константы, такие, что $\sum (a_i)^2 = n$. Второе свойство гарантирует, что это всегда возможно. Заменой для x мы получаем функцию, для минимизации, зависящую от собственного значения матрицы Лапласа λ_2 .

$f(x) = 0.25(a_2^2 \lambda_2 + a_3^2 \lambda_3 + \dots + a_n^2 \lambda_n)$ начиная с $\lambda_1 = 0$.

Очевидно, что

$(a_2^2 + a_3^2 + \dots + a_n^2) \lambda_2 \leq (a_2^2 \lambda_2 + a_3^2 \lambda_3 + \dots + a_n^2 \lambda_n)$

учитывая упорядоченность собственных величин $f(x) \geq n \lambda_2 / 4$.

Обратите внимание, что мы можем достигнуть $f(x) = n \lambda_2 / 4$, выбирая $x = \sqrt{n} u_2$. Обратите внимание также, что выбор x удовлетворяет ограничению равновесия $x^T \mathbf{1} = u_2^T u_i = 0$, (свойства II и III).

Поэтому, $x = \sqrt{n} u_2$ удовлетворяет ограничения и минимизирует $f(x)$, это – решение непрерывной проблемы. Если

$\lambda_2 \neq \lambda_3$, это решение уникально ($x = -\sqrt{nu_2}$ дает то же самое разбиение).

Остается решить задачу отображения решения непрерывной проблемы к дискретному разделению. В случае деления пополам имеется простой и естественный способ сделать это. Найдите медиану значений x_i , и затем отобразите вершины выше значения медианы в один набор, ниже в другой. Если несколько вершин имеют значение медианы, то они распределяются не нарушая равновесия. Это решение - самая близкая дискретная точка к непрерывному оптимуму.

После деления пополам, если значение медианы больше чем ноль, то подграф состоящий из вершин, оцененных больше медианы, связан. Подобно, если значение медианы - меньше чем ноль, то подграф, состоящий из вершин, оцененных ниже медианы, связан. В обоих случаях другой подграф может быть несвязным. Несвязный подграф проблематичен, потому что мы намереваемся применить метод рекурсивно, и метод может терпеть неудачу если применяется к разъединенному графу, так как Теорема 2.1 нарушается. Мы поэтому контролируем связность подграфов на каждой стадии рекурсии. Если несвязный подграф обнаружен, мы добавляем минимальное число фантомных граней, чтобы установить связность, делим этот подграф и затем удаляем фантомные грани. Мы находим, что в практике несколько подграфов станут несвязными в ходе деления типичного большого графа, и наша фантом-стратегия заметно улучшает полные результаты в этих случаях.