УДК 519.4

В. П. КИРЛИЦА

ТОЧНЫЕ *D*-ОПТИМАЛЬНЫЕ ПЛАНЫ ЭКСПЕРИМЕНТОВ ДЛЯ ЛИНИИ РЕГРЕССИИ С НЕРАВНОТОЧНЫМИ НАБЛЮДЕНИЯМИ

Рассмотрена модель неравноточных наблюдений

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon(x_i),\ i = 1, 2, \ldots, n,\ n \geq 2,$$

с некоррелированными ошибками наблюдений $\varepsilon(x_i)$ со средними значениями ноль и дисперсиями $D\{\varepsilon(x_i)\} = d(x_i) > 0,\ x_i \in [-1, 1]$. Показано, что точный *D*-оптимальный план экспериментов для модели неравноточных наблюдений (1) совпадает с планом для равноточных наблюдений $(d(x) = \text{const})$, если функция $d(x)$ удовлетворяет неравенству

$$d(x) \geq \frac{1}{4}\left((d_1 + d_2)x^2 + 2(d_2 - d_1)x + d_1 + d_2\right),\ d_1 = d(-1),\ d_2 = d(1).$$

Класс функций $d(x)$, удовлетворяющих (2), обширен: постоянные функции, с линейным изменением, вогнутые функции. Если неравенство (2) обращается в равенство не только в точках –1, 1, а и в других точках из отрезка (–1, 1), то точный *D*-оптимальный план экспериментов для модели неравноточных наблюдений (1) может отличаться от классического точного *D*-оптимального плана экспериментов для равноточных наблюдений.

Ключевые слова: линия регрессии; неравноточные наблюдения; *D*-оптимальные планы экспериментов; структура точных *D*-оптимальных планов.

In article the model of heteroscedastic observations

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon(x_i),\ i = 1, 2, \ldots, n,\ n \geq 2,$$

with not correlated errors of observations which have means zero and variances $D\{\varepsilon(x_i)\} = d(x_i) > 0,\ x_i \in [-1, 1]$ is considered. It is shown that the exact *D*-optimal design of experiments (1) coincides with the design for homoscedastic observations $(d(x) = \text{const})$, if function $d(x)$ satisfies to inequality

$$d(x) \geq \frac{1}{4}\left((d_1 + d_2)x^2 + 2(d_2 - d_1)x + d_1 + d_2\right),\ d_1 = d(-1),\ d_2 = d(1).$$

Class of functions satisfying to an inequality (2) is extensive: constant functions, with linear change, concave functions. If inequality (2) is transformed to equality not only in points –1, 1, but also in other points of interval (–1, 1), then *D*-optimal design of experiments for model of heteroscedastic observations can differ from the classical design for homoscedastic observations.

Key words: linear regression; heteroscedastic observations; *D*-optimal designs of experiments; structure of exact *D*-optimal designs.

Рассмотрим линейную модель наблюдений:

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon(x_i),\ i = 1, 2, \ldots, n,\ n \geq 2, \tag{1}$$

где y_i – наблюдаемые значения; θ_0, θ_1 – неизвестные параметры; x_i – контролируемые переменные из интервала $[-1, 1]$; $\varepsilon(x_i)$ – некоррелированные ошибки наблюдений с нулевым средним и ограниченной дисперсией $D\{\varepsilon(x_i)\} = d(x_i) > 0$ для каждой реализации $x_i \in [-1, 1]$.

Для равноточных наблюдений $(d(x) = \text{const})$ известно [1], что точный *D*-оптимальный план (*n* фиксировано) определяется следующим образом. Наблюдения надо проводить только в точках –1 либо 1. Для четного числа наблюдений $n = 2s$ в точке –1 проводится s наблюдений, остальные – в точке 1. Для нечетного $n = 2s + 1$ в точке –1 необходимо провести s или $s + 1$ наблюдений, остальные – в точке 1.

Покажем, что классические планы экспериментов, построенные для равноточных наблюдений, будут робастными (устойчивыми) и для определенного класса функций $d(x)$, описывающих изменение дисперсии. Класс функций $d(x)$ определим следующим образом:

$$d(x) \geq \frac{1}{4}\left\{[d_1 + d_2]x^2 + 2[d_2 - d_1]x + d_1 + d_2\right\},\ d_1 = d(-1) > 0,\ d_2 = d(1) > 0,\ x \in [-1, 1]. \tag{2}$$

Легко проверить, что построенному классу функций $d(x)$ удовлетворяют равноточные наблюдения $(d(x) = d = \text{const})$, линейное изменение дисперсии наблюдений $(d(x) = a + bx, a > 0, |b| < a)$, а также все вогнутые функции, неотрицательные на интервале $[-1, 1]$.

Лемма. *Для модели наблюдений* (1)*, у которой дисперсии наблюдений* $d(x)$ *удовлетворяют неравенству* (2)*, существует точный D-оптимальный план экспериментов, у которого все точки спектра лежат на концах интервала* [–1, 1].

Доказательство. Обозначим через k отношение дисперсий наблюдений в точках 1 и –1, т. е. $k = d_2/d_1$ Лемма доказана в [2] для случая $k = 1$. Убедимся в ее справедливости при $k \neq 1$. Для определенности будем считать, что $k > 1$. Для $k < 1$ доказательство проводится аналогично.

Рассмотрим произвольный план наблюдений ε, сосредоточенный в n точках $x_1, x_2, \ldots, x_n$. Покажем, что от плана ε можно перейти к плану ε_1 с точками спектра $x_i = \pm 1, i = \overline{1, n}$, для которого определитель информационной матрицы $M(\varepsilon_1)$ больше или равен определителю информационной матрицы $M(\varepsilon)$ плана ε, т. е. $|M(\varepsilon)| \leq |M(\varepsilon_1)|$. Действительно, точку x_1 спектра плана ε сделаем «плавающей», т. е. $x_1 = x \in [-1, 1]$. Остальные значения $x_i, i = \overline{2, n}$, фиксированы. Полученный план обозначим через ε_x. Вычисления показывают, что определитель информационной матрицы плана ε_x имеет вид

$$|M(\varepsilon_x)| = \frac{\alpha x^2 - 2\beta x + \gamma}{d(x)} + p, \tag{3}$$

где

$$d(x) = d(x_1),$$

$$\alpha = \sum_{i=2}^{n} \frac{1}{d(x_i)},\ \beta = \sum_{i=2}^{n} \frac{x_i}{d(x_i)},\ \gamma = \sum_{i=2}^{n} \frac{x_i^2}{d(x_i)},\ p = \sum_{i=2}^{n} \frac{1}{d(x_i)} \sum_{i=2}^{n} \frac{x_i^2}{d(x_i)} - \left(\sum_{i=2}^{n} \frac{x_i}{d(x_i)}\right)^2. \tag{4}$$

Используя (2), можно оценить $|M(\varepsilon_x)|$ сверху:

$$|M(\varepsilon_x)| \leq \frac{4}{d_1} f_1(x) + p, \tag{5}$$

где

$$f_1(x) = \frac{\alpha x^2 - 2\beta x + \gamma}{(1+k)x^2 + 2(k-1)x + 1 + k},\ k = \frac{d_2}{d_1}.$$

В точках –1 и 1 неравенство (5) обращается в равенство. Вычислим максимум $|M(\varepsilon_x)|$ по $x \in [-1, 1]$ и покажем, что максимум достигается на одном из концов этого интервала. Поскольку константы α, β, γ, p, d_1 не зависят от x, то в силу (5) максимизация $|M(\varepsilon_x)|$ сводится к максимизации $f_1(x)$ по x и выяснению того, что этот максимум достигается при $x = \pm 1$.

Производная функции $f_1(x)$ равна

$$\frac{df_1(x)}{dx} = \frac{2}{1+k} \frac{(\alpha\lambda + \beta)x^2 + (\alpha - \gamma)x - \beta - \gamma\lambda}{(x^2 + 2\lambda x + 1)^2},\ \lambda = \frac{k-1}{1+k}. \tag{6}$$

В (6) коэффициент $\lambda \in (0, 1)$ для $k > 1$. Из (4) следует, что $\alpha - \gamma \geq 0$. Если $\alpha\lambda + \beta = 0$ и $\alpha - \gamma = 0$, то производная (6) имеет постоянный знак на интервале $[-1, 1]$. Максимум $f_1(x)$ в этом случае достигается на одном из концов интервала $[-1, 1]$. Если $\alpha\lambda + \beta = 0$ и $\alpha - \gamma > 0$, поведение производной (6) зависит от положения корня μ линейного уравнения $(\alpha - \gamma)x - \beta - \gamma\lambda = 0$ относительно отрезка $[-1, 1]$. Если

$$\mu = \frac{\beta + \gamma\lambda}{\alpha - \gamma} \leq -1,$$

то производная $f_1'(x) > 0, x \in [-1, 1]$, и функция $f_1(x)$ достигает своего максимального значения при $x = 1$. Если $-1 < \mu < 1$, $\alpha\lambda + \beta = 0$, то на интервале $[-1, \mu)$ производная $f_1'(x) < 0$, а на интервале $(\mu, 1]$ производная $f_1'(x) > 0$. Это означает, что своего максимального значения функция $f_1(x)$ достигает на одном из концов интервала $[-1, 1]$. Если $\mu \geq 1$ и $\alpha\lambda + \beta = 0$, $\alpha - \gamma > 0$, то в этом случае производная (6) на интервале $[-1, 1]$ отрицательна, т. е. функция $f_1(x)$ достигает своего максимума на левом конце интервала $[-1, 1]$. Возможен случай, когда $\alpha\lambda + \beta = \alpha - \gamma = -\beta - \gamma\lambda = 0$. В этом случае функция $f_1(x)$ не зависит от x (можно считать, что ее максимальное значение достигается при $x = \pm 1$). В любом случае при $\alpha\lambda + \beta = 0$ максимум функции $f_1(x)$ достигается при $x = \pm 1$.

Если $\alpha\lambda + \beta < 0$, то знак производной (6) определяется знаком параболы $y = (\alpha\lambda + \beta)x^2 + (\alpha - \gamma)x - -\beta - \gamma\lambda$ на интервале $[-1, 1]$. Обозначим через $D = (\alpha - \gamma)^2 + 4(\alpha\lambda + \beta)(\beta + \gamma\lambda)$ дискриминант исследуемой параболы. Оказывается, что $D \geq 0$, если $\alpha\lambda + \beta < 0$. В этом случае парабола пересекает ось x в точках

$$x_1 = \frac{\alpha - \gamma - \sqrt{D}}{-2(\alpha\lambda + \beta)}, \ x_2 = \frac{\alpha - \gamma + \sqrt{D}}{-2(\alpha\lambda + \beta)}.$$

Покажем, что $-1 \leq x_1 < 0$, а $x_2 \geq 1$. Воспользуемся тем, что $\alpha \geq \gamma$, $\alpha > 0$, $\gamma \geq 0$. Поскольку $\alpha\lambda + \beta < 0$, то отсюда следует, что $\beta < 0$. Но тогда и $\beta + \gamma\lambda < 0$. Проверим это. Предположим противное, т. е. $\beta + \gamma\lambda \geq 0$. Отсюда вытекает, что $\beta \geq -\gamma\lambda$. С другой стороны, $\beta < -\alpha\lambda$. Имеем: $-\gamma\lambda \leq \beta < -\alpha\lambda$. Значит, $-\gamma\lambda < -\alpha\lambda$, или $\gamma > \alpha$. Получаем противоречие. Поскольку $(\alpha\lambda + \beta)(\beta + \gamma\lambda) > 0$, то $x_1 < 0$. Проверим теперь, что $x_1 \geq -1$. Если выполняется это неравенство, то оно сводится к неравенству

$$\sqrt{(\alpha - \gamma)^2 + 4(\alpha\lambda + \beta)(\beta + \gamma\lambda)} \leq \alpha - \gamma - 2(\alpha\lambda + \beta).$$

Правая часть последнего неравенства положительна. Обе части этого неравенства возведем в квадрат и получим неравенство

$$4(\alpha\lambda + \beta)(\beta + \gamma\lambda) \leq -4(\alpha - \gamma)(\alpha\lambda + \beta) + 4(\alpha\lambda + \beta)^2.$$

Поделим обе части этого неравенства на $\alpha\lambda + \beta < 0$. Приходим к неравенству $\gamma(\lambda - 1) \geq \alpha(\lambda - 1)$. Поскольку $\lambda - 1 < 0$, то окончательно получаем верное неравенство $\gamma \leq \alpha$. Итак, доказано, что $-1 \leq x_1 < 0$.

Покажем теперь, что $x_2 \geq 1$. Это неравенство сводится к следующему:

$$\sqrt{(\alpha - \gamma)^2 + 4(\alpha\lambda + \beta)(\beta + \gamma\lambda)} \geq \gamma - \alpha - 2(\alpha\lambda + \beta). \tag{7}$$

Если правая часть неравенства (7) меньше или равна нулю, то (7) выполняется. Если правая часть (7) положительна, то обе части неравенства можно возвести в квадрат. Получим

$$(\alpha\lambda + \beta)(\beta + \gamma\lambda) \geq -(\gamma - \alpha)(\alpha\lambda + \beta) + (\alpha\lambda + \beta)^2. \tag{8}$$

Разделим обе части (8) на $\alpha\lambda + \beta < 0$ и сведем его к верному неравенству $\gamma \leq \alpha$. Продолжим исследование производной (6) при условии, что $\alpha\lambda + \beta < 0$, $D \geq 0$. Если $x_1 = -1$, то на интервале $[-1, 1]$ производная (6) больше или равна нулю, т. е. максимум $f_1(x)$ достигается при $x = 1$. Если $-1 < x_1 < 0$, то на интервале $[-1, x_1)$ производная (6) отрицательна, а на интервале $(x_1, 1]$ она положительна, т. е. максимум $f_1(x)$ достигается в точке -1 либо 1.

Наконец, рассмотрим случай, когда $\alpha\lambda + \beta > 0$ и $D \geq 0$. Теперь корни уравнения $(\alpha\lambda + \beta)x^2 + +(\alpha - \gamma)x - \beta - \gamma\lambda = 0$ следующие:

$$x_1 = \frac{\gamma - \alpha - \sqrt{D}}{2(\alpha\lambda + \beta)}, \ x_2 = \frac{\gamma - \alpha + \sqrt{D}}{2(\alpha\lambda + \beta)}.$$

Аналогично тому, как исследовались корни квадратного уравнения в случае $\alpha\lambda + \beta < 0, D \geq 0$, можно показать, что $x_1 \leq -1$, $0 < x_2 \leq 1$. Если $x_2 = 1$, то на интервале $[-1, 1]$ производная (6) меньше или равна нулю, т. е. функция $f_1(x)$ достигает своего максимального значения в точке $x = -1$. Если $0 < x_2 < 1$, то на интервале $[-1, x_2)$ производная (6) отрицательна, а на интервале $(x_2, 1]$ она положительна. Следовательно, максимального значения функция $f_1(x)$ достигает в точке –1 либо 1.

Итак, максимум функции $f_1(x)$ достигается на концах интервала $[-1, 1]$. Зафиксируем значение x_1, при котором определитель информационной матрицы достиг своего максимального значения. Аналогичным образом повторим процесс максимизации определителя информационной матрицы последовательно по переменным $x_2, \ldots, x_n$. В итоге получим, что $|M(\varepsilon)| \leq |M(\varepsilon_1)|$, где ε_1 – план экспериментов, в котором все точки спектра равны ±1. Ясно, что один из точных D-оптимальных планов находится среди планов ε_1. Лемма доказана.

Опираясь на лемму, докажем следующую теорему.

Теорема 1. *Для модели неравноточных наблюдений* (1) *с дисперсией наблюдений* $d(x)$, *удовлетворяющей неравенству* (2), *точный D-оптимальный план экспериментов остается таким же, как и для равноточных наблюдений.*

Доказательство. Согласно доказанной лемме точный D-оптимальный план экспериментов должен иметь структуру

$$\varepsilon_n = \begin{Bmatrix} -1, & 1 \\ m, & n-m \end{Bmatrix},$$

где m – число наблюдений в точке –1, $0 < m < n$. Чтобы построить D-оптимальный план экспериментов, надо определить оптимальное значение m^0. Вычисления показывают, что определитель информационной матрицы плана ε_n равен

$$|M(\varepsilon_n)| = \frac{4(n-m)m}{d_1 d_2}.$$

Максимального значения этот определитель достигает в точке $m^0 = 2^{-1}n$. Если $n = 2s$, то $m^0 = s$. Если $n = 2s + 1$, то $m^0 = s + 1$ либо $m^0 = s$. Теорема 1 доказана.

Обозначим через K множество точек из интервала $[-1, 1]$, в которых неравенство (2) обращается в равенство. Если множество K, кроме точек –1, 1, содержит и другие точки из интервала $[-1, 1]$, то в этом случае точный D-оптимальный план экспериментов для нечетного числа наблюдений $n = 2s + 1$ может отличаться от плана для равноточных наблюдений.

Теорема 2. *Если множество* $K \setminus \{-1, 1\}$ *не пусто, то для нечетного числа наблюдений* $n = 2s + 1$ *D-оптимальный план имеет вид*

$$\varepsilon^0_{2s+1} = \begin{Bmatrix} -1, & 1, & x \\ s, & s, & 1 \end{Bmatrix}, \quad x \in K. \tag{9}$$

Доказательство. Если x принимает значения –1 либо 1, то план (9) совпадает с одним из D-оптимальных планов для равноточных наблюдений. Если $x \in K \setminus \{-1, 1\}$, то определитель информационной матрицы плана (9) равен $4s(s+1)(d_1 d_2)^{-1}$, т. е. совпадает с определителем D-оптимального плана для равноточных наблюдений. Это означает, что план (9) D-оптимален. Теорема 2 доказана.

Теорема 3. *Для модели неравноточных наблюдений* (1) *с дисперсией наблюдений* $d(x)$, *удовлетворяющей* (2), *наилучшие линейные несмещенные оценки неизвестных параметров, построенные по точным D-оптимальным планам, не зависят от дисперсий наблюдений* $d(x)$ *и совпадают с оценками для равноточных наблюдений.*

Доказательство. Для D-оптимальных планов введем следующие обозначения: t – число наблюдений в точке –1; $n - t$ – в точке 1. Для четных $n = 2s$, $t = s$, а для нечетных $n = 2s + 1$ значение t равно s либо $s + 1$. Обозначим через y_{1i} наблюдения в точке –1, а через y_{2i} – наблюдения в точке 1.

Непосредственные вычисления показывают, что для D-оптимальных планов согласно [1] оценки неизвестных параметров имеют вид

$$\tilde{\theta}_0 = \frac{1}{2t(n-t)}\left((n-t)\sum_{i=1}^{t} y_{1i} + t\sum_{i=1}^{n-t} y_{2i}\right),\ \tilde{\theta}_1 = \frac{1}{2t(n-t)}\left(t\sum_{i=1}^{n-t} y_{2i} - (n-t)\sum_{i=1}^{t} y_{1i}\right). \tag{10}$$

Оценки (10) не зависят от дисперсий d_1, d_2 и совпадают с оценками неизвестных параметров для случая, если бы наблюдения были равноточными. Теорема 3 доказана.

Перейдем теперь к рассмотрению ситуации, когда дисперсия наблюдений $d(x)$ может принимать нулевое значение на одном из концов интервала $[-1, 1]$. Для определенности будем считать, что дисперсия обращается в нуль на левом конце интервала. Пусть $d_1 = 0$, $d_2 > 0$. Класс функций, описывающих изменение дисперсий наблюдений, определим аналогично (2) неравенством

$$d(x) \ge \frac{d_2}{4}(x+1)^2,\ \ x \in [-1, 1]. \tag{11}$$

Очевидно, при $x = \pm 1$ неравенство (11) обращается в равенство. Пусть K_1 – множество точек из интервала $(-1, 1]$, в которых неравенство (11) обращается в равенство.

Теорема 4. *Для модели неравноточных наблюдений* (1) *с дисперсией наблюдений* $d(x)$*, удовлетворяющей* (11)*, с вероятностью 1, точный D-оптимальный план экспериментов имеет вид*

$$\varepsilon_n^0 = \begin{Bmatrix} -1, & x_i, & i = \overline{2, n} \\ 1, & 1, & \end{Bmatrix},\ \ x_i \in K_1. \tag{12}$$

Оценки неизвестных параметров, построенные по плану (12), следующие:

$$\tilde{\theta}_1 = \left(\sum_{i=2}^{n} \frac{(1+x_i)^2}{d(x_i)}\right)^{-1} \sum_{i=2}^{n} \frac{\bar{y}_i(1+x_i)}{d(x_i)},\ \tilde{\theta}_0 = \tilde{\theta}_1 + y_{(-1)}, \tag{13}$$

где $\bar{y}_i = y_i - y_{(-1)}$*;* y_i *– результаты наблюдений в точках* x_i*;* $y_{(-1)}$ *– результат наблюдения в точке –1. Дисперсии оценок* (13) *равны*

$$D\{\tilde{\theta}_1\} = D\{\tilde{\theta}_0\} = \frac{d_2}{4(n-1)}. \tag{14}$$

Доказательство. Поскольку в точке –1 ошибка наблюдений с вероятностью 1 равна нулю, то проведем одно наблюдение в этой точке: $y_{(-1)} = \theta_0 - \theta_1$. Это равенство, которое выполняется с вероятностью 1, означает, что неизвестные параметры линейно связаны. Используя этот факт, получаем

$$\bar{y}_i = \theta_1(1+x_i) + \varepsilon(x_i),\ \ i = \overline{2, n}. \tag{15}$$

Дисперсия значений $\bar{y}_i$ равна $d(x_i)$. По теореме Гаусса – Маркова [1] наилучшая линейная несмещенная оценка $\tilde{\theta}_1$ для параметра θ_1 имеет вид

$$\tilde{\theta}_1 = \left(\sum_{i=2}^{n} \frac{(1+x_i)^2}{d(x_i)}\right)^{-1} \sum_{i=2}^{n} \frac{\bar{y}_i(1+x_i)}{d(x_i)}. \tag{16}$$

Дисперсия оценки (16) равна

$$D\{\tilde{\theta}_1) = \left(\sum_{i=2}^{n} \frac{(1+x_i)^2}{d(x_i)}\right)^{-1}. \tag{17}$$

Для дисперсий наблюдений $d(x)$, удовлетворяющих (11), минимальное значение дисперсии (17) достигается, когда $x_i \in K_1, i = \overline{2, n}$. Действительно, в силу (11)

$$\frac{\left(1+x_i\right)^2}{d\left(x_i\right)} \leq \frac{4}{d_2}, \; i=\overline{2, n}. \tag{18}$$

При $x_i \in K_1, i=\overline{2, n},$ неравенства (18) обращаются в равенства. В силу (18)

$$D\left\{\tilde{\theta}_1\right\} \geq \frac{d_2}{4(n-1)}, \; i=\overline{2, n}. \tag{19}$$

Нижняя граница в (19) достигается, когда $x_i \in K_1, i=\overline{2, n}.$ Утверждение (14) доказано, и с вероятностью 1 точный D-оптимальный план экспериментов имеет вид (12). При $x_i \in K_1, i=\overline{2, n},$ оценка (16) обращается в оценку (13) для параметра θ_1. Доказательство теоремы завершено.

Следствие. *Если множество* $K_1 \setminus \{1\}$ *пусто, то в условиях теоремы 4 D-оптимальный план экспериментов принимает вид*

$$\varepsilon_n^0=\left\{\begin{matrix} -1, & 1 \\ 1, & n-1 \end{matrix}\right\}. \tag{20}$$

Оценки неизвестных параметров, построенные по плану (20)*, не зависят от значений дисперсий и равны*

$$\tilde{\theta}_1=\frac{1}{2}\left(\frac{1}{n-1} \sum_{i=2}^n y_i-y_{(-1)}\right), \; \tilde{\theta}_0=\tilde{\theta}_1+y_{(-1)},$$

где y_i – значения наблюдений в точке 1.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Федоров В. В.* Теория оптимального эксперимента. М., 1971.
2. *Кирлица В. П. D*-оптимальные планы экспериментов, робастные относительно изменения дисперсии наблюдений, для линии регрессии // Вестн. БГУ. Сер. 1, Физика. Математика. Информатика. 2008. № 3. С. 89–92.

Поступила в редакцию 03.02.2015.

Валерий Петрович Кирлица – кандидат физико-математических наук, доцент кафедры математического моделирования и анализа данных факультета прикладной математики и информатики БГУ.

УДК 004.021

А. А. КОЗИК, А. П. ПОБЕГАЙЛО

УСКОРЕНИЕ ПАРАЛЛЕЛЬНОГО ПОИСКА В ШИРИНУ В ГРАФАХ ПРИ ВЫЧИСЛЕНИЯХ НА GPU

Решается задача ускорения работы алгоритма поиска в ширину на графах при вычислениях с помощью графических процессоров. Ускорение алгоритма достигается посредством кодирования матрицы смежности графа специальным образом, учитывающим программно-аппаратную архитектуру CUDA, что дает возможность значительно улучшить производительность рассматриваемого алгоритма при исполнении на этой платформе. Представленный способ кодирования матрицы смежности позволяет снизить накладные расходы на операции чтения/записи из глобальной памяти, необходимый объем памяти для представления графа, а также существенно уменьшить время, затрачиваемое на загрузку данных, которое в других случаях может превышать время выполнения самого алгоритма.

Предложенная реализация параллельного алгоритма поиска в ширину может быть использована при решении задач, включающих поиск кратчайшего пути в графах, для случайных графов с высокой вероятностью появления ребра. Несмотря на избыточность вычислений, при высоких значениях вероятности *p* – появления ребра графа – полезная нагрузка на GPU остается на уровне 96,7 %.

Ключевые слова: алгоритмы на графах; параллельный поиск в ширину; кодирование матрицы смежности; GPU; CUDA.

The article solves the problem of accelerating parallel breadth first search algorithm on GPU. Acceleration of the algorithm is obtained by special coding of graph adjacency matrix, taking into account CUDA architecture. This coding improves the algorithm efficiency on CUDA platform because reduces the number of read/write operations from the global memory and the time needed to load data. The proposed implementation of the algorithm can be used for solving problems on graph, which use breadth first search as basis computation and have not very sparse adjacency matrix. Despite the redundancy calculation, if the probability of graph edges occurrences is high then the payload on GPU remains at the level 96,7 %.

Key words: algorithms on graphs; parallel breadth first search; coding adjacency matrix; GPU; CUDA.