УДК 519.85

П.И. СТЕЦЮК

УСКОРЕНИЕ СУБГРАДИЕНТНОГО МЕТОДА ПОЛЯКА

Исследуются свойства субградиентного метода Поляка для нахождения точки минимума выпуклой функции. Показано, что для овражных функций сходимость метода можно ускорить за счет линейного преобразования пространства переменных. Изложен субградиентный метод Поляка с преобразованием пространства в случае тупого угла между двумя последовательными субградиентами, что существенно сокращает количество итераций для гладких и негладких овражных функций.

Введение. В этом году исполнилось бы 75 лет украинскому ученому, академику НАН Украины Н.З. Шору. На протяжении многих лет статьи Н.З.Шора и его учеников публиковались в сборнике научных трудов "Теория оптимальных решений". Первая статья "Алгебраический подход к исследованию задачи о четырех красках" опубликована в соавторстве с Г.А.Донцом в 1967 г. Последняя статья "Двойственные оценки для специальной оптимизационной задачи квадратичного типа на многообразии Штиффеля" (соавторы – П.И.Стецюк и О.А.Березовский) опубликована в 2004 г. Названия этих статей лишь подчеркивают центральный результат Н.З.Шора за последние 20 лет – двойственный метод получения и уточнения оценок целевой функции в невыпуклых квадратичных моделях. Этот метод нашел красивое воплощение в NP-трудной задаче нахождения взвешенного максимального независимого множества вершин графа, где двойственные оценки Шора равны известным числам Ловаса. Эта задача используется в теории информации и кодировании, проектировании различных устройств при определенных условиях несовместности; она тесно связана с известными задачами выбора, разбиения множеств, раскраски графов и другими комбинаторными задачами, имеющими, в свою очередь, массу приложений.

Большое количество статей в "Теории оптимальных решений" посвящено исследованию модификаций *r*-алгоритмов и их применений в сложных экстремальных задачах. Данная статья связана с использованием линейных неортогональных преобразований пространства для улучшения обусловленности овражных функций – идеи Н.З.Шора, на которой построены эффективные модификации *r*-алгоритмов. В ней рассмотрены субградиентные методы с преобразованием пространства, основанные на монотонном уменьшении расстояния к точке минимума. Сделано это на примере известного субградиентного метода, предложенного Б.Т.Поляком в 1969 г. для нахождения точки минимума выпуклой функции при известном оптимальном значении функции.

1. Субградиентный метод Поляка. Пусть $f(x)$ – выпуклая функция векторного аргумента $x \in R^n$ и ее субградиент $\partial f(x)$ удовлетворяет условию:

$$(x - x^*, \partial f(x)) \ge m(f(x) - f^*), \quad \forall x \in R^n, \forall x^* \in X^*, \quad m \ge 1. \qquad (1)$$

Здесь R^n – евклидово пространство размерности n со скалярным произведением (x, y); X^* – множество точек минимума функции $f(x)$; f^* – минимальное значение функции $f(x)$: $f^* = f(x^*)$, $x^* \in X^*$. Параметр m введен для учета специальных классов выпуклых функций: например, для кусочно-линейной негладкой функции $m = 1$, для квадратичной гладкой функции $m = 2$.

Пусть известно f^*. Для нахождения точки $x^* \in X^*$ применяется субградиентный метод Поляка [1]:

$$x_{k+1} = x_k - h_k \frac{\partial f(x_k)}{\|\partial f(x_k)\|}, \quad h_k = \frac{m\left(f(x_k) - f^*\right)}{\|\partial f(x_k)\|}, \quad k = 0, 1, 2, \ldots. \qquad (2)$$

Величину h_k называют шагом Поляка или шагом Агмона-Моцкина-Шенберга. Этот шаг тесно связан с результатами И. И. Еремина о сходимости итерационных методов аппроксимации неподвижных точек с помощью операторов, обладающих свойством квазисжимаемости (фейеровости) [2].

Шаг Поляка обладает интересными свойствами, которые ему обеспечивает априорное знание минимального значения функции и связанное с ним неравенство (1).

Теорема 1. Для всех точек, генерируемых методом (2), справедливы неравенства

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - \left(\frac{m(f(x_k) - f^*)}{\|\partial f(x_k)\|}\right)^2, \quad \forall x^* \in X^*, k = 0, 1, \ldots, \qquad (3)$$

и неравенства

$$(x^* - x_{k+1}, -\partial f(x_k)) \ge 0, \quad \forall x^* \in X^*, k = 0, 1, \ldots, \qquad (4)$$

Доказательство. Для любого $x^* \in X^*$ и произвольного k ($k \ge 0$) имеем

$$\|x_{k+1} - x^*\|^2 = \left\| x_k - x^* - h_k \frac{\partial f(x_k)}{\|\partial f(x_k)\|} \right\|^2 = \|x_k - x^*\|^2 - 2h_k \frac{(x_k - x^*, \partial f(x_k))}{\|\partial f(x_k)\|} + h_k^2.$$

Учитывая, что из (1) следует неравенство

$$\frac{(x_k - x^*, \partial f(x_k))}{\|\partial f(x_k)\|} \ge \frac{m(f(x_k) - f^*)}{\|\partial f(x_k)\|} = h_k,$$

имеем

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\| - h_k^2 = \|x_k - x^*\|^2 - \left(\frac{m(f(x_k) - f^*)}{\|\partial f(x_k)\|} \right)^2,$$

что дает неравенства (3).

Неравенства (4) следуют из того, что на основании (1) имеем

$$(x^* - x_{k+1}, -\partial f(x_k)) = (x_{k+1} - x^*, \partial f(x_k)) = (x_k - x^* - h_k \frac{\partial f(x_k)}{\|\partial f(x_k)\|}, \partial f(x_k)) =$$

$$= (x_k - x^*, \partial f(x_k)) - h_k \|\partial f(x_k)\| = (x_k - x^*, \partial f(x_k)) - m(f(x_k) - f^*) \ge 0.$$

Теорема 1 доказана.

В теореме отражены два центральных свойства шага Поляка. Первое свойство следует из неравенств (4). Оно означает, что шаг h_k определяет величину максимального сдвига в направлении нормированного антисубградиента, при котором угол между антисубградиентом и направлением из точки x_{k+1} на произвольную точку из множества минимумов будет нетупым. Второе свойство следует из неравенств (3) и означает, что если множество X^* состоит из единственной точки x^*, то шаг h_k выбирается таким, чтобы расстояние от точки x_{k+1} к точке минимума x^* было минимальным.

2. Проблема овражности. Медленную скорость сходимости метода Поляка проиллюстрируем на примере трех выпуклых функций от двух переменных, первые две функции – негладкие (овражная и существенно овражная), а третья – квадратичная овражная функция. В табл. 1 приведено количество итераций метода (2) для нахождения последовательно уточняемых приближений к единственной точке минимума $x^* = (0,0)$. Эти приближения заданы десятью последовательно уменьшающимися (на порядок) значениями epsf = ε_f (первая колонка в табл. 1). Метод Поляка прекращал работу или на итерации $k = itn$, для которой $f(x_{itn}) - f^* \le \varepsilon_f$, или если превышено максимальное количество итераций, равное 20000.

Колонки itn1, itn2 и itn3 отвечают овражной кусочно-линейной функции $f_1(x_1,x_2)=|x_1|+t|x_2|$ для значений t равных 3, 9 и 27 соответственно. При расчетах использовалась стартовая точка $x_0=(1,1)$ и параметр $m=1$. Из табл. 1 видим, что количество итераций существенно увеличивается с ростом t. Для того чтобы найти точку, где значение функции $f_1(x_1,x_2)=|x_1|+27|x_2|$ отличается от минимального $f_1^*=0$ не более, чем на $\varepsilon_f=10^{-10}$, методу Поляка потребовалось 8634 итерации. Такую же точность для функции $f_1(x_1,x_2)=|x_1|+t|x_2|$ при $t=100$ получить нереально даже при ста тысячах итераций метода Поляка.

Для $f_2(x_1,x_2)=\max\left\{x_1^2+(2x_2-2)^2-3, x_1^2+(x_2+1)^2\right\}$ – существенно овражной кусочно-квадратичной функции зигзаго-образная траектория последовательных приближений метода Поляка дана на рис. 1. Здесь метод Поляка находит приближение, где значение функции $f_2(x_1,x_2)$ отличается от минимального $f_2^*=1$ не более, чем на 0.0001, только за 16004 итераций (см. колонку itn4 в табл. 1) Медленная сходимость проявляется со второй итерации, за одну итерацию метод спускается на дно оврага, а далее метод зигзагообразно движется вдоль русла оврага.

ТАБЛИЦА 1. Сходимость метода Поляка для двумерных овражных функций

epsf	Itn1	itn2	itn3	itn4	itn5	Itn6	itn7
1.0e-01	14	119	1080	16	6	6	6
1.0e-02	24	212	1919	162	10	10	10
1.0e-03	34	305	2759	1604	12	12	12
1.0e-04	45	398	3598	16004	16	16	16
1.0e-05	55	492	4437	20000	20	20	20
1.0e-06	65	585	5277	20000	22(42)	22(42)	22(42)
1.0e-07	76	678	6116	20000	26(48)	26(52)	26(52)
1.0e-08	86	771	6955	20000	28(54)	30(56)	30(56)
1.0e-09	96	865	7795	20000	32(62)	32(62)	32(62)
1.0e-10	107	958	8634	20000	36(68)	36(70)	36(70)

Для функций $f_1(x_1,x_2)$ и $f_2(x_1,x_2)$ использовался параметр $m=1$, который можно применять для произвольной выпуклой функции. Для квадратичной функции $f_3(x_1,x_2)=x_1^2+tx_2^2, t>0$ использовался параметр $m=2$. Он в два раза увеличивает длину шага Поляка, соответствующего параметру $m=1$, и обеспечивает более быструю скорость сходимости с различных начальных приближений. Количество итераций метода Поляка со стартовой точки $x_0 = (0,0)$ приведено в колонках itn5, itn6 и itn7 и соответствует значениям t равным 100, 10000 и 1000000. Количество итераций для всех 10 значений epsf одинаково и не зависит от степени вытянутости поверхности квадратичной функции. Отличия наблюдаются лишь при достаточно малых $\varepsilon_f = 10^{-12}, 10^{-14}, 10^{-16}, 10^{-18}, 10^{-20}$, для которых количество итераций приведено в скобках после количеств итераций для шести последних значений epsf. Чтобы найти точку минимума квадратичной функции с такой же точностью, как и точку минимума функции $f_1(x_1, x_2)$, нужно взамен ε_f использовать ε_f^2.

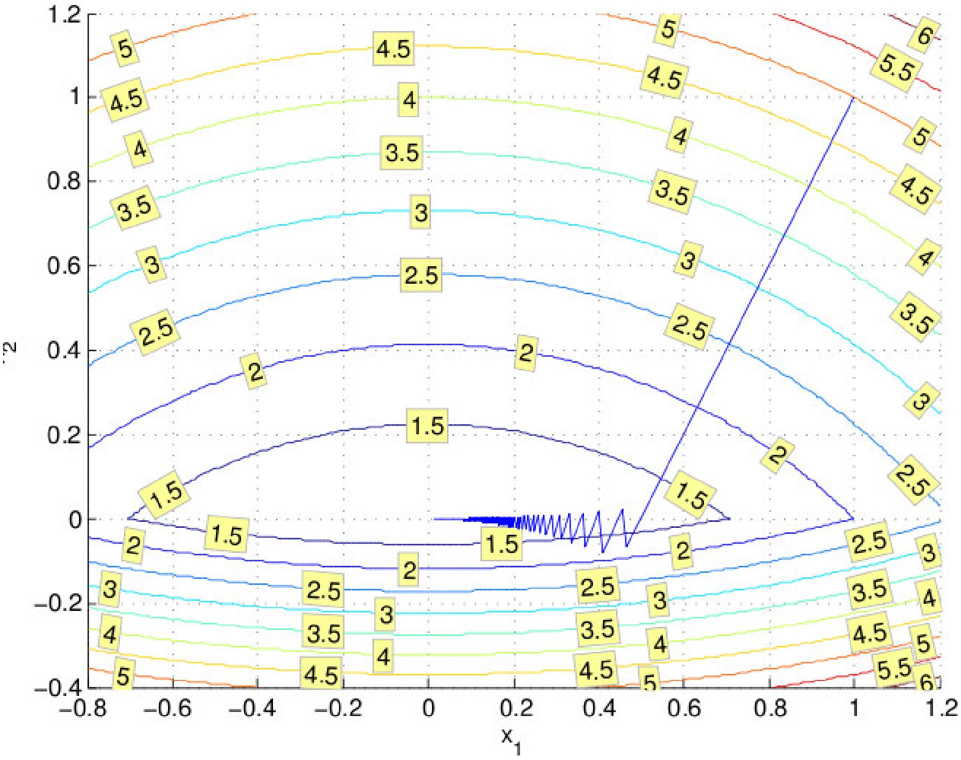

РИСУНОК 1. Траектория метода Поляка для $f_2(x_1, x_2)$ ($x_0 = (1,1)$)

3. Метод Поляка с преобразованием пространства. Пусть произведена замена переменных $x = By$, где B – неособенная $n \times n$-матрица (т.е. существует обратная матрица $A = B^{-1}$). Субградиент выпуклой функции $f(x)$ в точке x_k удовлетворяет неравенству

$$f(x) \geq f(x_k) + (\partial f(x_k), x - x_k) \quad \forall x \in R^n,$$

откуда, осуществляя замену переменных $x = By$, получаем

$$\varphi(y) \geq \varphi(y_k) + (B^T \partial f(x_k), y - y_k) \quad \forall y \in R^n.$$

Вектор $\partial \varphi(y_k) = B^T \partial f(x_k)$ удовлетворяет неравенству

$$\varphi(y) \geq \varphi(y_k) + (\partial \varphi(y_k), y - y_k) \quad \forall y \in R^n$$

и является субградиентом выпуклой функции $\varphi(y) = f(By)$ в точке $y_k = Ax_k$ преобразованного пространства переменных $y = Ax$ [3].

Для нахождения точки $x^* \in X^*$ субградиентный метод Поляка с преобразованием пространства (определяется невырожденной матрицей B) имеет следующий вид:

$$x_{k+1} = x_k - h_k \frac{B^T \partial f(x_k)}{\left\| B^T \partial f(x_k) \right\|}, \quad h_k = \frac{m\left(f(x_k) - f^*\right)}{\left\| B^T \partial f(x_k) \right\|}, \quad k = 0, 1, 2, \ldots. \quad (5)$$

Здесь величина h_k – шаг Поляка (шаг Агмона-Моцкина-Шенберга), но в преобразованном пространстве переменных $y = Ax$. Это следует из того, что в преобразованном пространстве переменных метод (5) записывается как субградиентный процесс

$$y_{k+1} = y_k - h_k \frac{\partial \varphi(y_k)}{\| \partial \varphi(y_k) \|}, \quad h_k = \frac{m(\varphi(y_k) - \varphi^*)}{\| \partial \varphi(y_k) \|}, \quad k = 0, 1, 2, \ldots. \quad (6)$$

Шаг Поляка в преобразованном пространстве переменных обладает такими же свойствами как и шаг Поляка в исходном пространстве. Их обеспечивает ему априорное знание минимального значения функции и связанное с ним неравенство (1).

Теорема 2. Для всех точек, генерируемых методом (5), справедливы неравенства

$$\| A(x_{k+1} - x^*) \|^2 \leq \| A(x_k - x^*) \|^2 - \left(\frac{m(f(x_k) - f^*)}{\| B^T \partial f(x_k) \|} \right)^2, \forall x^* \in X^*, k = 0, 1, \ldots, \quad (7)$$

и неравенства

$$(A(x^* - x_{k+1}), -B^T \partial f(x_k)) \geq 0, \quad \forall x^* \in X^*, k = 0, 1, \ldots. \quad (8)$$

Доказательство теоремы 2 аналогично доказательству теоремы 1.

Очевидно, что если матрицу B выбрать такой, чтобы овражная функция в преобразованном пространстве переменных становилась менее овражной, то субградиентный метод Поляка с преобразованием пространства (5) окажется эффективнее, чем этот метод без преобразования пространства (2). Это подтверждает количество итераций метода (5) для нахождения десяти последовательно уточняемых приближений функции $f_1(x_1, x_2) = |x_1| + 10|x_2|$

для различных матриц B, которые получены в результате растяжения пространства переменных в направлении x_2 с коэффициентами растяжения $\alpha = 1;\ 1.5;\ 2;\ 3;\ 4;\ 5$. Из табл. 2 легко видеть, что количество итераций метода Поляка с преобразованием пространства увеличивается по мере того, как уменьшается степень овражности функции $\varphi_1(y_1, y_2) = |\, y_1 \,| + \dfrac{10}{\alpha} |\, y_2 \,|$ в преобразованном пространстве переменных.

ТАБЛИЦА 2. Сходимость метода Поляка с преобразованием пространства

epsf	itn1	itn2	itn3	itn4	itn5	itn6
1.0e-01	147	63	33	6	10	9
1.0e-02	262	114	62	19	17	13
1.0e-03	377	165	91	31	24	18
1.0e-04	492	216	119	44	31	22
1.0e-05	607	268	148	57	38	27
1.0e-06	722	319	177	70	45	31
1.0e-07	837	370	206	82	53	36
1.0e-08	952	421	234	95	60	40
1.0e-09	1000	472	263	108	67	45
1.0e-10	1000	523	292	121	74	49

4. Ускоренный субградиентный метод Поляка. Медленную сходимость метода Поляка для овражных функций определяет угол между двумя последовательными субградиентами: $\partial f(x_k)$ и $\partial f(x_{k+1})$. Чем ближе этот угол к 180 градусам, тем более медленной будет скорость сходимости. Тупой угол между векторами $\partial f(x_k)$ и $\partial f(x_{k+1})$ можно преобразовать в прямой с помощью "одноранговового эллипсоидального оператора" [4]. Если на каждой итерации избавляться от тупого угла между последовательными субградиентами в преобразованном пространстве переменных, то для овражных функций повысится скорость сходимости метода. Этот принцип реализован в изложенной ниже модификации метода Поляка.

Ускоренный субградиентный метод Поляка имеет следующий вид:

$$x_{k+1} = x_k - h_k B_k \frac{B_k^T \partial f(x_k)}{\left\| B_k^T \partial f(x_k) \right\|}, \quad h_k = \frac{m(f(x_k) - f^*)}{\left\| B_k^T \partial f(x_k) \right\|}, \quad k = 0, 1, 2, \ldots, \qquad (9)$$

где матрица $B_0 = I$, а матрица B_{k+1} размерности $n \times n$ вычисляется по правилу:

$$B_{k+1} = \begin{cases} B_k, & \text{если } \mu_k \geq 0 \\ B_{k+1} = B_k + (B_k\eta)\,\xi_{k+1}^T, & \text{иначе} \end{cases} \tag{10}$$

где

$$\mu_k = (\xi_k, \xi_{k+1}), \quad \xi_k = \frac{B_k^T \partial f(x_k)}{\left\| B_k^T \partial f(x_k) \right\|}, \quad \xi_{k+1} = \frac{B_k^T \partial f(x_{k+1})}{\left\| B_k^T \partial f(x_{k+1}) \right\|},$$

$$\eta = \left(\frac{1}{\sqrt{1-\mu_k^2}} - 1 \right) \xi_{k+1} - \frac{\mu_k}{\sqrt{1-\mu_k^2}} \xi_k .$$

Метод (9), (10) можно назвать ускоренным методом Поляка за счет антиовражного приема, подобного тому, который использован в r-алгоритмах Шора [3]. Действительно, на k-й итерации растяжение пространства производится в направлении разности нормированных последовательных субградиентов в преобразованном пространстве переменных: $y = A_k x = B_k^{-1} x$, где B_k – невырожденная матрица размерности $n \times n$. Если нормы субградиентов одинаковы, то это направление совпадает с разностью двух последовательных субградиентов, по которой реализуется растяжение пространства в r-алгоритмах. Отличие состоит в том, что в преобразованном пространстве переменных для r-алгоритмов второй субградиент определяется согласно шагу наискорейшего спуска в направлении антисубградиента, а в методе (9), (10) – согласно шагу Поляка.

Теорема 3. *Пусть* $A_k = B_k^{-1}$, $A_{k+1} = B_{k+1}^{-1}$. *Для всех точек, генерируемых методом (9), (10), справедливы неравенства*

$$\| A_{k+1}(x_{k+1} - x^*) \|^2 \leq \| A_k (x_k - x^*) \|^2 - \left(\frac{m(f(x_k) - f^*)}{\| B_k^T \partial f(x_k) \|} \right)^2, \forall x^* \in X^*, k = 0,1,\ldots \tag{11}$$

Доказательство теоремы 3 аналогично доказательству теоремы 2 [4].

Преобразование пространства в методе (9), (10) направлено на уменьшение степени вытянутости поверхностей уровня овражных выпуклых функций. Неравенства (11) означают, что в каждом очередном преобразованном пространстве переменных гарантируется уменьшение расстояния до множества точек минимума. В связи с этим для каждой итерации $k > 1$ имеет место неравенство

$$\| A_k (x_k - x^*) \|^2 \leq \| x_0 - x^* \|^2 - \sum_{i=0}^{k-1} \left(\frac{m(f(x_i) - f^*)}{\| B_i^T \partial f(x_i) \|} \right)^2 .$$

Для овражных функций детерминант матрицы B_k стремится к нулю, так как, если на k-м шаге реализуется преобразование пространства, то

$$\det(B_{k+1}) = \det(B_k)\det(I + \eta\xi_{k+1}^T) = \det(B_k)\sqrt{1-\mu_k^2} = \det(B_k)\sqrt{1-\cos^2\varphi_k},$$

где φ_k – тупой угол между двумя последовательными субградиентами.

Для овражных негладких функций метод (9), (10) обладает ускоренной сходимостью по отношению к методу (2). Так, например, для кусочно-линейной функции двух переменных $f_1(x_1,x_2) = |x_1| + t|x_2|$ при любом значении параметра $t > 1$ и произвольной стартовой точке x_0 методом (9), (10) можно найти точку минимума $x^* = (0,0)$ не более, чем за две итерации. Метод (2) сходится к x^* со скоростью геометрической прогрессии со знаменателем $\sqrt{1-1/t^2}$ и требует существенного количества итераций даже при сравнительно небольших значениях t. Траектория ускоренного метода Поляка для кусочно-квадратичной функции $f_2(x_1,x_2)$ дана на рис. 2, здесь метод находит точку, где $f_2(x) < 1+10^{-5}$ всего за 16 итераций, а где $f_2(x) < 1+10^{-10}$ – за 31 итерацию.

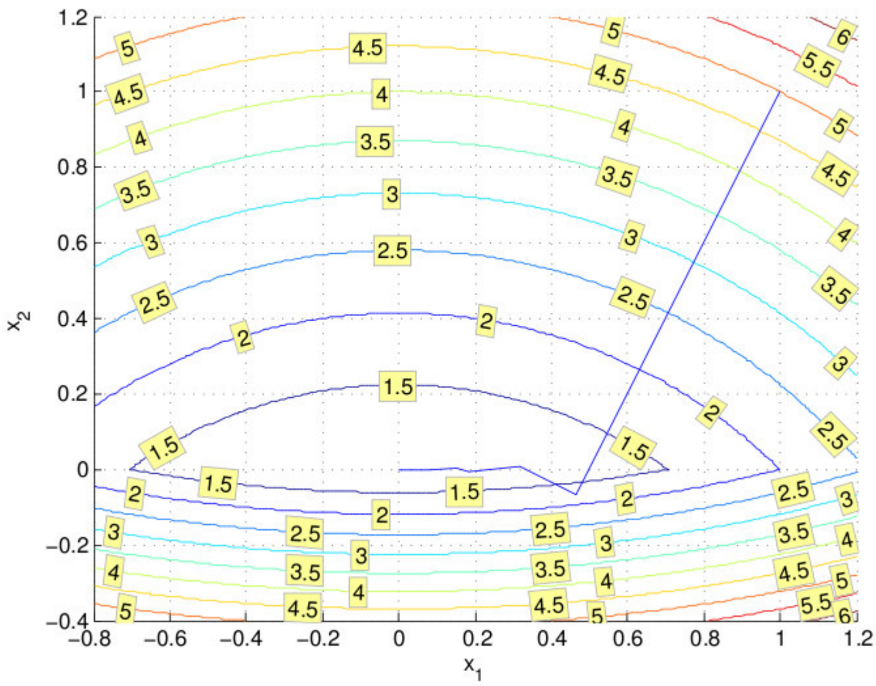

РИСУНОК 2. Траектория ускоренного метода Поляка для $f_2(x_1,x_2)$ ($x_0 = (1,1)$)

Заключение. Несмотря на то, что r-алгоритмы используются уже 40 лет, проблема обоснования их сходимости для всего класса выпуклых функций остается открытой и в настоящее время. Еще в 1982 г. Н.З.Шор и В.И.Гершович в работе [5] отметили: "Теория всего класса алгоритмов с растяжением пространства далека от совершенства. Нам кажется достаточно реалистичной целью – построение такого алгоритма, который по своей практической эффективности не уступал бы r-алгоритму и был столь же хорошо обоснован, как метод эллипсоидов". Шагом в этом направлении можно считать алгоритм (9), (10), где для преобразования специального эллипсоида в шар используется антиовражный прием, близкий к тому, который имеет место в r-алгоритмах. Однако, здесь растяжение пространства реализуется в направлении разности двух нормированных субградиентов, и близким к направлению разности двух субградиентов оно будет только тогда, когда нормы субградиентов близки.

П.І. Стецюк

ПРИСКОРЕННЯ СУБГРАДІЄНТНОГО МЕТОДУ ПОЛЯКА

Досліджуються властивості субградієнтного методу Поляка для знаходження точки мінімуму опуклої функції. Показано, що для яружних функцій збіжність методу можна прискорити за рахунок лінійного перетворення простору змінних. Розглянуто субградієнтний метод Поляка з перетворенням простору у випадку тупого кута між двома послідовними субградієнтами, що істотно скорочує кількість ітерацій для гладких і негладких яружних функцій.

P.I. Stetsyuk

ACCELERATION OF POLYAK’S SUBGRADIENT METHOD

The properties of Polyak’s subgradient method for finding the minimum point of a convex function is investigated. It is shown that for ravine functions the convergence of the method can be accelerated by a linear transformation of the space of variables. Polyak’s subgradient method with the transformation of the space in the case of the obtuse angle between two successive subgradients is considered. It significantly reduces the number of iterations for smooth and nonsmooth ravine functions.

1. *Поляк Б.Т.* Введение в оптимизацию. – М.: Наука, 1983. – 384 с.
2. *Васин В.В., Еремин И.И.* Операторы и итерационные процессы фейеровского типа (теория и приложения). – Москва; Ижевск: Институт компьютерных исследований, НИЦ «Регулярная и хаотическая динамика», 2005. – 200 с.
3. *Шор Н.З.* Методы минимизации недифференцируемых функций и их приложения. – Киев: Наук. думка, 1979. – 199 с.
4. *Стецюк П.И.* Ортогонализующие линейные операторы в выпуклом программировании // Кибернетика и системный анализ. – 1997. – № 3. – С. 97–119.
5. *Шор Н.З., Гершович В.И.* Метод эллипсоидов, его обобщения и приложения // Кибернетика. – 1982. – № 5. – С. 61–69.

Получено 21.05.2012