УДК 512.25/26

УСТОЙЧИВЫЙ АЛГОРИТМ МЕТОДА СОПРЯЖЕННЫХ НАПРАВЛЕНИЙ ДЛЯ КВАДРАТИЧНОГО ПРОГРАММИРОВАНИЯ

В.И.Шмырёв

Изложенный в работе [I] алгоритм метода сопряженных направлений для решения задачи квадратичного программирования предполагает точное выполнение условий сопряженности некоторой системы векторов. Однако при практической реализации алгоритма вследствие ошибок округления эти условия будут систематически нарушаться, что повлечет за собой накопление погрешности решения. В настоящей работе рассмотрена модификация процесса, предполагающая лишь приближенное выполнение условий сопряженности. Эта модификация основана на идее специальной организации процесса субоптимизации, примененной в алгоритмах [2 , 3]. Попутно показывается применимость этой идеи к более широкому классу алгоритмов.

I. Постановка вопроса

Алгоритм [I] относится к классу алгоритмов субоптимизации (см. [4,5]), реализующих процесс направленного перебора аффинных многообразий, определяемых ограничениями задачи, проходящими через текущую точку процесса (активными ограничениями). Для полноты изложения остановимся кратко на общей схеме этого процесса.

Будем рассматривать задачу в следующей формулировке:

$$f(x) \longrightarrow \min!$$

$$(a_i, x) \geq b_i, \quad i = 1, \ldots, m, \qquad (1)$$

где x_i, $a_i \in R^n$. Предполагается, что функция f выпукла и непрерывно дифференцируема, а для ограничений (1) выполняется <u>условие невырожденности</u>:

для любого вектора x, удовлетворяющего системе (1), векторы a_i, при $i \in I(x) = \{i \mid (a_i, x) = b_i\}$ линейно-независимы.

Процесс субоптимизации для такой задачи будет состоять в следующем. К началу k-го шага процесса имеется некоторый вектор x^k, удовлетворяющий ограничениям (1), и множество $I_k \subset I(x^k)$. Рассматривается подзадача минимизации функции f на аффинном многообразии

$$L(I_k) = \{x \in R^n \mid (a_i, x) = b_i, \; i \in I_k\}.$$

Для простоты изложения будем считать, что при любом I_k такая задача разрешима. Пусть $\bar{x}^k$ - решение рассматриваемой подзадачи. Пытаемся осуществить перемещение из точки x^k в точку $\bar{x}^k$, не нарушая ограничений (1). Если такое перемещение осуществить нельзя, то номер i ограничения, препятствующего перемещению, присоединяется к множеству I_k - множество I_k расширяется, что ведет к сужению многообразия $L(I_k)$. Так продолжаем до тех пор, пока не удастся продвинуться в точку минимума на рассматриваемом аффинном многообразии. Для полученной точки $x^s \in L(I_s)$ определяются множители Лагранжа λ_i из условия

$$\sum_{i \in I_s} \lambda_i a_i = \nabla f(x^s). \qquad (3)$$

Если все λ_i неотрицательны, точка x^s решает исходную задачу. Если же для некоторого i_0 оказалось $\lambda_{i_0} < 0$, то i_0 исключается из множества I_s, что ведет к расширению многообразия $L(I_s)$. На новом многообразии мы повторяем описанную процедуру с самого начала.

Таким образом, весь процесс можно рассматривать состоящим из макрошагов, каждый из которых в свою очередь состоит из

расширения имеющегося многообразия и, возможно, последующих сужений.

В работе [1] для случая выпуклой квадратичной функции $f(x) = (p,x) + \frac{1}{2}(x, Cx)$ решение подзадач минимизации f на аффинных многообразиях осуществляется на основе схемы сопряженных направлений. На каждом шаге процесса для рассматриваемого аффинного многообразия имеется базис, состоящий из сопряженных векторов, который при расширении многообразия достраивается выполнением одного шага процесса C -ортогонализации, а при сужении корректируется посредством ортогонального преобразования.

Теоретически такая схема обладает тем свойством, что можно получить точку минимума на очередном аффинном многообразии путем минимизации функции f лишь по одному направлению. Однако вследствие ошибок округления и неустойчивости процесса C -ортогонализации сопряженность векторов, образующих базис текущего аффинного многообразия, может оказаться нарушенной настолько, что вызовет существенное отклонение начальной точки макрошага x^s от истинной точки минимума в соответствующем аффинном многообразии $L(J_s)$, и указанное свойство будет утрачено.

Ясно, что положение точки минимума в $L(J_s)$ можно уточнять, проводя, например, процесс последовательного спуска вдоль направлений, определяемых базисными векторами. Но независимо от того, какой процедурой уточнения точки минимума мы будем пользоваться, возникают два взаимосвязанных вопроса: 1) когда процедуру уточнения подключать и 2) до какой степени точности уточнение доводить? Обеспечение предельно возможной степени точности на каждом шаге не является необходимым и явно нецелесообразно, ибо ведет к увеличению объема вычислений.

II. Общая схема процесса

Поставленные выше вопросы не связаны с тем, что функция квадратичная, а для ее минимизации на аффинных многообразиях используется метод сопряженных направление, и могут быть сформулированы в более общем виде: каким об-

разом следует организовать процесс субоптимизации и каким требованиям должен удовлетворять метод, используемый для минимизации функции f на аффинных многообразиях, чтобы обеспечить сходимость последовательности x^s к решению исходной задачи? В связи с другими алгоритмами эти вопросы исследовались в работах [2, 3]. В [2] рассмотрена задача минимизации строго выпуклой функции, и хотя предложенный там процесс также использует идею сопряженных направлений, но способ получения и преобразования этих направлений принципиально отличен от примененного в [1]. Поэтому алгоритм [2] не позволяет достаточно просто использовать специальную структуру ограничений задачи, если таковая имеется. Однако сама идея организации процесса субоптимизации, примененная в [2], может быть обоснована для более общего класса алгоритмов и, в частности, позволяет получить требуемую модификацию алгоритма [1]. По существу эта же идея была ранее предложена в работе [3], где рассматривался процесс субоптимизации на основе некоторой комбинации методов Ньютона и сопряженных направлений. Кратко способ организации процесса субоптимизации, о котором идет речь, может быть сформулирован следующим образом: выбрав по некоторому правилу в начале макрошага одну из представляющихся возможностей - расширять имеющееся многообразие или проводить уточнение точки минимума в нем, - осуществляем последующие сужения до тех пор, пока не будет выполнен полный шаг используемого метода минимизации на аффинных многообразиях, т.е. шаг, длина которого не лимитируется ограничениями задачи (1); после этого начинается новый макрошаг.

Ясно, что такой процесс может сходиться к решению исходной задачи лишь при некоторых условиях, налагаемых на используемый метод минимизации на аффинных многообразиях и упомянутое правило выбора.

Будем предполагать, что метод минимизации функции f на аффинных многообразиях обладает следующими двумя свойствами, характерными для процессов градиентного типа:

(i) Существует такая положительная константа θ, что выполняется неравенство

$$(\nabla f(x^k),\ x^k - x^{k+1}) \geq \theta \,\|\nabla f(x^k)\| \cdot \|x^k - x^{k+1}\|, \quad (4)$$

где x^k – текущая точка процесса, а $\nabla f(x^k)$ – градиент функции f в точке x^k.

(ii) Метод обеспечивает сходимость (по значению функции) в сочетании с любым другим процессом, гарантирующим лишь невозрастание значения функции. Более точно, при любом выборе точек x^{2k} ($k=1,2,\dots$) таких, что $f(x^{2k})\le f(x^{2k-1})$, будет

$$\lim_{i\to\infty} f(x^i) = \inf_x f(x),$$

если только переход от x^{2k} к x^{2k+1} осуществляется посредством основного метода минимизации.

Прежде чем сформулировать требование к правилу выбора, введем некоторые обозначения. Для краткости обозначим многообразие $L(I_s)$ через L_s, а параллельное ему подпространство через P_s. Кроме того, через $\nabla_{P_s} f(x)$ будем обозначать проекцию градиента $\nabla f(x)$ на подпространство P_s.

Многообразие L_s может быть расширено до многообразия L_s^i исключением из множества I_s какого-либо из его элементов: $L_s^i = L(I_s\setminus\{i\})$. Для определения того, следует ли проводить расширение, и если да, то какой элемент i при этом выбрать, возьмем для каждого $j\in I_s$ некоторый вектор q^{sj} из множества $P_s^j\setminus P_s$ (P_s^j – подпространство, параллельное многообразию L_s^j). Каждый такой вектор должен удовлетворять условиям:

$$\begin{aligned}&(q^{sj}, a_i)=0, \quad i\in I_s\setminus\{j\},\\ &(q^{sj}, a_j)\ne 0.\end{aligned} \tag{5}$$

Пусть, для определенности,

$$(q^{sj}, a_j)=1. \tag{6}$$

Для одного и того же многообразия L, фигурирующего в качестве L_s, наборы векторов q^{sj} при разных значениях s могут быть разными. Однако мы будем предполагать ограниченность множества этих векторов:

$$\|q^{sj}\|\le\gamma, \quad j\in I_s, \quad s=1,2,\dots .$$

Это условие вместе с условием (6) обеспечивает ограниченность снизу положительной константой величины угла между подпространством P_s и векторами q^{sj}.

Теперь каждую из имеющихся возможностей можно характеризовать некоторым числовым значением: расширению многообразия L_s до многообразия L_s^i соответствует $\partial f / \partial q^{i} (x^s)$, а фиксации многообразия L_s без изменения соответствует величина $-\|\nabla_{P_s} f(x^s)\|$. Обозначим через d_s значение той из указанных величин, которая соответствует выбираемой возможности, т.е.

$$d_s = \frac{\partial f}{\partial q^{i}} (x^s), \tag{7}$$

если L_s расширяется до L_s^i, и

$$d_s = -\|\nabla_{P_s} f(x^s)\|, \tag{8}$$

если на данном шаге мы решаем оставить L_s без изменения, т.е. проводить уточнение точки минимума в нем (возможно, с последующими сужениями до выполнения полного шага применяемого метода минимизации на аффинных многообразиях).

Наряду с d_s введем величину M_s, характеризующую лучшую из имеющихся возможностей:

$$M_s = \min\{-\|\nabla_{P_s} f(x^s)\|, \min_{i \in I_s} \frac{\partial f}{\partial q^{i}} (x^s)\}.$$

Теперь мы можем сформулировать требование к правилу выбора. Оно состоит в выполнении условия:

если $d_s \xrightarrow[s \in S]{} 0$ для некоторого множества $S \subset \{1,2,\dots\}$, (9)

то и $M_s \xrightarrow[s \in S]{} 0$.

Доказательству основного утверждения предпошлем следующую лемму.

ЛЕММА. Если $x^s \xrightarrow[s \in S]{} x^*$ и $M_s \xrightarrow[s \in S]{} 0$, то x^* является решением исходной задачи.

ДОКАЗАТЕЛЬСТВО. Не ограничивая общности, можно считать, что множество I_s, а следовательно, многообразие L_s и подпространство P_s при всех $s \in S$ одни и те же: $I_s = \hat{I}$, $L_s = \hat{L}$, $P_s = \hat{P}$. Кроме того, ввиду ограниченности множества

векторов q^{sj}, можно считать, что для всех $j \in \hat{I}$ имеет место сходимость векторов q^{sj} : $q^{sj} \xrightarrow[s \in S]{} q^j$.

По непрерывности будем иметь $\| \nabla_{\hat{p}} f(x^*) \| = 0$, т.е. точка x^* доставляет минимум функции f на многообразии $\hat{L}$. Ввиду этого, найдутся такие числа λ_i, что

$$\nabla f(x^*) = \sum_{i \in \hat{I}} \lambda_i a_i .$$

Для оптимальности x^* в исходной задаче требуется лишь показать неотрицательность чисел λ_i. Если бы для некоторого $j \in \hat{I}$ оказалось, что $\lambda_j < 0$, то для соответствующего вектора q^j, ввиду (5)–(6), имели бы

$$\frac{\partial f}{\partial q^j}(x^*) = (\nabla f(x^*), q^j) = \lambda_j .$$

Тогда по непрерывности для достаточно больших значений $s \in S$ было бы

$$\frac{\partial f}{\partial q^{sj}}(x^s) < \delta < 0,$$

где δ – некоторая константа. А это противоречит предположению, что M_s при $s \in S$ стремится к нулю. Лемма доказана.

ТЕОРЕМА. Если последовательность x^s ограничена, применяемый метод минимизации на аффинных многообразиях обладает приведенными выше свойствами (i), (ii), а для принятого правила выбора выполняется условие (9), то любая предельная точка последовательности x^s является решением исходной задачи.

ДОКАЗАТЕЛЬСТВО. Утверждение достаточно очевидно, если в процессе происходит стабилизация аффинного многообразия, т.е., начиная с некоторого момента, множество I_s не меняется. В этом случае процесс превращается в поиск минимума функции f на $L(I_s)$, а величина d_s задается формулой (8). Ясно, что $d_s \to 0$, а тогда, ввиду (9), $M_s \to 0$, и требуемое следует из леммы.

Предположим, что стабилизация многообразия не происходит. Но так как различных многообразий лишь конечное число, то в последовательности многообразий L_s будут содержаться циклы, т.е. такие отрезки последовательности, в которых первое и последнее многообразия совпадают. Какой-то из таких циклов будет повторяться бесконечно много раз. При этом может оказаться, что на таком цикле встречается ситуация, когда, в случае расширения многообразия, удается сделать сразу полный шаг, т.е. данный макрошаг не содержит сужений. Пусть, для определенности, речь идет о многообразии L и его расширении L^{κ}. Пусть $L = L_{s-1}$, $L_{\kappa} = L_s$ при $s \in S$. Не ограничивая общности, ввиду ограниченности, можно считать, что подпоследовательность $x^s (s \in S)$ сходится. На последовательности точек $x^s, x^{s+1} (s \in S)$ функция f убывает, причем x^{s+1} получается из x^s выполнением одного шага используемого метода минимизации. Тогда, по свойству (ii), будем иметь

$$\lim_{s \in S} f(x^s) = \min_{x \in L^{\kappa}} f(x),$$

и, следовательно, $\|\nabla_{p^{\kappa}} f(x^s)\| \xrightarrow[s \in S]{} 0$. Поэтому производная по любому направлению в L^{κ} должна стремиться к нулю. В частности,

$$d_s = \frac{\partial f}{\partial q^{s\kappa}} (x^s) \xrightarrow[s \in S]{} 0,$$

и по лемме заключаем, что предел подпоследовательности $x^s (s \in S)$ является решением исходной задачи. Ввиду монотонного убывания значений $f(x^s)$ ($s = 1, 2, \ldots$), из этого следует утверждение теоремы.

Предположим теперь, что на рассматриваемом цикле указанных ситуаций нет: каждый макрошаг содержит хоть одно сужение. Но если бы было более одного сужения, то на цикле должно бы быть по крайней мере два последовательных расширения, т.е. макрошаг без сужений, что противоречит предположению. Таким образом, в этом случае каждый макрошаг состоит из расширения и ровно одного сужения. Пусть на цикле встречаются последовательно многообразия L', L, L'' и L получено расширением L', а L'' — сужением L. Пусть $L' = L_{s-1}$, $L = L_s$, $L'' = L_{s+1}$ при $s \in S$. Тогда $x^{s+1} = x^s - t_s h^s \in L''$, а $x^s = x^{s-1} \in L'$. Каждое из мно-

гообразий L' и L'' будет обладать тем свойством, что на них будет содержаться бесконечно много полных шагов применяемого метода минимизации на аффинных многообразиях: на L' полным шагом будет переход от x^{s-2} к x^{s-1}, а на L'' – переход от x^{s+1} к x^{s+2}. Следовательно, сходящаяся подпоследовательность на x^s, $s \in S$, будет сходиться к точке минимума f на L'. Прореживая ее, если нужно, можно добиться того, чтобы соответствующая ей подпоследовательность точек x^{s+1} сходилась к точке минимума на L''. Чтобы не усложнять обозначения, считаем, что уже последовательность x^s, $s \in S$, обладает требуемым свойством. Пусть $x^s \xrightarrow[s\in S]{} x^*$. Достаточно показать, что x^* является точкой минимума на L. Тогда, повторяя рассуждения предыдущего случая, мы получим, что x^* – точка минимума на многогранном множестве, определяемом условиями (I).

Предположим противное:

$$|\nabla_p f(x^*)| \neq 0$$

(P – подпространство, параллельное L).

Из предположения (i) об используемом методе минимизации на аффинных многообразиях имеем:

$$(\nabla_p f(x^s), h^s) \geq \theta\, |\nabla_p f(x^s)| \cdot |h^s|. \tag{10}$$

Предполагая некоторую нормировку векторов h^s, получаем, что правая часть этого неравенства в пределе строго положительна. Тогда ясно, что $t_s \xrightarrow[s\in S]{} 0$. Таким образом, $x^{s+1} \xrightarrow[s\in S]{} x^*$, т.е. x^* – точка минимума функции f на многообразии L''.

Далее, пусть $L' = L(I')$, $L = L(I)$, $L'' = L(I'')$ и $I = I' \setminus \{k\}$. Так как x^* является точкой минимума функции f на L', то

$$\nabla f(x^*) = \sum_{i \in I'} \lambda_i a_i . \tag{11}$$

При этом $\lambda_k \neq 0$, ибо в противном случае x^* являлась бы точкой минимума и на L. Умножая скалярно (11) на вектор q^{sk}, получим, ввиду (5)–(6),

$$\frac{\partial f}{\partial a^{sk}}(x^*) = \lambda_k . \tag{12}$$

С другой стороны, имеем

$$\frac{\partial f}{\partial q^{sk}}(x^k) = d_s < 0 .$$

По непрерывности получаем $\lambda_k < 0$. Теперь, умножая (II) скалярно на вектор h^s , аналогично (I2) получим

$$\frac{\partial f}{\partial h^s}(x^k) = (h^s, a_k)\lambda_k .$$

Левая часть этого равенства при больших значениях $s \in S$ будет мало отличаться от левой части неравенства (I0) и, следовательно, будет положительной. В итоге получаем, что при больших значениях $s \in S$ будет $(h^s, a_k) < 0$ и значит $k \notin I''$.

Таким образом, множества I' и I'' различны. Ввиду условия невырожденности, отсюда следует различие многообразий L' и L'' и, следовательно, их аффинная оболочка совпадает с L . Но тогда x^* - точка минимума функции f на L . Полученное противоречие завершает доказательство теоремы.

Ш. Конкретизация общей схемы для квадратичного программирования

I. Прежде чем переходить к конкретизации вышеизложенной общей схемы для интересующего нас алгоритма квадратичного программирования [I], сделаем несколько замечаний общего характера относительно выбора системы векторов q^{si} и величины d_s .

Простейший способ выбора d_s , обеспечивающий выполнение условия (9), состоит в том, что принимается $d_s = M_s$. Возможен, однако, и более эффективный способ, используемый в процедурах симплекс-метода, - так называемый "барьерный принцип", который в данном случае будет состоять в следующем. Зафиксировав некоторую положительную величину B (в начале процесса достаточно большую), проверяем неравенство

$$|\nabla_{q_s} f(x^s)| < B .$$

Если оно нарушено, то d_s определяется согласно (8). В противном случае в качестве d_s принимается первая из производных $\frac{\partial f}{\partial q^{si}}(x^s)$, $i \in I_s$, меньшая величины $-B$. Если же ока-

залось, что таких нет, то величина B уменьшается: принимаем, например, $B = \tau M_s$ при $\tau \in (0, 1)$. Ясно, что такое правило обеспечивает выполнение условия (9).

Относительно выбора системы векторов q^{si} следует заметить, что нас интересуют не сами векторы q^{si}, а производные $\frac{\partial f}{\partial q^{si}}(x^s)$. Выбирая векторы q^{si} надлежащим образом, можно вычислять указанные производные, не получая явно самих векторов q^{si}. Конкретизация способа выбора зависит от особенностей используемого метода минимизации на аффинных многообразиях и способа проверки текущего решения x^s на оптимальность. Например, если при проверке оптимальности предполагается получать разложение

$$\nabla f(x^s) = \nabla_{P_s} f(x^s) + \sum_{i \in I_s} \lambda_i a_i$$

(как в методе проекций градиента), то целесообразно принять q^{si} ортогональными к P_s. Тогда из (5) и (6) будет следовать

$$\frac{\partial f}{\partial q^{si}}(x^s) = \lambda_i \tag{13}$$

В интересующем нас алгоритме [1] для проверки оптимальности текущего приближения x^s решается система линейных уравнений

$$\sum_{i \in \mathcal{B}_s} \lambda_i a_i = \nabla f(x^s), \tag{14}$$

где множество $\mathcal{B}_s$ такое, что $\mathcal{B}_s \supset I_s$, и соответствующие векторы a_i $(i \in \mathcal{B}_s)$ образуют базис пространства R^n; для полученного решения системы (14) проверяются условия:

$$\lambda_i = 0, \quad i \in \mathcal{B}_s \setminus I_s,$$
$$\lambda_i \geq 0, \quad i \in I_s.$$

В данном случае в качестве векторов q^{si} целесообразно взять решения систем уравнений

$$\begin{cases} (a_i, q^{si}) = 1, \\ (a_j, q^{si}) = 0, \quad j \in \mathcal{B}_s \setminus \{i\}. \end{cases} \tag{15}$$

Для таких векторов q^{si} при $i \in I_s$ снова будет выполнено равенство (13).

аналогично для вычисления величины $|\nabla_{P_s} f(x^s)|$ нет необходимости в выполнении трудоемкой операции проектирования вектора $\nabla f(x^s)$ на подпространство P_s, если надлежащим образом ввести норму. Заметим, что векторы q^{si}, являющиеся решениями систем (15) для $i \in \mathcal{B}_s \setminus I_s$, будут образовывать базис подпространства P_s. Поэтому можно принять

$$|\nabla_{P_s} f(x^s)| = \max_{i \in \mathcal{B}_s \setminus I_s} \left| \frac{\partial f}{\partial q^{si}}(x^s) \right|.$$

Учитывая (13), получаем

$$|\nabla_{P_s} f(x^s)| = \max_{i \in \mathcal{B}_s \setminus I_s} |\lambda_i|.$$

Хотя таким образом введенная норма в подпространстве P_s будет зависеть от выбора множества $\mathcal{B}_s$, однако, ввиду конечности числа различных множеств $\mathcal{B}_s$, утверждение леммы сохранится, а это лишь и требуется.

2. Не уточняя до конца выбор d_s (к этому вопросу мы вернемся ниже), рассмотрим вопрос минимизации функции f на аффинных многообразиях в предположении, что для каждого аффинного многообразия L_s нам известна система "почти сопряженных" векторов, образующих базис соответствующего подпространства P_s. Предполагается, что эта система векторов преобразуется от шага к шагу, как в алгоритме [1]: при расширении многообразия L_s до многообразия L_s^k строится расширяющее направление $\hat{q}$ путем C-ортогонализации вектора q^{sk} к подпространству P_s, а при сужении аффинного многообразия упомянутая система векторов корректируется при помощи некоторого ортогонального преобразования.

Пусть L - аффинное многообразие и $q^1, \ldots, q^r$ - соответствующая система векторов. Рассмотрим сначала случай, когда функция f на многообразии L строго выпукла. Это означает, что $(q^i, Cq^i) > 0$, $i = 1, \ldots, r$. Не ограничивая общности, можно считать, что $(q^i, Cq^i) = 1$ ($i = 1, \ldots, r$).

Для минимизации функции f на многообразии L можно воспользоваться процедурой спуска по наиболее перспективному из направлений q^i, т.е. по направлению с максимальным модулем производной функции f. Ясно, однако, что если сопряженность

векторов q^i нарушена не очень сильно, то целесообразнее проводить процесс одновременного спуска по всем направлениям. А именно речь идет о следующем процессе.

Пусть векторы q^i образуют матрицу Q, являясь ее столбцами. Процесс одновременного спуска по такой системе направлений будет описываться формулой

$$x^{k+1} = x^k - QQ^T \nabla f(x^k).$$

Вводя замену неизвестных по формуле $x = Qy$ и подставляя выражение для градиента функции f, получаем иное представление процесса:

$$Qy^{k+1} = Qy^k - QQ^T (CQy^k + p). \tag{16}$$

Если бы векторы q^i были в точности сопряженными, то было бы $Q^T C Q = E$ (E - единичная матрица). Потребуем, чтобы для имеющейся системы векторов q^i выполнялось при некотором фиксированном числе $\theta \in (0,1)$ неравенство

$$| E - Q^T C Q | \le \theta. \tag{17}$$

Ясно, что из этого следует линейная независимость векторов q^i, а тогда из (16) имеем:

$$y^{k+1} = (E - Q^T C Q) y^k - Q^T p.$$

Таким образом, рассматриваемый процесс представляет собой процесс метода итераций для решения системы линейных уравнений

$$Q^T C Q y = - Q^T p,$$

который, ввиду (17), будет сходящимся: $y^k \xrightarrow[k\to\infty]{} \hat{y}$. Очевидно, что соответствующий вектор $\hat{x} = Q\hat{y}$ будет являться решением задачи минимизации функции f на аффинном многообразии L .

3. Проверим для описанного метода выполнение условия (i) . В данном случае направление смещения на k -м шаге процесса определяется вектором

$$h = QQ^T \nabla f(x^k).$$

Ясно, что $Q^T \nabla f(x^k) = Q^T \nabla_p f(x^k)$. Поэтому

$$(\nabla_p f(x^k), h) = (\nabla_p f(x^k),\ QQ^T \nabla_p f(x^k)) = |Q^T \nabla_p f(x^k)|^2. \tag{18}$$

Так как векторы q^i ($i = 1,\dots, r$) образуют базис подпространства P, то вектор $\nabla_p f(x^k)$ можно представить в виде $\nabla_p f(x^k) = Q\alpha$, где α - некоторый ненулевой вектор (предполагается, что $\nabla_p f(x^k) \neq 0$). Таким образом,

$$Q^T \nabla_p f(x^k) = Q^T Q \alpha . \tag{19}$$

Как отмечалось выше, векторы q^i ($i = 1,\dots, r$) линейно-независимы и, следовательно, матрица $Q^T Q$ неособенная. Получим оценку снизу для собственных чисел этой матрицы.

Из (17) следует, что собственные числа матрицы $Q^T C Q$ принадлежат отрезку $[1-\sigma, 1+\sigma]$. Поэтому для любого y выполняется неравенство

$$(1-\sigma)^2 |y|^2 \leqslant (y, Q^T C Q y) \leqslant (1+\sigma)^2 |y|^2 . \tag{20}$$

Кроме того, так как квадратичная форма (x, Cx) на подпространстве P строго положительна, то существуют константы M, $m > 0$, такие, что для любого вектора $x \in P$ будет

$$m^2 |x|^2 \leqslant (x, Cx) \leqslant M^2 |x|^2 . \tag{21}$$

Теперь из (20) и (21) получаем

$$\frac{1-\sigma}{M} |y| \leqslant |Qy| \leqslant \frac{1+\sigma}{m} |y| . \tag{22}$$

Таким образом,

$$|Q| \leqslant \frac{1+\sigma}{m} , \tag{23}$$

а собственные числа матрицы $Q^T Q$ принадлежат отрезку $[(1-\sigma)/M, (1+\sigma)/m]$.

Вернемся к равенству (19). Теперь имеем

$$|Q^T \nabla_p f(x^k)| \geqslant \frac{1-\sigma}{M} |\alpha| \geqslant \frac{1-\sigma}{M|Q|} |Q\alpha| = \frac{1-\sigma}{M|Q|} |\nabla_p f(x^k)| .$$

Окончательно, учитывая (18), получим

$$(\nabla_p f(x^k), h) \geqslant \frac{(1-\sigma)^2}{M^2 |Q|^2} |\nabla_p f(x^k)|^2 \geqslant \frac{(1-\sigma)^2}{M^2 |Q|^2 |Q Q^T|} |\nabla_p f(x^k)| \cdot |h| .$$

Ввиду того, что константы m и M не зависят от выбора конкретной матрицы Q, удовлетворяющей условию (17), оценка (23) определяется лишь многообразием L. Так как различных многообразий $L_s = L(I_s)$ конечное число, то тем самым можно утверждать существование такой положительной константы θ, не зависящей от выбора многообразия L_s, что для метода одновременного спуска будет выполняться условие (i).

Проверим теперь выполнение условия (ii). Как и для градиентных процессов, в данном случае оно следует из достаточно быстрого убывания значения функции на каждом шаге метода. Имеем:

$$f(z^k) - f(z^{k+1}) = (h, \nabla f(z^k)) - \frac{1}{2}(h, Ch) =$$

$$= (\nabla f(z^k), QQ^T \nabla f(z^k)) - \frac{1}{2}(\nabla f(z^k), QQ^T C QQ^T \nabla f(z^k)).$$

Обозначив вектор $Q^T \nabla f(z^k)$ для краткости через y^k, можем это равенство переписать в виде:

$$f(z^k) - f(z^{k+1}) = (y^k, y^k) - \frac{1}{2}(y^k, Q^T C Q y^k) =$$

$$= \frac{1}{2}\left[(y^k, y^k) - (y^k, (Q^T C Q - E) y^k)\right].$$

Теперь, учитывая, что собственные числа матрицы $Q^T C Q - E$ не превосходят σ, получаем

$$f(z^k) - f(z^{k+1}) \geq \frac{1}{2}(y^k, y^k)\sigma^2. \qquad (24)$$

В результате из сходимости последовательности $f(z^k)$ следует сходимость вектора y^k к нулю, что эквивалентно сходимости к нулю вектора $\nabla_p f(z^k)$, а это и требуется.

4. Мы рассмотрели метод минимизации функции f на таких многообразиях L, на которых функция f строго выпукла. Однако в процессе могут встретиться многообразия L, для которых это условие нарушается. Такие многообразия в дальнейшем будут именоваться "особенными".

Ясно, что неособенное многообразие L_s может перейти в особенное лишь при его расширении. Пусть L_s расширяется до особенного многообразия L_s^k. Тогда для нового сопряженного направления $\hat{q}$, полученного C-ортогонализацией вектора q^{sk} к подпространству P_s, окажется $(\hat{q}, C\hat{q}) = 0$. Это

означает, что $C\hat{q}=0$, и производная функции f по направлению $\hat{q}$ постоянна и равна $(p,\hat{q})$. Если $(p,\hat{q})<0$, то смещаем текущую точку x^s в направлении $\hat{q}$. При этом либо окажется, что направление $\hat{q}$ не выводит из множества допустимых векторов исходной задачи, и, следовательно, исходная задача неразрешима, либо произойдет сужение многообразия L_s^k до некоторого, как легко показать (см. [I]), неособенного многообразия. Если же $(p,\hat{q})>0$, то многообразие L_s^k снова сужаем до L_s. В соответствии с общим правилом после любого из указанных сужений следует производить уточнение точки минимума на суженном многообразии (до выполнения полного шага процесса одновременного спуска).

Таким образом, точки последовательности x^s будут принадлежать лишь неособенным из многообразий L_s. Из этого, ввиду монотонного убывания последовательности $f(x^s)$ и конечности числа различных многообразий L_s, следует ограниченность последовательности x^s.

Покажем, что для изложенного процесса будет выполняться утверждение приведенной выше теоремы, и более того, последовательность x^s будет сходящейся.

В доказательстве теоремы нужно дополнить лишь последнюю часть рассмотрением случая, когда многообразие L особенное. В принятых обозначениях в этом случае будем иметь $x^{s+1}=x^s+t_s\hat{q}$ при $s\in S$. Как и ранее, нужно показать, что точка $x^*=\lim\limits_{s\in S}x^s$ является точкой минимума функции f на многообразии L. Предполагая противное, на основании прежних рассуждений получим, что в разложении (II) для вектора $\nabla f(x^*)$ будет $\lambda_k<0$. Но, как легко видеть, $(\hat{q},a_k)=(q^{sk},a_k)=1$, и потому

$$(p,\hat{q})=(\hat{q},\nabla f(x^*))=\lambda_k(\hat{q},a_k)=\lambda_k<0.$$

Теперь, так как производная функции f по направлению $\hat{q}$ постоянна и отрицательна, то на многообразии L функция f не ограничена снизу. С другой стороны, по этой же причине из ограниченности последовательности x^s следует, что $t_s \xrightarrow[s\in S]{} 0$. Кроме того, так как $(\hat{q},a_k)>0$, то $k\notin I^*$. Повторяя рассуждения заключительной части доказательства теоремы, получим, как и прежде, что x^* является точкой минимума функции f на многообразии L. Полученное противоречие завершает доказательство.

Покажем, что последовательность x^s сходится. Так как каждая предельная точка является решением исходной задачи, то на неособенном многообразии не может быть более одной предельной точки: ввиду выпуклости множества решений исходной задачи, на отрезке, соединяющем две предельные точки, функция f была бы постоянной, что противоречило бы неособенности многообразия. Ввиду конечности числа различных многообразий L_s, требуемое утверждение получается из стремления к нулю разности $x^{s+1}-x^s$: если перемещение из x^s в x^{s+1} происходит по особенному многообразию, то это следует из приведенных выше рассуждений, а в случае неособенного многообразия это следует из неравенства (24).

5. С целью повышения эффективности описанного процесса следует несколько уточнить правило выбора d_s, ибо если не предусматривать никаких дополнительных мер, то может встретиться такая ситуация: хотя принятое правило выбора d_s диктует необходимость расширения многообразия L_s, однако после построения расширяющего направления $\hat{q}$ может оказаться, что производная функции f по этому направлению положительна и, следовательно, спуск по этому направлению невозможен. Это повлечет за собой возврат к многообразию L_s. Формально алгоритм будет работать и в этой ситуации: последуют сужения многообразия до выполнения полного шага процедуры одновременного спуска. Но чтобы исключить лишние вычисления, связанные с построением направления $\hat{q}$, можно предупредить такую ситуацию, проводя модификацию процесса, использованную в [3].

Пусть многообразие L_s расширяется до многообразия L_s^k. Тогда вектор $\hat{q}$ ищем в виде

$$\hat{q} = q^{sk} - Q\beta\ , \tag{25}$$

где Q, как и выше, – матрица "почти сопряженных" направлений многообразия L_s. Вектор β определяется из системы линейных уравнений

$$Q^T C Q \beta = Q^T C q^{sk} . \tag{26}$$

Из условия (17) и неравенства (23) следует ограниченность компонент вектора $\beta : |\beta_i| \leq R$.

Из (25) получаем

$$(\hat{q}, \nabla f(x^s)) = (q^{sk}, \nabla f(x^s)) - \sum_i \beta_i (q^i\ \nabla f(x^s)). \tag{.7}$$

Теперь, если определение α_s производить по барьерному принципу, то

$$(q^{sk}, \nabla f(x^s)) < -B, \tag{28}$$

и из неравенства

$$R \sum_i |(q^i, \nabla f(x^s)| < \frac{1}{2} B \tag{29}$$

будет следовать

$$(\hat{q}, \nabla f(x^s)) < -\frac{1}{2} B. \tag{30}$$

Таким образом, можно гарантировать возможность спуска по направлению $\hat{q}$, если принять такое правило организации процесса: при нарушении неравенства (29) расширять многообразие L_t на данном шаге не следует; если же неравенство (29) выполнено, то вопрос о расширении многообразия решается в соответствии с барьерным принципом. Ясно, что такая организация процесса укладывается в рассмотренную ранее схему: речь идет просто об ином способе введения нормы при вычислении $|\nabla_s f(x^s)|$.

Указать заранее значение константы R, не очень завышая ее, весьма затруднительно. Целесообразнее подбирать ее в ходе процесса: если при некотором значении R окажется, что неравенство (30) нарушается, в то время как неравенства (28) и (29) выполнены, то это означает, что значение R недостаточно велико; следует продолжить процесс в соответствии с общим правилом, увеличив, однако, значение R (например, в два раза).

В заключение остановимся кратко на вопросе практической реализации описанного процесса.

Может оказаться, что неравенство (I7) нарушится при проведении сужения текущего аффинного многообразия. В этом случае можно улучшить сопряженность векторов q^i, используя процесс C-ортогонализации. Для того чтобы эта ситуация не возникала слишком часто, можно зафиксировать два значения величины σ: $\sigma_1 < \sigma_2$. При проведении процесса C-ортогонализации будем пользоваться, проверяя неравенство (I7), значением σ_1, а в остальных случаях - значением σ_2. Для построения вектора $\hat{q}$ в случае расширения текущего многообразия мы предполагали, что система (26) решается точно. При практической реализации алгоритма можно поступать следующим образом. Задавшись некоторой границей для величины $|\hat{q}| / (\hat{q}, C\hat{q})$, проводим реше-

ние системы (26) по методу итераций (что возможно ввиду (17)) до тех пор, пока для текущего приближения $\hat{q}$ либо нарушится упомянутая граница, и в этом случае мы считаем $(\hat{q}, C\hat{q}) = 0$, либо вектор $\tilde{q} = \hat{q}/(\hat{q}, C\hat{q})$ с достаточной степенью точности C - ортогонален к векторам q^i, т.е. если присоединить вектор $\tilde{q}$ к матрице Q, то для полученной матрицы неравенство (17) будет выполнено. Ясно, что на некотором шаге метода итераций одна из указанных возможностей непременно реализуется.

Литература

1. ШМЫРЕВ В.И. Об алгоритмах метода последовательного улучшения для задачи квадратичного программирования на основе схемы сопряженных направлений. - В кн.: Оптимизация, Вып.14(31), Новосибирск, 1974, с. 71-82.
2. RITTER K. A method of conjugate directions for linearly constrained nonlinear programming problems. - "SIAM Jornal of numerical analisis", 1975, v.12,N3
3. БУЛАВСКИЙ В.А., ЯКОВЛЕВА М.А. Метод последовательного улучшения для выпуклых задач с линейными ограничениями. - В кн.: Оптимизация, Вып. 5(22), Новосибирск, 1972, с. 37-62.
4. РУБИНШТЕЙН Г.Ш., ШМЫРЕВ В.И. Методы минимизации квазивыпуклой функции на многограннике. - В кн.: Оптимизация. Вып. 5(22). Новосибирск, 1972, с. 133-197.
5. ЗАНГВИЛЛ У.И. Нелинейное программирование. М., "Сов. радио", 1973, с. 165-173.

Поступила в ред.-изд. отд.
17. X. 1976 г.