УДК 512.25/26

ОБ АЛГОРИТМАХ МЕТОДА ВОЗМОЖНЫХ НАПРАВЛЕНИЙ С КВАДРАТИЧНОЙ НОРМАЛИЗАЦИЕЙ

В.И.Шмырёв

В задаче определения подходящего направления, решаемой на каждом шаге метода возможных направлений [1], ввиду нетелесности в общем случае конуса возможных направлений, линейные и нелинейные ограничения учитываются различно. Отсутствие телесности обусловливает необходимость точного решения задач определения направления, которые в зависимости от используемой нормализации направлений являются задачами линейного или квадратичного программирования.

Предлагаемый в настоящей заметке подход, основанный на вариации правых частей линейных ограничений, позволяет учитывать ограничения единообразно и ограничиться лишь приближенным решением задач определения направления. Использование двойственной задачи при квадратичной нормализации направлений приводит в этом случае к легко реализуемым итеративным алгоритмам. В частности, на этом пути получается итеративный алгоритм и для задач линейного программирования. Попутно устанавливается связь алгоритмов, предложенных для задач выпуклого программирования в [2], с методом возможных направлений.

Будем рассматривать задачу выпуклого программирования в обычной формулировке: минимизировать значение выпуклой функции f на выпуклом множестве $\Omega \subset R^n$, определяемом условиями

$$g_i(x) \le 0, \quad i \in I_1 \cup I, \tag{1}$$

где g_i - выпуклые функции для $i \in I_1$ и аффинные - для $i \in I_2$. Наиболее трудоёмкой частью при реализации схемы возможных направлений, предложенной Г.Зойтендейком в [I], является определение для имеющейся точки $x^k \in \Omega$ подходящего направления h^k. В предположении, что выполняется условие Слейтера[ж)], а функции f и g_i непрерывно дифференцируемы, для определения h^k в [I] формулируется вспомогательная экстремальная задача, состоящая в минимизации величины ρ при условиях:

$$1)\ (\operatorname{grad} g_i(x^k), h) \leq \rho\theta_i, \quad i \in I(x^k, \varepsilon), \tag{2}$$

$$2)\ (\operatorname{grad} f(x^k), h) \leq \rho, \tag{3}$$

$$3)\ \|h\| \leq 1, \tag{4}$$

где θ_i равны нулю при $i \in I_2$ и произвольные положительные числа, выбираемые из дополнительных соображений, при $i \in I_1$;

ε - устремляемый к нулю положительный параметр, вводимый для того, чтобы обеспечить сходимость к оптимуму (преодоление "заедания"); $I(x^k, \varepsilon) = \{ i \in I_1 \cup I_2 \mid -\varepsilon \leq g_i(x^k) \leq 0 \}$; $\|\cdot\|$ - какая-либо норма в R^n.

После определения оптимального решения ρ_k, h^k этой задачи дальнейший ход итерации сводится к следующему. Если $\rho_k < 0$, то имеющуюся точку x^k можно улучшить, осуществляя сдвиг в направлении вектора h^k до границы области Ω или до точки минимума функции f по найденному направлению в зависимости от того, какой из этих случаев реализуется при меньшей величине сдвига. Ход всего процесса регулируется двумя устремляемыми к нулю параметрами ε и μ, именно: при фиксированном $\varepsilon > 0$ величина ρ_k с ходом процесса стремится к нулю и, следовательно, для достаточно больших значений k станет больше $-\mu$ при любом $\mu > 0$. В этом случае говорят, что точка x^k (ε, μ)-стационарная. Таким образом, при любых $\varepsilon > 0$ и $\mu > 0$ (ε, μ)- стационарная точка получается за конечное число шагов. Получив (ε, μ)-стационарную точку, параметры ε и μ уменьшают. При надлежащих предположениях лю-

---

ж) Условие Слейтера состоит в существовании точки $x^0 \in \Omega$, для которой нелинейные из неравенств (I) выполняются как строгие.

бая последовательность (ε,μ) -стационарных точек при $\varepsilon\to 0$ и $\mu\to 0$ сходится "по функционалу" к решению исходной задачи, т.е. если $x(\varepsilon,\mu)$ - (ε,μ)-стационарные точки, то

$$f(x(\varepsilon,\mu)) \xrightarrow[\substack{\varepsilon\to 0\\ \mu\to 0}]{} f(x^*),$$

где x^* - решение исходной задачи.

В [I] показывалось, что в случае квадратичной нормализации вектора h, т.е. когда $\|h\| = (h,h)^{1/2}$, задача (2-4) при $\rho_\kappa < 0$ эквивалентна задаче квадратичного программирования вида (для простоты считается $\theta_i = 1$, $i\in I_1$): минимизировать

$$(S,S) \tag{5}$$

при условиях:

$$(d_i,s)\geqslant 1,\quad i\in\hat{I}_1, \tag{6}$$

$$(d_i,s)\geqslant 0,\quad i\in\hat{I}_2, \tag{7}$$

где через d_0 обозначен вектор $\operatorname{grad} f(x^\kappa)$ и через d_i - векторы $\operatorname{grad} g_i(x^\kappa)$, $i\in I_1\cup I_2$, а $\hat{I}_1 = \{0\}\cup(I(x^\kappa,\varepsilon)\cap I_1)$, $\hat{I}_2 = I(x^\kappa,\varepsilon)\cap I_2$. Связь между решениями задач (2-4) и (5-7) задаётся соотношением

$$s^\kappa = h^\kappa/\rho_\kappa.$$

Несколько иная задача квадратичного программирования для определения подходящего направления была предложена в [2]. В принятых обозначениях эта задача состоит в минимизации функции

$$\psi(\lambda) = \Big\|\sum_{i\in\hat{I}_1\cup\hat{I}_2} \lambda_i d_i\Big\|^2 \tag{8}$$

при условиях:

$$\sum_{i\in\hat{I}_1} \lambda_i = 1, \tag{9}$$

$$\lambda_i\geqslant 0,\quad i\in\hat{I}_1\cup\hat{I}_2. \tag{10}$$

По решению этой задачи подходящее направление определяется вектором $-\sum_{i\in\hat{I}_1\cup\hat{I}_2} \lambda_i d_i$. Несложные выкладки, однако, показывают, что задача (8-10) эквивалентна задаче, двойственной к задаче (5-7), полученной по схеме, предложенной Г.Ш.Рубинштейном в [3].

Тем самым алгоритмы для задачи выпуклого программирования, рассмотренные в [2], являются алгоритмами возможных направлений с квадратичной нормализацией. Эти алгоритмы основаны на том, что в применении к задаче (8-10) существенно упрощается процедура метода условного градиента [4], которая в случае $I_2 = \varnothing$ за конечное число шагов позволит получить подходящее направление. Получаемая реализация схемы возможных направлений имеет определенные преимущества по сравнению с прочими ввиду малого объема требуемой памяти ЭВМ. Однако при $I_2 \neq \varnothing$ лишь точное решение задачи (8-10) позволяет получить требуемое подходящее направление. Учитывая медленную сходимость метода условного градиента, в этом случае такой путь определения подходящего направления оказывается непригодным.

Указанные затруднения в случае $I_2 \neq \varnothing$ обусловлены отсутствием в общем случае телесности множества Ω. Предлагаемый приём преодоления этих затруднений состоит в следующем.

Зададимся некоторым положительным числом σ и наряду с исходной задачей рассмотрим проварьированную задачу (σ-задачу), заменив ограничения $g_i(x) \leq 0$, $i \in I_2$, на

$$g_i(x) \leq \sigma, \quad i \in I_2. \tag{11}$$

Теперь если для исходной задачи выполнялось условие Слейтера, то множество допустимых точек в σ-задаче является телесным. Тем самым для σ-задачи можно проводить процесс, уже не выделяя аффинных ограничений, т.е. рассматривая при определении подходящего направления задачу минимизации ρ при ограничениях (3-4) и

$$(\mathrm{grad}\, g_i(x^k), h) \leq \rho, \quad i \in I^\sigma(x^k, \varepsilon),$$

где $I^\sigma(x^k,\varepsilon) = \{i \in I_1 \mid -\varepsilon \leq g_i(x^k) \leq 0\} \cup \{i \in I_2 \mid \sigma - \varepsilon \leq g_i(x^k) \leq \sigma\}$.

В соответствующей задаче (8-9) это приведет к таким изменениям: множество $I_1 \cup I_2$ заменится на $I^\sigma = I^\sigma(x^k,\varepsilon) \cup \{0\}$, а в уравнении (9) суммирование будет осуществляться по $i \in I^\sigma$. Теперь уже решение этой задачи нет необходимости получать точно: все векторы h из достаточно малой окрестности вектора h^k, решающего задачу определения направления, будут определять подходящие направления. Получение (ε, μ)-стационарной точки за конечное число шагов можно гарантировать, если вектор $\tilde{h}^k$, по которому производить смещение, обеспечивает определенную часть оптимального значения ρ, например, полсвину:

$$(d_i, \tilde{h}^{\kappa}) \leq \rho_\kappa / 2 , \qquad i \in \hat{I}^{\sigma}.$$

Эти неравенства будут заведомо выполняться, если выполняются более жесткие неравенства, получающиеся из приведенных заменой ρ_κ на его оценку снизу. Такая оценка легко следует из двойственности задач (5-7) и (8-10) и связи задачи определения направления с задачей (5-7):

$$\rho_\kappa \geqslant -\sqrt{\psi(\lambda)}. \tag{12}$$

На основании этой же оценки можно проверить имеющуюся точку x^κ на (ε, μ) -стационарность.

Так как для вектора λ, решающего задачу (8-10), оценка (12) реализуется, то, применяя к задаче (8-10) какую-либо из итеративных процедур (в частности, метод условного градиента), мы за конечное число шагов на основании оценки (12) либо убедимся, что имеющаяся точка x^κ (ε, μ)-стационарна, либо убедимся, что полученное к данному шагу направление $\tilde{h}^\kappa = -\sum_{i \in \hat{I}^{\sigma}} d_i \lambda_i$ является подходящим с требуемой точностью.

Проведение процесса возможных направлений по описанной выше схеме гарантирует при любых $\varepsilon > 0$, $\mu > 0$ и $\sigma > 0$ получение за конечное число шагов (ε, μ)-стационарной точки в σ -задаче. Рассмотрим теперь последовательность σ -задач с уменьшающимся значением параметра σ. Предполагая компактность множества Ω, докажем следующее утверждение.

ТЕОРЕМА. Пусть $\varepsilon_\kappa \to 0$, $\mu_\kappa \to 0$, $\sigma_\kappa \to 0$ и $\{x^\kappa\}$ - последовательность $(\varepsilon_\kappa, \mu_\kappa)$-стационарных точек в σ_κ -задачах. Тогда если $\varepsilon_\kappa < \sigma_\kappa$, то

$$\lim_{\kappa \to \infty} f(x^\kappa) = \min_{x \in \Omega} f(x).$$

ДОКАЗАТЕЛЬСТВО. Из компактности множества Ω следует компактность множества допустимых точек в любой σ_κ -задаче. Ввиду этого без ограничения общности можно считать последовательность $\{x^\kappa\}$ сходящейся: $x^\kappa \to x^*$. Ясно, что $x^* \in \Omega$, ибо $\sigma_\kappa \to 0$. Нужно доказать, что точка x^* будет и оптимальной, т.е.

$$f(x^*) = \min_{x \in \Omega} f(x.$$

Предположим противное - точка x^* не является оптимальной. Тогда должен существовать такой ненулевой вектор z и такое число $\rho < 0$, что выполняются неравенства:

$$
\begin{aligned}
&(\mathrm{grad}\, g_i(x^*), z) < \rho, && i \in I_1(x^*),\\
&(\mathrm{grad}\, f(x^*), z) < \rho, &&\\
&(\mathrm{grad}\, g_i(x^*), z) \leq 0, && i \in I_2(x^*).
\end{aligned}
$$

Среди множеств $I^{\sigma_k}(x^k, \varepsilon_k)$ какое-то будет встречаться при бесконечном числе значений k. Пусть таковым является множество I^0. Множество соответствующих значений k, для которых $I^{\sigma_k}(x^k, \varepsilon_k) = I^0$, обозначим через $K(I^0)$. Для таких k и $i \in I^0$ будет либо $-\varepsilon_k \leq g_i(x^k) \leq 0$, либо $\sigma_k - \varepsilon_k \leq g_i(x^k) \leq \sigma_k$. Переходя по $k \in K(I^0)$ к пределу, получим $g_i(x^*) = 0$, т.е. $I^0 \subset I(x^*)$.

Далее, взяв произвольную точку $x^0 \in \Omega$, будем иметь $g_i(x^0) \leq 0$, $i \in I_2$. С другой стороны, для точки x^k при $k \in K(I^0)$ и $i \in I_2^0 = I_2 \cap I^0$ имеем $\sigma_k - \varepsilon_k \leq g_i(x^k) \leq \sigma_k$. Таким образом, $g_i(x^0) < g_i(x^k)$, т.е.

$$
(\mathrm{grad}\, g_i, v^k) < 0, \quad i \in I_2^0, \quad k \in K(I^0), \tag{13}
$$

где $v^k = x^0 - x^k$. Теперь, учитывая, что $I^0 \subset I(x^*)$, и добавляя вектор v^k с небольшим множителем к z, получим, что при $k \in K(I^0)$ вектор $z^k = z + \alpha v^k$ удовлетворяет системе неравенств

$$
\begin{aligned}
&(\mathrm{grad}\, g_i(x^*), z^k) < 0, \qquad i \in I^0,\\
&(\mathrm{grad}\, f(x^*), z^k) < 0.
\end{aligned}
$$

Это означает, что существуют такой вектор $\bar{z}$ и число $\bar{\rho} < 0$, что

$$
(\mathrm{grad}\, g_i(x^*), \bar{z}) < \bar{\rho}, \qquad i \in I^0, \tag{14}
$$

$$
(\mathrm{grad}\, f(x^*), \bar{z}) < \bar{\rho}. \tag{15}
$$

Ввиду предполагаемой непрерывной дифференцируемости функций g_i и f, система неравенств (14-15) будет выполняться не только для $x = x^*$, но и для всех x из некоторой окрестности ϑ точки x^*. Так как $x^k \to x^*$, то для достаточно больших значений k будет $x^k \in \vartheta$ и, следовательно, при больших $k \in K(I^0)$

будет выполняться система неравенств

$$(grad\, g_i(x^k), \bar{z}) < \bar{p}, \quad i \in I^{\sigma_k}(x^k, \varepsilon_k) - I^0,$$
$$(grad\, f(x^k), \bar{z}) < \bar{p},$$

что противоречит (ε_k, μ_k) -стационарности точек x^k и условию $\mu_k \longrightarrow 0$. Теорема доказана.

Остановимся теперь на переходе от σ_k к σ_{k+1}. Если $\sigma_k < \sigma_{k+1}$, то полученная (ε_k, μ_k) -стационарная точка в σ_k -задаче уже не будет допустимой для σ_{k+1} -задачи. Для получения требуемой допустимой точки можно, отправляясь от полученной (ε_k, μ_k)-стационарной точки, решать задачу минимизации величины σ при ограничениях

$$g_i(x) \leq 0, \quad i \in I_1,$$
$$g_i(x) \leq \sigma, \quad i \in I_2.$$

Эта задача с телесной областью, и для её решения можно воспользоваться уже описанной процедурой. Процесс следует проводить до тех пор, пока не будет получено значение σ, не превосходящее σ_{k+1}. После этого можно перейти к решению σ_{k+1} -задачи.

В заключение остановимся кратко на задаче линейного программирования в канонической постановке: минимизировать (c, x) при условиях

$$(a_i, x) = b_i, \quad i = 1, \ldots, m, \tag{16}$$
$$x \geq 0 \tag{17}$$

Ясно, что в предположении существования допустимых векторов в этой задаче мы можем добиться телесности допустимой области, производя вариацию лишь ограничений общего вида, т.е. рассматривая в качестве ограничений σ -задачи систему неравенств

$$b_i - \sigma \leq (a_i, x) \leq b_i + \sigma, \quad i = 1, \ldots, m,.$$
$$x \geq 0.$$

Сходимость (ε_k, μ_k)-стационарных точек к оптимуму (по функционалу) в этом случае можно гарантировать, если параметры σ_k и ε_k связаны условием $\varepsilon_k < \sigma_k/(R+1)$ при $R = max\,|(a_i, e)|$, $e = (1, \ldots, 1)$. Доказательство этого факта проводится по той же схеме, что и доказательство приведенной теоремы. Отличие будет лишь в построении векторов v^k. Для фиксированного k вектор v^k должен быть внутренним для конуса, определяемого системой неравенств:

$$(a_i, h) \leq 0, \quad i \in I_+^{\sigma_k}(x^k, \varepsilon_k),$$
$$(a_i, h) \geq 0, \quad i \in I_-^{\sigma_k}(x^k, \varepsilon_k),$$
$$h_j > 0, \quad j \in J(x^k, \varepsilon_k),$$

где h_j - компоненты вектора h и

$$I_+^{\sigma_k}(x^k, \varepsilon_k) = \{i \mid b_i + \sigma_k - \varepsilon_k \leq (a_i, x^k) \leq b_i + \sigma_k\},$$
$$I_-^{\sigma_k}(x^k, \varepsilon_k) = \{i \mid b_i - \sigma_k \leq (a_i, x^k) \leq b_i - \sigma_k + \varepsilon_k\},$$
$$J(x^k, \varepsilon_k) = \{j \mid 0 \leq x_j^k \leq \varepsilon_k\}.$$

Несложно проверить, что требуемым свойством будет обладать вектор $v^k = w^k - x^k$, где w^k строится следующим образом: взяв произвольную допустимую (т.е. удовлетворяющую (I6,I7)) точку x^0 и число $t_k \in (\varepsilon_k, (\sigma_k - \varepsilon_k)/R)$, полагаем $w_j^k = x_j^0 + t_k$. Заключительная часть доказательства при этом даже упростится ввиду отсутствия нелинейных ограничений.

Литература

1. ЗОЙТЕНДЕЙК Г. Методы возможных направлений. М., ИЛ, 1963.
2. ДЕМЬЯНОВ В.Ф., МАЛОЗЁМОВ В.Н. Введение в минимакс, М., "Наука", 1972.
3. РУБИНШТЕЙН Г.Ш. Двойственные экстремальные задачи.-"Докл. АН СССР", 1963, т.152, № 2, с. 288-291.
4. M.Frank, P.Wolfe. An algorithm for quadratic programming. Naval Res. Log. Quart., 3(1956), 95-110.

Поступила в ред.-изд. отд.
31. XII. 1975 г.