УДК 512.25/26

К ВОПРОСУ О РЕАЛИЗАЦИИ МЕТОДА ВОЗМОЖНЫХ НАПРАВЛЕНИЙ

А.А. Каплан

Трудоемкость методов возможных направлений связана с тем обстоятельством, что для определения подходящего направления возрастания функции (в задаче на максимум) на каждом шаге процесса приходится решать задачу линейного или квадратичного программирования. Комбинирование этих методов с методами, в которых поиск экстремума ведется с использованием системы фиксированных направлений сдвига, имеет целью сокращение числа указанных вспомогательных задач.

Ниже предлагается алгоритм решения задачи выпуклого программирования, в котором отыскание экстремума осуществляется посредством покоординатного спуска[ж)] с переходом к методу возможных направлений на тех шагах, когда эффективный покоординатный спуск неосуществим. Показана сходимость приведенного алгоритма, частично исследован вычислительный аспект.

Рассмотрим задачу максимизации вогнутой функции f на выпуклом ограниченном множестве Ω n-мерного пространства R^n, задаваемом системой неравенств

$$g^j(x) \geq 0 \ , \ j \in J = \{1,2,\dots,m\},$$

с вогнутыми функциями g^j. Будем предполагать, что функции

ж) Известно, что метод покоординатного спуска в чистом виде, как правило, непригоден для решения задач с ограничениями.

f и $g^j (j \in \mathcal{J})$ заданы на всем R^n, непрерывно дифференцируемы и для указанной задачи выполнено условие Слейтера : существует точка x^* такая, что $g^j(x^*) > 0$ для всех $j \in \mathcal{J}$.

Допустим, что к началу процесса мы имеем некоторую точку $x^0 \in \Omega$ и положительное число δ_0. Пусть в результате K-того шага получены $x^K \in \Omega$ и $\delta_K > 0$. Обозначим $\mathcal{J}(\delta_K) = \{j : g^j(x^K) \leq \delta_K\}$.

В процессе отыскания точки x^{K+1} на ($K+1$)-ом шаге решается одна из двух следующих задач.

ЗАДАЧА G. Максимизировать σ при ограничениях:

$$(\nabla f(x^K), s) \geq \sigma ;$$
$$(\nabla g^j(x^K), s) \geq \sigma , \quad j \in \mathcal{J}(\delta_K);$$
$$s \in E.$$

Здесь $E = \{\pm e^i : i = 1, 2, \ldots, n\}$, $e^i = (0, 0, \ldots, \overset{i}{1}, 0, \ldots, 0)$.

ЗАДАЧА Z. Максимизировать σ при ограничениях:

$$(\nabla f(x^K), s) \geq \sigma ;$$
$$(\nabla g^j(x^K), s) \geq \sigma , \quad j \in \mathcal{J}(\delta_K);$$
$$s \in S.$$

Здесь $S = \{\xi \in R^n : (\xi, \xi) \leq 1\}$ либо $S = \{\xi \in R^n : |\xi_i| \leq 1, \ i = 1, 2, \ldots, n\}$.

Отметим сразу, что решение задачи G тривиально и оно определяет подходящее координатное направление, если таковое существует (здесь термин "подходящее" имеет тот же смысл, что и у Зойтендейка [I], если только вместо конуса возможных направлений рассматривать множество возможных координатных направлений). Действительно, оптимальному σ соответствует вектор s такой, что

$$\max_{\xi \in E} \cdot \min \left\{ (\nabla f(x^K), \xi) ; (\nabla g^j(x^K), \xi), \ j \in \mathcal{J}(\delta_K) \right\} =$$
$$= \min \left\{ (\nabla f(x^K), s) ; (\nabla g^j(x^K), s), \ j \in \mathcal{J}(\delta_K) \right\},$$

то есть если $\sigma > 0$, $s = \pm e^i$, то

$$\max_{\xi \in E} \min \left\{ (\nabla f(x^K), \xi) ; (\nabla g^j(x^K), \xi), \ j \in \mathcal{J}(\delta_K) \right\} =$$
$$= \min \left\{ \left| \frac{\partial f(x^K)}{\partial x_i} \right| ; \left| \frac{\partial g^j(x^K)}{\partial x_i} \right| , \ j \in \mathcal{J}(\delta_K) \right\},$$

причем числа $\frac{\partial f(x^k)}{\partial x_i}$; $\frac{\partial g^j(x^k)}{\partial x_i}$ ($j \in \mathcal{J}(\delta_k)$) - все одного знака.

Иными словами, для отыскания решения задачи G надо лишь из совокупности тех векторов

$$\left\{ \frac{\partial f(x^k)}{\partial x_i} ; \frac{\partial g^j(x^k)}{\partial x_i} \quad (j \in \mathcal{J}(\delta_k)) \right\},$$

у которых все компоненты одного знака, выбрать вектор[ж)], у которого

$$\min \left\{ \left|\frac{\partial f(x^k)}{\partial x_i}\right|, \left|\frac{\partial g^j(x^k)}{\partial x_i}\right| \quad (j \in \mathcal{J}(\delta_k)) \right\}$$

больше. Соответствующий вектор e^i или $-e^i$ определяет подходящее координатное направление.

Процесс начинается с решения задачи G при данных $x^0 \in \Omega$ и δ_0. В зависимости от того, как был реализован k -тый шаг, на $(k+1)$ -ом шаге возможен один из двух следующих вариантов:

а) Если на k -том шаге полученное значение σ (будем обозначать его σ_k) оказалось не меньше δ_{k-1}, на $(k+1)$ -ом шаге решается задача G.

Если при этом получается, что $\sigma_{k+1} \geq \delta_k$, мы определяем точку

$$x^{k+1} = x^k + \lambda s^k \qquad (!)$$

такую[жж)], что

$$x^{k+1} \in \Omega, \; f(x^{k+1}) = \max_{\lambda > 0} f(x^k + \lambda s^k), \qquad (!!)$$

и полагаем $\delta_{k+1} = \delta_k$. Если же $\sigma_{k+1} < \delta_k$, полагаем $x^{k+1} = x^k$, $\delta_{k+1} = \delta_k$. Через s^k обозначен вектор S, соответствующий σ_{k+1}.

б) Если на k -том шаге имели $\sigma_k < \delta_{k-1}$, на $(k+1)$ -ом шаге решается задача Z. Точка x^{k+1} определяется по формулам (!)-(!!). В случае, когда $\sigma_{k+1} \geq \delta_k$, полагаем $\delta_{k+1} = \delta_k$. Если же $\sigma_{k+1} < \delta_k$, берем $\delta_{k+1} = \frac{1}{2}\delta_k$.

ж) Если таких векторов несколько, можно брать любой из них.

жж) Для определенности можно считать, что точка x^{k+1} соответствует наименьшему значению λ, при котором достигается максимум $f(x^k + \lambda s^k)$.

Важной для практики особенностью приведенного метода является то обстоятельство, что при переходе от решения задачи G к задаче Z (т.е. тогда, когда пользоваться подходящим координатным направлением нецелесообразно) используются все вычисления, проведенные при решении задачи G. Следует также отметить, что в зависимости от того, насколько часто используются в процессе решения координатные направления, можно в определенной степени судить о том, удачен ли выбранный способ дробления δ_k, и варьировать выбор δ_k.

Представляются целесообразными следующие два приема, направленные на уменьшение трудоемкости процесса, эффективность которых, вероятно, можно проверить лишь экспериментально. Первый прием, тесно связанный с приведенным в [I] способом устранения так называемого зигзагообразного движения – AZ-2. Если в процессе решения задачи некоторый номер j часто "попадает" в множества $J(\delta_k)$, причем ограничение $g^j(x) \geqslant 0$ мешает осуществлению покоординатного спуска, следует временно ввести дополнительное ограничение, препятствующее попаданию точек x^k в множества $\{x : g^j(x) \leqslant \delta_k\}$. Оно может быть очевидным образом введено как в задачу Z, так и в задачу G. Накопленные к $(\ell+1)$-ому шагу дополнительные ограничения должны быть отброшены, если на ℓ-том шаге решалась задача Z и полученное при этом значение σ оказалось меньше $\delta_{\ell-1}$. Вероятно, во избежание роста размерности вспомогательных задач следует отбрасывать данное дополнительное ограничение и в том случае, если оно в течение ряда последовательных шагов оказывается несущественным.

Второй прием, которым предлагается пользоваться в завершающей стадии решения задачи, состоит в периодическом изменении системы координат. Если, начиная с некоторого шага K', до K-того шага мы движемся преимущественно по градиентным направлениям, новую систему координат имеет смысл построить так, чтобы некоторые из новых координатных направлений были близкими к определенным на этом интервале градиентным направлениям. Для этого следует из множества векторов S^ℓ, $\ell = K', \ldots, K-1$, выделить подмножества $L_1, \ldots, L_T$ так, что $\frac{(S^i, S^j)}{\|S^i\| \|S^j\|} \geqslant$
$\geqslant 1 - \frac{\delta_{K'}}{2}$, если $S^i, S^j \in L_t$ при некотором t;
$\frac{|(S^i, S^j)|}{\|S^i\| \cdot \|S^j\|} \leqslant \frac{\delta_{K'}}{2}$, если $S^i \in L_p$, $S^j \in L_{p'}$, $p \neq p'$

(Ясно, что в таком выделении допускается значительный произвол. Следует отметить, что T может равняться I. В этом случае второе условие теряет смысл.)

Рассматривая затем лишь те множества L_t, которые содержат значительное число элементов*), часть новых ортов $\bar{e}^t$ нужно выбрать так, чтобы

$$\frac{(s^i, \bar{e}^t)}{\|s^i\| \cdot \|\bar{e}^t\|} \geqslant 1-\delta_k \qquad \text{при} \quad s^i \in L_t^*;$$

$$(\bar{e}^t, \bar{e}^\tau)=0 \quad \text{при} \quad t \neq \tau ,$$

а остальные – произвольным образом, соблюдая лишь условия ортогональности.

Отметим, что четкая обоснованная формализация описанных приемов, уточняющая употребленные выше обороты "ряд последовательных шагов", "множества L_t, которые содержат значительное число элементов" и т.п., может быть проведена лишь на основе больших экспериментальных расчетов.

Следует иметь в виду, что применение данного алгоритма в более общем случае, когда в исходной задаче имеются линейные ограничения в форме равенств**), требует предварительного исключения части переменных. Если это исключение не будет выполнено при наличии, скажем, одного ограничения вида $\sum_{i=1}^{n} a_i x_i = b_i$ $(a_i \neq 0,\ i=1,2,\ldots,n)$, покоординатный сдвиг, очевидно, нельзя осуществить ни в одной точке множества Ω.

Приводимое ниже доказательство сходимости описанного алгоритма использует без сколь-нибудь существенных изменений обычную схему доказательства сходимости "чистого" метода возможных направлений (см., например, [1], [2]).

Покажем вначале, что в процессе решения задачи $\delta_k \to 0$ при $k \to \infty$. Предполагая противное, можно утверждать, что, начиная с некоторого K, имеем $\delta_k = \delta > 0$ при $k \geqslant K$.

---

*) При стабилизации процесса решения в его заключительной стадии существование таких множеств весьма вероятно. В дальнейшем мы будем помечать их звездочкой.

**) Вместо условия Слейтера можно потребовать, чтобы существовала допустимая точка, "слейтеровская" относительно ограничений – неравенств.

По смыслу алгоритма это означает, что при $k \geq K$

$$(\nabla f(x^k), s^k) \geq \delta;$$
$$(\nabla g^j(x^k), s^k) \geq \delta, \qquad j \in \mathcal{J}(\delta_k);$$
$$s^k \in S.$$

Учитывая равномерную непрерывность функций $(\nabla f(x), s)$, $(\nabla g^j(x), s)$ $(j \in \mathcal{J})$ на множестве $\Omega \times S$ можно указать $\bar\lambda > 0$ такое, что при $k \geq K$

$$(\nabla f(x^k + \bar\lambda s^k), s^k) \geq \frac{\delta}{2};$$
$$(\nabla g^j(x^k + \bar\lambda s^k), s^k) \geq 0, \quad j \in \mathcal{J}(\delta_k);$$
$$|(\nabla g^j(x^k + \bar\lambda s^k), \bar\lambda s^k)| \leq \delta.$$

Отсюда ввиду вогнутости функций f, $g^j (j \in \mathcal{J})$

$$f(x^k + \bar\lambda s^k) - f(x^k) \geq (\nabla f(x^k + \bar\lambda s^k), \bar\lambda s^k) \geq \frac{\bar\lambda \delta}{2};$$
$$g^j(x^k + \bar\lambda s^k) - g^j(x^k) \geq 0, \quad j \in \mathcal{J}(\delta_k);$$
$$g^j(x^k + \bar\lambda s^k) \geq g^j(x^k) + (\nabla g^j(x^k + \bar\lambda s^k), \bar\lambda s^k) \geq$$
$$\geq \delta - |(\nabla g^j(x^k + \bar\lambda s^k), \bar\lambda s^k)| \geq 0, \quad j \notin \mathcal{J}(\delta_k).$$

Следовательно, при любом $k \geq K$ точка $x^k + \bar\lambda s^k$ является допустимой и

$$f(x^{k+1}) - f(x^k) \geq f(x^k + \bar\lambda s^k) - f(x^k) \geq \frac{\delta \bar\lambda}{2},$$

то есть $f(x^k) \to +\infty$ при $k \to \infty$, что невозможно.

Остается показать, что $\lim\limits_{k\to\infty} f(x^k) = \max\limits_{x \in \Omega} f(x)$ (точнее, нетривиальное неравенство $\lim\limits_{k\to\infty} f(x^k) \geq \max\limits_{x\in\Omega} f(x)$). Допустим противное: $\lim\limits_{k\to\infty} f(x^k) < \max\limits_{x\in\Omega} f(x)$. Пусть k_μ $(\mu = 1, 2, \dots)$ – номера шагов, на которых произошло изменение δ_k, такие, что последовательность x^{k_μ} сходится к некоторой точке $\bar x$. Для дальнейших рассуждений существенно, что по структуре алгоритма на каждом шаге k_μ решалась вспомогательная задача $\mathcal{Z}$.

Так как $f(\bar x) < \max\limits_{x\in\Omega} f(x)$, найдется вектор $\bar s \in S$, для которого

$$(\nabla f(\bar x), \bar s) > 0;$$
$$(\nabla g^j(\bar x), \bar s) > 0, \quad j \in \bar{\mathcal{J}} = \{j \in \mathcal{J} : g^j(\bar x) = 0\}.$$

Положим

$$\bar{\sigma} = \frac{1}{2} \min \left\{ (\nabla f(\bar{x}), \bar{s}); (\nabla g^j(\bar{x}), \bar{s}) \ (j \in \bar{J}); \ g^j(\bar{x}) (j \in J \setminus \bar{J}) \right\}.$$

Тогда при достаточно больших μ

$$(\nabla f(x^{k_\mu}), \bar{s}) > \frac{\bar{\sigma}}{2};$$

$$(\nabla g^j(x^{k_\mu}), \bar{s}) > \frac{\bar{\sigma}}{2} \quad (j \in \bar{J}),$$

причем $J(\delta_{k_\mu}) \cap (J \setminus \bar{J}) = \emptyset$, т.е. $J(\delta_{k_\mu}) \subset \bar{J}$. Последнее очевидно, если $\delta_{k_\mu} \le \bar{\sigma}$. Отсюда следует, что при достаточно больших μ имеем $\sigma_{k_\mu} > \frac{\bar{\sigma}}{2}$, а это невозможно, т.к. σ_{k_μ} стремится к 0 вместе с δ_{k_μ} .

Нетрудно видеть, что приведенное доказательство полностью сохраняется, если рассматривать алгоритм с учетом описанных выше модификаций.

Литература

1. Зойтендейк Г. Методы возможных направлений, ИЛ,М., 1963.
2. Зуховицкий С.И., Авдеева Л.И. Линейное и выпуклое программирование, Наука, М., 1964.

Поступила в редакцию
21.I. 1972 г.