

УДК 519.87:656.7

МОДИФИЦИРОВАННЫЙ МЕТОД ПРОЕКЦИИ ГРАДИЕНТА В ЛИНЕЙНОМ ПРОГРАММИРОВАНИИ

¹Безряков В.В., ²Солодухин В.А., ³Цибизова Т.Ю.

¹Управление государственного надзора за деятельностью в гражданской авиации
Федеральной службы по надзору в сфере транспорта, Москва, e-mail: bezryakov@list.ru;
²ФГБОУ ВО «Санкт-Петербургский государственный университет гражданской авиации»,
Санкт-Петербург, e-mail: SolodukhinVA@yandex.ru;
³ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)», Москва, e-mail: mumc@bmstu.ru

Необходимость решения прикладных задач в реальном масштабе времени определяет интерес к разработке методов оптимизации с гарантийными оценками трудоемкости вычислений, а также дает возможности повышения эффективности известных методов путем их модификации. Методы симплексного типа остаются по-прежнему основными для решения задач линейного программирования, в связи с этим в теории и практике решения задач математического программирования относительно мало уделяется внимания исследованию вычислительной сложности решения задач линейного программирования другими методами. Целью данной работы является обоснование модификации метода проекции градиента, эквивалентной методу наискорейшего спуска, и дальнейшей модификации с целью ускорения сходимости модифицированного метода к решению задачи линейного программирования. Показано, что метод проекции градиента при изменении правила выбора проектирующей матрицы на очередной итерации позволяет определить наилучшее направление движения по границе множества допустимых решений, которое совпадает с направлением наискорейшего спуска. Представлены эффективные допустимые направления, являющиеся существенными при реализации модифицированного метода проекции градиента. Сделан вывод о том, что предложенный подход синтеза методов градиентного спуска и симплексного типа в линейном программировании при незначительном увеличении вычислительной сложности компенсируется существенным уменьшением общего количества итераций.

Ключевые слова: линейное программирование, метод проекции градиента, модификация, сходимость, алгоритм

MODIFIED METHOD OF GRADIENT PROJECTION IN LINEAR PROGRAMMING

¹Bezryakov V.V., ²Solodukhin V.A., ³Tsibizova T.Yu.

¹Department of State Supervision of Civil Aviation Activities of the Federal Service
for Supervision in the Sphere of Transport, Moscow, e-mail: bezryakov@list.ru;
²Federal State Budget Educational Institution Higher Education
«St. Petersburg State University of Civil Aviation», St. Petersburg, e-mail: SolodukhinVA@yandex.ru;
³Federal State Budgetary Education Institution of Higher Education
«Bauman Moscow State Technical University» (BMSTU), Moscow, e-mail: mumc@bmstu.ru

The need to solve applied problems in real time determines the interest in the development of optimization methods with warranty estimates of the laboriousness of computations, and also look for ways to improve the efficiency of known methods by modifying them. Simplex methods remain basic for solving linear programming problems, therefore, in theory and practice of solving problems of mathematical programming, relatively little attention is paid to the study of the computational complexity of solving linear programming problems by other methods. The purpose of this paper is to substantiate the modification of the gradient projection method, equivalent to the steepest descent method, and further modification in order to accelerate the convergence of the modified method to solving the linear programming problem. It is shown that the gradient projection method when changing the selection rule of the projecting matrix to the next iteration allows us to determine the best direction of motion along the boundary of the set of admissible solutions, which coincides with the direction of steepest descent. Effective permissible directions are presented that are essential for implementing the modified method of gradient projection. The conclusion is made that the proposed approach to synthesis of gradient descent and simplex type methods in linear programming with an insignificant increase in computational complexity is compensated by a significant decrease in the total number of iterations.

Keywords: linear programming, gradient projection method, modification, convergence, algorithm

Потребности решения прикладных задач в реальном масштабе времени обуславливают интерес к разработке методов оптимизации с гарантийными оценками трудоемкости вычислений, а также к поиску возможностей повышения эффективности известных методов путем их модификации [1].

Применительно к задачам линейного программирования (ЛП) в качестве первой успешной попытки решения проблемы можно рассматривать результаты работы [2], распространенные в дальнейшем на некоторые классы задач нелинейного программирования. Несмотря на это, мето-

ды симплексного типа (МСТ) остаются по-прежнему основными для решения задач ЛП [3, 4], благодаря их более высокой статистической эффективности практически на всем многообразии прикладных задач, хотя и известны «искусственные» примеры, когда применение МСТ равносильно полному перебору опорных планов задачи [5, 6].

По-видимому, в связи с этим в теории и практике решения задач математического программирования относительно мало уделяется внимания исследованию вычислительной сложности решения задач ЛП другими методами.

В частности, еще в [7] отмечалось, что метод проекции градиента (МПП) часто требует для решения задачи ЛП меньшего числа итераций по сравнению с МСТ. Интуитивно высокая эффективность МПП объясняется тем, что при его реализации переход между точками на очередной итерации может осуществляться по внутренней части многогранника допустимых решений или по его граням, в то время как при МСТ переходы всегда происходят по ребру многогранника.

Однако в известных алгоритмах МПП эти потенциальные возможности реализованы не полностью. Основной их недостаток состоит в том, что размерность пространства (граней), в котором осуществляется движение, на каждой итерации уменьшается на единицу до получения опорного плана задачи ЛП. Далее вычислительный процесс становится эквивалентным использованию двойственного симплекс-метода [8]. Это приводит в результате к таким же сложностям получения оценок эффективности МПП, как и при рассмотрении МСТ.

Целью настоящей работы является обоснование модификации МПП, эквивалентной методу наискорейшего спуска, и дальнейшей модификации с целью ускорения сходимости модифицированного метода к решению задачи ЛП.

Модифицированный метод проекции градиента

Задачу ЛП в R^n без нарушения общности рассмотрения можно представить в следующем виде:

$$\max f(x) = c^T x \quad (1)$$

при ограничениях

$$\begin{aligned} a_i^T x &\leq b_i, i \in I = 1, \dots, m; \\ -x_j &\leq 0, j \in J = 1, \dots, n_1 \leq n, \end{aligned} \quad (2)$$

где R^n – n -мерное арифметическое пространство, $f(x)$ – целевая функция; c – матрица коэффициентов a_i и b_i ; x – матрица переменных.

Будем считать, что условия (2) линейно независимы.

Для приведения к такому виду общей задачи ЛП с линейно-независимыми условиями вида $a_i^T x = b_i, i = m + 1, \dots, m + m_1$ достаточно разрешить эту систему уравнений относительно любых m_1 переменных и исключить их из системы условий (2).

Для решения задачи ЛП методом ПГ на каждой итерации определяется [9, 10]: $N^k (m_k \times n)$ – матрица, образованная транспонированными векторами-градиентами ограничений, которые по предположению на протяжении k -й итерации являются существенными, то есть выполняются как строгие равенства; $R^k (n \times n)$ – матрица проектирования; r^k – n -вектор направления движения; δ^k – m_k -вектор двойственных переменных, соответствующих существенным ограничениям.

При этом имеют место следующие соотношения

$$r^k = P^k c, \delta^k = Q^k c, P^k = E - (N^k)^T Q^k, \quad (3)$$

где $Q^k = [N^k (N^k)^T]^{-1} N^k = G^k N^k$.

Вычисления начинаются с допустимого решения системы (2) x^0 .

Значение вектора переменных после выполнения k итераций определяется выражением

$$x^{k+1} = x^k + \lambda^k r^k, \quad (4)$$

где $\lambda^k \geq 0$ – величина максимально допустимого шага в направлении r^k .

Условия оптимальности решения x^k имеют вид

$$r^k = 0, \delta_s^k \geq 0, s \in S_k = I_k \cup J_k, \quad (5)$$

где $I_k \subset I, J_k \subset J$ – соответственно множества индексов существенных ограничений на k -й итерации.

В классической процедуре МПП анализ существенности ограничений производится только в том случае, когда выполняется первое из необходимых условий оптимальности (5). При этом если x^k не оптимально, одно из ограничений, для которого $\delta_s^k < 0$ переводится в несущественные (из матрицы N^k исключается строка) [11]. Затем осуществляется $k + 1$ -я итерация.

Суть предлагаемой модификации МПП состоит в том, что для перехода от x^k к x^{k+1} определяется наилучшее из возможных направление движения по границе множества допустимых решений сокращенной задачи, которое далее используется для определения допустимого направления движения.

Один из возможных алгоритмов реализации модификации МПП состоит из по-

следовательности итераций, на каждой из которых ($k \geq 0$) необходимо:

- 1) определить λ^k и x^{k+1} ;
- 2) сформировать N^{k+1} путем присоединения к матрице N^k строки l^T , соответствующей вектору градиента ограничения, определяющего λ^k , определить r^{k+1} , δ^{k+1} ;
- 3) проверить выполнение условий оптимальности.

Из матрицы N^{k+1} последовательно исключить все строки, соответствующие несущественным ограничениям, для которых $\delta_s^{k+1} < 0$ ($s \in S_k$), и скорректировать значения r^{k+1} , δ^{k+1} , перейти к п.1.

Обоснование алгоритма модификации МПГ

Исследуем две вспомогательные задачи оптимизации, решение которых определяют направление движения: наилучшее в градиентных методах (задача А) и в методе ПГ (задача Б).

Задача А имеет вид [10]

$$\max z = \sigma^T r \quad (6)$$

при ограничениях

$$a_i^T r \leq 0, i \in I_k; r_j \geq 0, j \in J_k; r^T r \leq 1. \quad (7)$$

Задача Б отличается от задачи А только характером ограничений, которые имеют вид:

$$a_i^T r = 0, i \in I_k = I_k; -r_j = 0, j \in J_k; r^T r \leq 1. \quad (8)$$

Если задача Б разрешима, то ее решение \hat{r}^k может быть найдено по алгоритму МПГ. При этом выполнение последнего из условий (8) обеспечивается нормировкой (путем деления \hat{r}^k на $\sqrt{(\hat{r}^k)^T \hat{r}^k}$). Оптимальность \hat{r}^k для задачи Б следует непосредственно из определения проекции вектора на множество.

Понятно, что решения задач А и Б отличаются в том случае, когда среди ограничений $a_i^T r \leq 0, i \in I_k$ и $r_j \geq 0, j \in J_k$ есть несущественные для выбора наилучшего направления. Обозначим $I_k^< = \{i / a_i^T r^k < 0\}, J_k^> = \{j / r_j^k > 0\}$.

По теореме Куна – Таккера необходимые и достаточные условия оптимальности решения r задач А и Б состоят в существовании векторов (α, α_r) и $r_s \geq 0$ таких, что [10]:

$$N^k r + r_s = 0; (N_k)^T \alpha + 2\alpha_r r = \sigma, \alpha^T r_s = 0. \quad (9)$$

При этом для задачи Б первое условие имеет вид $N^k r^k = 0$, а последнее удовлетворяется при любых α , так как $r = 0$.

Учитывая выражения для P^k и δ^k , легко видеть, что переменные $\alpha = \delta^k$ удовлетворяют второму условию из (9) при $\alpha_r = 0,5$. Следовательно, δ^k являются множителями Лагранжа для задачи Б.

Представим линейные ограничения (8) в виде $N^k r = \beta$.

Известно, что $\frac{\partial z}{\partial \beta_s} = \delta_s^k, s \in S_k$.

Отсюда следует, что уменьшение любого β_s ($s \in S_k^< = \{s | \delta_s^k < 0\}$) приводит к увеличению значения целевой функции. Поэтому соответствующее ограничение является несущественным и его закрепление может быть снято. Исключив его из системы (8), приходим к задаче Б меньшей размерности.

Если имеются другие $\delta_s^k < 0$, процесс исключения может быть повторен. В результате получим задачу Б для которой все δ_s^k неотрицательны. Решение последней задачи совпадает с решением задачи А.

В противоположном случае среди ограничений (8) должно найтись ограничение ($i \in I_k$) или ($j \in J_k$), которое в задаче А несущественно. То есть, по крайней мере, для одного из этих ограничений должно быть $\delta_s^k < 0$.

Геометрически увеличение значения целевой функции в задаче Б или А означает увеличение значения косинуса угла между вектором-градиентом целевой функции и направлением движения, определяемым текущим решением $r \left(\cos = \frac{\sigma^T r}{\sigma^T \sigma} \right)$ и, следо-

вательно, увеличение длины проекции вектора-градиента целевой функции на множество допустимых решений в задаче Б.

В силу единственности проекции вектора на множество – наилучшее направление также единственно. Следовательно, процесс исключения несущественных ограничений приводит к одной и той же задаче независимо от порядка их рассмотрения. При этом процесс может быть начат с любого ограничения, для которого $\delta_s^k < 0$. Отсюда следует, что все такие ограничения могут быть исключены одновременно.

Таким образом, МПГ при изменении правила выбора проектирующей матрицы на очередной итерации позволяет определить наилучшее направление движения по границе множества допустимых решений, которое совпадает с направлением наискорейшего спуска.

Определение эффективных допустимых направлений по данным модифицированного МПГ

Эффективность модифицированного МПГ в значительной степени определяется структурой множества допустимых решений и прежде всего количеством ограничений (2), которые могут оказаться существенными при реализации МПГ. Сокращение количества исследуемых гиперплоскостей

может быть достигнуто при использовании полученных результатов для построения допустимых направлений движения во внутренней области МДР (максимально достижимым кодовым расстоянием) [12], которые приводят на каждом шаге к возможно наибольшему приращению целевой функции.

Пусть выполнено k итераций с использованием таких допустимых направлений и определено множество $X = \{xs | s = 0, \dots, k\}$.

На $(k + 1)$ -й итерации очередная точка определяется в виде

$$x^{k+1} = x^s + \mu(x^k - x^s) + \lambda r^k, \quad xs \in X, \mu, \lambda \geq 0. \quad (10)$$

Очевидно, найдутся такие значения μ, λ , что x^{k+1} , определяемое таким образом, является допустимой точкой. В частности, при $\mu = 1, \lambda = \lambda k$ получаем точку, соответствующую МПГ.

Для определения направления, приводящего к наибольшему приращению целевой функции, рассмотрим задачу

$$\max_{x^s, \mu, \lambda} \sigma = c^T x^{k+1}(x^s, \mu, \lambda) \quad (11)$$

при ограничениях

$$a_i^T x^{k+1}(x^s, \mu, \lambda) \leq b_i, i \in I;$$

$$x^{k+1}(x^s, \mu, \lambda) \leq 0, j \in J.$$

Возможны два случая решения этой задачи.

В первом случае $\mu^* = 1, \lambda^* = \lambda k$. Это означает, что нельзя найти допустимое направление выбранного вида, приводящее к большему значению целевой функции по сравнению с модифицированным МПГ. При этом выделяется одно новое существенное ограничение.

Во втором случае $\mu^* < 1, \lambda^* > \lambda k$. При этом выделяются два новых существенных ограничения.

Пусть при $\mu = 1, \lambda = \lambda k$ существенным становится ограничение с номером q . Тогда для всех $xs \in X$ имеем $a_q^T x^{k+1}(x^s, 1, \lambda^k) = bq$.

Обозначим $\lambda = 1 - \varepsilon, \mu = \lambda k + \delta$. Очевидно существуют такие $\varepsilon, \delta > 0$, что имеет место

$$a_q^T x^{k+1}(x^s, 1 - \varepsilon, \lambda^k + \delta) = bq,$$

$$cT x^{k+1}(x^s, 1 - \varepsilon, \lambda^k + \delta) > cT x^{k+1}(x^s, 1, \lambda^k).$$

Для таких ε, δ получаем

$$a_q^T (-\varepsilon(xk - xs) + \delta rk) = 0,$$

$$cT (-\varepsilon(xk - xs) + \delta rk) > 0.$$

Исключая из равенства одну из переменных ε, δ и подставляя ее значение в неравенство, получаем, что при фиксированных значениях ε, δ наибольшее приращение целевой функции при движении в q -й гиперплоскости достигается тогда, когда

$$x^s = \operatorname{argmax}_{x^i \in X} \frac{c^T(x^k - x^i)}{a_q^T(x^k - x^i)}.$$

Выбор такого направления обеспечивает наилучшее приближение к направлению движения в q -й гиперплоскости при использовании модифицированного МПГ и поэтому может считаться наиболее целесообразным.

Задача (11) при фиксированном xs сводится к задаче линейного программирования с двумя переменными. Ее решение вычислительных трудностей не представляет.

Решение исходной задачи сводится, таким образом, к решению последовательно задачи вида (11).

При этом, если $\mu^* = 1$ определение r^k и проверка условий оптимальности осуществляются по алгоритму модифицированного МПГ. При $\mu^* < 1$ матрица M^{k+1} формируется по двум ограничениям, определяющим решение задачи (11) и выражения (3) для определения r^k, δ^k можно записать в явном виде.

Заключение

Сопоставление детального анализа оценок вычислительной сложности отдельных итераций и результатов многократной практической апробации алгоритма, реализующего предложенный подход, показывает, что незначительное увеличение вычислительной сложности, связанное с решением на каждой итерации задачи вида (11), компенсируется существенным уменьшением общего количества итераций по сравнению с алгоритмом МПГ. Предложенная схема формирования допустимых направлений не претендует на получение безусловно точного наиболее эффективного метода, а является только примером возможного синтеза методов градиентного спуска и симплексного типа в линейном программировании.

Список литературы

1. Фам С.Ф., Цибизова Т.Ю. Методы построения математических моделей: генетические алгоритмы // Достижения вузовской науки. Труды международной научно-практической конференции. – М.: Изд-во МГОУ, 2014. – С. 158–162.
2. Хачиян Л.Г. Полиномиальные алгоритмы в линейном программировании // Журнал вычислительной математики и математической физики. – 1980. – Т. 20, № 1. – С. 51–68.
3. Батенков А.А., Батенков К.А., Яковлев А.В. Разработка алгоритма декодирования Евклидова кода на основе

симплексного базиса // Информационные системы и технологии. – 2010. – № 3 (59). – С. 134–142.

4. Саушев А.В. Сеточный метод построения областей работоспособности технических объектов на основе алгоритма симплексного поиска // Журнал университета водных коммуникаций. – 2010. – № 1. – С. 58–69.

5. Моисеев О.В., Яковлев А.В. Алгоритм приема большого ансамбля сложных сигналов на основе симплексного базиса // Телекоммуникации. – 2012. – № 4. – С. 13–19.

6. Цибизова Т.Ю., Карпунин А.А. Применение метода анализа иерархий в оценке качества процессов управления // Современные проблемы науки и образования. – 2015. – № 2.

7. Черняев Ю.А. Сходимость метода проекции градиента для одного класса невыпуклых задач математического программирования // Известия высших учебных заведений. Математика. – 2005. – № 12. – С. 76–79.

8. Безряков В.В., Крыжановский Г.А., Солодухин В.А. Прямые и обратные задачи оптимизации управления потоками воздушного движения в районе аэродрома // Научный вестник Московского государствен-

ного технического университета гражданской авиации. – 2011. – № 171. – С. 109–113.

9. Тимофеев П.А., Ягудаев Г.Г., Горячкин Б.С., Строганов Д.В. Алгоритм управляемого имитационного процесса в системах массового обслуживания // Прикаспийский журнал: управление и высокие технологии. – 2011. – № 3. – С. 62–70.

10. Мясников Е.А. Нелинейное и динамическое программирование: уч. пособие. – Хабаровск: Изд-во Хабаровская гос. акад. экономики и права, 2011. – 139 с.

11. Перепелкин Е.А. Задача параметрического синтеза многосвязной динамической системы как задача нелинейного программирования // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2008. – № 1. – С. 23–27.

12. Гусев М.И. Метод динамического программирования в задаче построения множеств достижимости нелинейных управляемых систем // Известия Института математики и информатики Удмуртского государственного университета. – 2012. – № 1. – С. 42–43.