АЛГОРИТМЫ ВНУТРЕННИХ ТОЧЕК: ИСТОРИЯ, РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ, ПРИЛОЖЕНИЯ И ПЕРСПЕКТИВЫ[1]

В.И. ЗОРКАЛЬЦЕВ,
Лимнологический институт СО РАН,
Иркутск, Россия
vizork@mail.ru, https://www.zorkaltsev.com

Аннотация. *Рассматриваются семейства прямых и двойственных алгоритмов внутренних точек применительно к решению задач линейного и нелинейного программирования. Излагается история создания и развития алгоритмов. Приводятся новые модификации алгоритмов внутренних точек, содержащие как частный случай разработанные ранее алгоритмы.*

Ключевые слова: *математическое программирование, линейные неравенства, алгоритмы внутренних точек.*

Введение. В докладе рассматриваются алгоритмы решения задач математического программирования, осуществляющих ввод в область допустимых решений и оптимизацию путем итеративного улучшения в множестве векторов, удовлетворяющих ограничениям-неравенствам в строгой форме. Приводятся семейства прямых, двойственных, самосопряженных алгоритмов, полиноминальные алгоритмы оптимизации в конусе пути аналитических центров задачи линейного программирования. Представлены полученные автором результаты по теоретическому обоснованию алгоритмов. В частности выделены подмножества алгоритмов обладающих линейной и сверхлинейной скоростью сходимости, асимтотически не зависящей от исходных данных решаемой задачи. Излагаются результаты экспериментальных исследований вариантов алгоритмов на тестовых примерах и при реализации ряда математических моделей энергетики. Особое внимание уделено перспективам развития алгоритмов внутренних точек.

[1] Работа выполнена при поддержке РФФИ проект № 19-07-00322
и в рамках проекта РАН № 0279-2019-0003.

Исходным импульсом создания рассматриваемых алгоритмов послужила идея 1965 г. Л.В. Канторовича определения двойственных оценок из неоптимального плана задачи линейного программирования путем аппроксимации методом наименьших квадратов условия дополняющей нежесткости. На основе этой идеи был создан в 1967 г. алгоритм И.И. Дикина решения задач линейного и квадратичного программирования. Проведенные И.И. Дикиным совместно с С.М. Анцызом в Институте математики СО АН СССР экспериментальные исследования позволили выявить эффективный вариант алгоритма и показали его хорошие вычислительные перспективы.

В 1972 г. И.И. Дикиным, работавшим в то время уже в Сибирском энергетическом институте СО АН СССР (г. Иркутск) была подготовлена и защищена в Иркутском государственном университете кандидатская диссертация по данному алгоритму. В диссертации небольшого объема (менее 40 страниц) были представлены описание алгоритма, его геометрические интерпретации, результаты экспериментальных исследований и полученные И.И. Дикиным результаты по теоретическому обоснованию. А именно было доказано, что при оптимизации в области допустимых решений невырожденных задач линейного программирования алгоритм с линейной скоростью сходимости приводит к оптимальному решению с минимальным набором активных ограничений. То есть, доказано, что алгоритм приводит к относительно внутренней точке множества оптимальных решений, что является важной особенностью рассматриваемых алгоритмов.

Научным руководителем был Л.В. Канторович. Оппонировали известные российский и украинский ученые В.Л. Макаров и Н.З. Шор. С 70-х годов этот метод активно развивался только российскими математиками (Ю.Г. Евтушенко, В.И. Зоркальцев, В.Г. Жадан) и нашел активное применение при реализации ряда моделей экономки и энергетики [1]. В других странах повышенный интерес к алгоритмам рассматриваемого типа возник после публикации в 1984 году алгоритма Н. Кармаркара, являющегося фактически ухудшенной версией алгоритма И.И. Дикина.

Комбинированный алгоритм. В [1] рассматривался алгоритм, сочетающий в едином вычислительном процессе ввод в область допустимых решений и оптимизацию. Приведем новую модификацию такого комбинированного алгоритма для задачи линейного программирования в стандартной форме и двойственной к ней:

$$c^T x \to \min, Ax=b, x \geq 0; \tag{1}$$

$$b^T u \to \max, g(u) \equiv c - A^T u \geq 0. \tag{2}$$

Заданы матрица A размера $m \times n$, векторы $b \in R^m$, $c \in R^n$. Переменные задач (1), (2) составляют векторы $x \in R^n$, $u \in R^m$.

Стартовой точкой может служить любой вектор x^0 из R^n со всеми положительными компонентами. Например, $x_j^0 = 1$, $j=1,\ldots,n$.

Алгоритм вырабатывает последовательность векторов x^k из R^n также со всеми положительными компонентами, где $k = 0, 1, 2, \ldots$ – номер итерации. На каждой итерации осуществляется следующая последовательность действий.

1. Вычисляем вектор невязок ограничений-равенств задачи (1):

$$r^k = b - Ax^k. \tag{3}$$

2. Определяем вектор положительных весовых коэффициентов d^k из R^n, компоненты которого должны удовлетворять неравенствам

$$\bar{\sigma}(x_j^k) \geq d_j^k \geq \underline{\sigma}(x_j^k), \; j=1,\ldots,n. \tag{4}$$

Здесь $\bar{\sigma}$, $\underline{\sigma}$ некоторые функции от положительного вещественного аргумента такие, что при любом $\alpha > 0$

$$\bar{\sigma}(\alpha) \geq \underline{\sigma}(\alpha) > 0. \tag{5}$$

Требуется также выполнение следующего условия

$$\frac{\bar{\sigma}(\alpha)}{\underline{\sigma}(\tau)} = \mathrm{O}\left(\frac{\alpha}{\tau}\right). \tag{6}$$

Например, можно воспользоваться правилом

$$d_j^k = (x_j^k)^p \tag{7}$$

где $p \geq 1$ – заданный параметр. В этом случае при $\alpha \geq 0$

$$\bar{\sigma}(\alpha) = \underline{\sigma}(\alpha) = \alpha^p. \tag{8}$$

3. Определим вектор-функцию $s^k(\beta)$ со значениями из R^n, зависящую от вещественного параметра β, как решение вспомогательной задачи:

$$\sum_{j=1}^{n} c_j^k(\beta)s_j + \frac{1}{2}\sum_{j=1}^{n}(s_j^2 / d_j^k) \to \min,\ As=r^k. \qquad (9)$$

с вектором переменных $s\in R^n$. Здесь

$$c^k(\beta)=c- \beta y^k, \qquad (10)$$

при

$$y_j^k = 1/ x_j^k,\ j=1,\dots, n. \qquad (11)$$

Для задания вектор-функции $s^k(\beta)$ достаточно вычислить ее значения в двух точках, так как при любом β

$$s^k(\beta)= s^k(0)+ \beta\, s^k(1). \qquad (12)$$

Вычисление вектора $s^k(\beta)$ при двух значениях β сводится к задаче решения двух систем линейных уравнений с одной и той же симметричной неотрицательно определенной матрицей и двумя векторами в правой части системы. В вычислительном отношении такая задача равносильна поиску решения такой же системы с одним вектором в правой части.

Действительно, при любом заданном β

$$s_j^k(\beta) = d_j^k g_j^k(u^k(\beta),\beta),\ \ j=1,\dots, n. \qquad (13)$$

где вектор $u^k(\beta)$ является решением задачи безусловной минимизации квадратичной выпуклой функции относительно вектора переменных $u\in R^m$:

$$\frac{1}{2}\sum_{j=1}^{n} d_j^k (g_j^k(u,\beta))^2 - \sum_{i=1}^{m} r_i^k u_i \to \min. \qquad (14)$$

В (13), (14)

$$g^k(u, \beta)=c^k(\beta)\text{-}A^T u. \qquad (15)$$

Приравняв градиент целевой функции вспомогательной задачи (14) нулевому вектору получим систему линейных уравнений, у которой от параметра β зависит только вектор в правой части. Причем эта зависимость линейная.

В качестве пояснения отметим, что вспомогательная задача (14) является равносильной, симметрично двойственной к исходной вспомогательной задаче (9).

4. Вычисление шага корректировки решения. Пусть γ заданный параметр из открытого интервала (0, 1). Например, γ=0,9. Задано также верхнее значение $\bar{\beta}$ для определяемого по приводимым

ниже правилам значения β. Обозначим λ_k, β_k решение одной из приводимых ниже задач относительно переменных λ и β.

Если $r^k \neq 0$, то решается задача:

$$\lambda \to \max,\ \lambda s^k(\beta)+\gamma x^k \geq 0,\ 1 \geq \lambda,\ \bar{\beta} \geq \beta \geq 0. \qquad (16)$$

Если $r^k=0$, то решается задача:

$$\lambda \sum_{j=1}^{n} c_j s_j^k(\beta) \to \min,\ \lambda s^k(\beta)+\gamma x^k \geq 0,\ \bar{\beta} \geq \beta \geq 0. \qquad (17)$$

В приводимых ниже результатах расчета для решения задач (16), (17) использовался метод золотого сечения. Отметим, что максимальное значение целевой функции задачи (16), рассматриваемое как неявная функция от переменной β, является вогнутой функцией. Целевая функция задачи (17) при β, с которыми она достигает неположительных значений, является выпуклой неявной функцией от β (причем такие значения β составляют интервал с нижней границей равной нулю). Эти два факта обосновывают правомочность использования метода золотого сечения для решения задач (16), (17).

5. Осуществляется итеративный переход:

$$x^{k+1} = x^k + \lambda_k s^k(\beta_k). \qquad (18)$$

Замечания. 1. Из ограничений задач (16), (17) и итеративного перехода (18) следует, что

$$x^{k+1} \geq (1-\gamma)\, x^k. \qquad (19)$$

Поскольку все компоненты вектора x^k положительные и $\gamma<1$, то положительными будут и все компоненты вектора x^{k+1}.

2. Из условия вспомогательной задачи (9) и итеративного перехода (18) следует, что

$$r^{k+1}=(1-\lambda_k)\, r^k. \qquad (20)$$

Этим объясняется почему в (16) величина шага λ_k ограничена сверху единицей. Пока $r^k \neq 0$ приведенный алгоритм осуществляет ввод в область допустимых решений. Абсолютные значения каждой компоненты вектора невязок балансовых ограничений сокращаются в $(1-\lambda_k)$ раз при $\lambda_k \in (0, 1]$. Это будет этап ввода в область допустимых решений задачи (1). Учет целевой функции позволяет получать первое допустимое решение более близким к оптимальному.

При r^k=0 согласно (20), после итеративного перехода невязки балансовых ограничений должны оставаться нулевыми, r^{k+1}=0. На этом этапе будет происходить оптимизация в области допустимых решений, будут выполняться неравенства

$$c^T x^{k+1} < c^T x^k. \tag{21}$$

3. Можно априори зафиксировать значение $\bar{\beta}$=0. Тогда приведенный алгоритм станет одним из вариантов рассматривавшихся ранее алгоритмов метода внутренних точек. В частности, при правиле вычисления весовых коэффициентов (7) для p=2 получим наиболее известный алгоритм внутренних точек, за которым закрепилось название «affine scaling method».

Для вариантов алгоритмов с $\bar{\beta}$=0 применительно к процессу оптимизации в области допустимых решений при предположении о невырожденности задачи (1) доказано, что вырабатываемые последовательности векторов x^k , $x_0^k = u^k(0)$ сходятся не менее чем линейно к относительно внутренним точкам оптимальных решений задач (1) и (2). Причем скорость сходимости двойственных переменных будет быстрее, чем скорость сходимости переменных двойственной задачи: $\dfrac{\left\|u^k - \bar{u}\right\|}{\left\|x^k - \bar{x}\right\|} \to 0$, при k→∞.

Можно отметить, что более быстрая сходимость сначала была выявлена на основе опыта использования алгоритмов внутренних точек при реализации моделей энергетики. Затем было получено теоретическое обоснование. Этот факт делает целесообразным для более быстрого получения решения исходной задачи (1) использование двойственных аналогов изложенных алгоритмов в том числе «dual affine scaling method». Для алгоритмов с весовыми коэффициентами (7) при $p \in [1,\ 3]$ дано доказательство без предположения о невырожденности задачи. Причем для $p \in (1,\ 3]$ было доказано, что скорость сходимости асимтотически не зависит от исходных данных, в том числе, размерности задачи, так как

$$\frac{\left\|x^{k+1} - \bar{x}\right\|}{\left\|x^k - \bar{x}\right\|} \to (1-\gamma), \text{ при } k\to\infty.$$

Исследования показали, что это неверно для алгоритмов с весовыми коэффициентами (7) при p=1. В этом случае линейная скорость сходимости зависит от всех исходных данных задачи.

При доказательствах указанных и других фактов нельзя было использовать стандартную технику обоснования алгоритмов оптимизации. Вместе с тем разработанную технологию обоснования алгоритмов при $\bar{\beta}$ =0 нельзя легко перенести на изложенное здесь обобщение таких алгоритмов.

4. В приводимой ниже таблице представлены осуществленные А.Ю. Филатовым сравнительные расчеты «affine scaling method» и изложенного здесь алгоритма с весовыми коэффициентами (7) при p=2. Отметим, что на каждой итерации в обоих случаях осуществляется примерно один и тот же объем вычислений. Представленные результаты расчетов показывают, что есть смысл в использовании изложенного здесь комбинированного алгоритма. В таблице 1 под термином комбинированный алгоритм понимается алгоритм с $\bar{\beta}$ =1, под термином комбинированный расширенный понимается алгоритм с $\bar{\beta}$ =2.

Для каждой рассмотренной размерности рассматривалось двадцать случайно сгенерированных задач линейного программирования в стандартной форме. В таблице 1 представлены средние арифметические значения числа итераций потребовавшихся для решения задач и среднеквадратические отклонения от средних значений числа итераций.

Таблица 1.

Среднее значение и среднеквадратическое отклонение от среднего значения числа итераций, потребовавшихся для решения задач линейного программирования разных размерностей

Алгоритмы\ Размерность задач	20×40	40×80	100×200	200×500
Аффинно-масштабирующий	it=31,1 σ=6,39	it=33,0 σ=6,42	it=29,1 σ=11,62	it=28,6 σ=2,76
Комбинированный	it=25,1 σ=4,44	it=24,1 σ=4,97	it=23,1 σ=7,99	it=22,6 σ=0,66
Комбинированный расширенный	it=23,6 σ=5,28	it=21,7 σ=2,87	it=23,6 σ=8,49	it=22,5 σ=0,67

Из представленных в таблице 1 результатов видно, что использование изложенных в данной статье комбинированных алгоритмов приводит к сокращению примерно на 25% времени решения задач линейного программирования по сравнению с исходным «аффинно маштабирующем» (affine scaling method), наиболее известным вариантом алгоритма внутренних точек. Можно отметить также более устойчивое поведение по времени счета (более низкое среднеквадратического отклонения от среднего значения числа итераций) изложенных здесь алгоритмов на задачах одной и той же размерности.

Существенный положительный эффект был получен также в вычислительных экспериментах (проведенных также А.Ю. Филатовым) решения нелинейных систем уравнений и неравенств. В таком виде представляется задача расчета допустимых режимов электроэнергетических систем. Использование алгоритмов внутренних точек сопровождалось процедурами итеративной линеаризации. Рассматривались предоставленные О.Н. Войтовым четыре схемы электроэнергетических систем (с 152 узлами, с 118 узлами, 207 узлами и 211 узлами). Во всех четырех случаях изложенным здесь алгоритмом потребовалось существенно меньше (на 15-50%) числа обращений матрицы (количества требовавшихся решений задачи (14)). При этом только в одном примере потребовалось больше итераций линеаризации, каждая из которых требует некоторого дополнительного времени, но не сопоставимого с временем обращения матрицы методом квадратного корня. Этим методом решалась вспомогательная задача (14) в примерах, представленных в таблице 1 и при поиске допустимых режимов электроэнергетических систем.

Литература.

1. Дикин И.И., Зоркальцев В.И. Итеративное решение задач математического программирования: алгоритмы метода внутренних точек. – Новосибирск: Наука, 1980, 144 с.