УДК 519.8

И.В. СЕРГИЕНКО, П.И. СТЕЦЮК

О ТРЕХ НАУЧНЫХ ИДЕЯХ Н.З. ШОРА

Ключевые слова: *недифференцируемая оптимизация, субградиентный метод, оператор растяжения пространства, метод эллипсоидов, r-алгоритм, экстремальная квадратичная задача, двойственная оценка, функционально избыточные ограничения.*

ВВЕДЕНИЕ

В этом году исполняется 50 лет со времени опубликования Н.З. Шором работы [1], к которой восходит идея субградиентного метода минимизации негладких выпуклых функций (1962). Тогда Науму Зуселевичу исполнилось всего 25 лет. Для решения двойственных задач к транспортным задачам в матричной и сетевой формах он предложил использовать метод спуска в пространстве потенциалов. Впоследствии этот метод получил название метода обобщенного градиентного спуска (ОГС). Метод ОГС сначала применялся для минимизации кусочно-гладких выпуклых функций, появляющихся при решении транспортных и транспортно-производственных задач, а затем для класса произвольных выпуклых функций и задач выпуклого программирования. Последующие усовершенствования метода ОГС дали ряд численных методов оптимизации негладких (недифференцируемых) функций, которые существенно повлияли на развитие линейного, нелинейного, дискретного и стохастического программирования.

Настоящая статья посвящена 75-летию со дня рождения Наума Зуселевича Шора. В ней акцент сделан на трех его важнейших идеях: обобщенном градиентном спуске (1962), использовании линейных неортогональных преобразований пространства для улучшения обусловленности овражных функций (1969), двойственном подходе к получению и уточнению оценок в невыпуклых квадратичных моделях (1985). На основе первых двух идей получен ряд фундаментальных результатов в негладкой оптимизации [2], которая как самостоятельный раздел математического программирования сформировалась в конце прошлого века. На основе третьей идеи получен ряд важных результатов в многоэкстремальной оптимизации и теории графов [3].

Предложенный Н.З. Шором в [1] алгоритм, позволяющий минимизировать выпуклые функции с разрывным градиентом, приобрел большую значимость в силу многочисленных практических приложений. Именно с него началось планомерное исследование метода обобщенного градиентного спуска, который получил в дальнейшем название субградиентного метода [4]. До того времени вопросы минимизации негладких функций не привлекали большого внимания математиков, а рассматривались лишь эпизодически (например, в теории чебышевских приближений, в теории линейных неравенств). Развитие вычислительной техники, связанное с широким распространением электронных вычислительных машин (ЭВМ), стимулировало интерес к проблемам оптимизации. Оно привело к бурному развитию математического программирования, так как, с одной стороны, давало возможность реализовывать сложные алгоритмы и решать задачи высокой размерности, встречающиеся на практике; с другой стороны, появилась возможность проводить в короткие сроки проверку эффективности новых методов и алгоритмов.

О сути проблем, связанных с минимизацией недифференцируемых функций, и о вкладе Н.З. Шора в их разрешение замечательно написал Б.Т. Поляк [5]:

«Основные алгоритмы минимизации гладких функций — градиентный и Ньютона — были построены на использовании линейной и квадратичной аппроксимации функции, задаваемой первыми членами ряда Тейлора. Однако для недифференцируемой функции эта идея неприменима — такая функция не может быть хорошо аппроксимирована ни линейной, ни квадратичной функциями... Поэтому разработка методов минимизации негладких функций требует привлечения новых идей. Одна из них, принадлежащая Н.З. Шору, выглядит несколько неожиданно. Пишется прямой аналог градиентного метода с заменой градиента на произвольный субградиент $g_f(x)$ функции $f(x)$:

$$x_{k+1} = x_k - \gamma_k g_f(x_k) . \tag{1}$$

... Значения функции в методе (1) не могут убывать монотонно. Оказывается, однако, что при этом монотонно убывает другая функция — расстояние до точки минимума, и в этом-то заключается основная идея субградиентного метода (1).»

Анализу первой идеи Шора (1962) мы посвятим следующий раздел, где приведем ряд результатов по субградиентному методу (методу обобщенного градиентного спуска) и укажем на их связь с другими известными результатами.

1. СУБГРАДИЕНТНЫЙ МЕТОД

Пусть $f(x)$ — выпуклая функция, определенная в евклидовом пространстве E^n, X^* — множество минимумов (которое может быть и пустым), $x^* \in X^*$— точка минимума; $\inf\limits_{x \in E^n} f(x) = f^*$; $g_f(x)$ — субградиент (произвольный) функции $f(x)$ в точке x.

Определение 1. Субградиентом выпуклой функции $f(x)$ в точке $\bar{x}$ называется вектор $g_f(\bar{x})$, удовлетворяющий неравенству

$$f(x) - f(\bar{x}) \geq (g_f(\bar{x}), x - \bar{x}) \text{ для всех } x \in E^n . \tag{2}$$

Здесь $(\cdot,\cdot)$ — скалярное произведение векторов из E^n. Если $f(x)$ — непрерывно дифференцируема в точке $\bar{x}$, то субградиент $g_f(\bar{x})$ определяется однозначно и совпадает с $\nabla f(\bar{x})$ — градиентом функции $f(x)$ в точке $\bar{x}$. В точках негладкости функции $f(x)$ субградиент $g_f(\bar{x})$ определяется неоднозначно.

Из неравенства (2) следует, что если $f(x) < f(\bar{x})$, то субградиент $g_f(\bar{x})$ удовлетворяет неравенству

$$(-g_f(\bar{x}), x - \bar{x}) > 0 . \tag{3}$$

Геометрически формула (3) означает, что антисубградиент в точке $\bar{x}$ образует острый угол с произвольным направлением, проведенным из точки $\bar{x}$ в точку x с меньшим значением $f(x)$. Отсюда если X^* непусто и $\bar{x} \notin X^*$, то при сдвиге из точки $\bar{x}$ в направлении $-g_f(\bar{x})$ с достаточно малым шагом расстояние до X^* убывает. Этот простой факт является центральной идеей субградиентного метода минимизации негладких функций, и именно о нем идет речь в приведенной выше цитате Б.Т. Поляка.

Определение 2. Субградиентным методом называется процедура построения последовательности $\{x_k\}_{k=0}^{\infty}$ по правилу

$$x_{k+1} = x_k - h_k \frac{g_f(x_k)}{\| g_f(x_k) \|}, \; k = 0, 1, 2, \ldots , \tag{4}$$

где x_0 — начальное приближение, h_k — шаговый множитель, $g_f(x_k)$ — произвольный субградиент функции $f(x)$ в точке x_k. Если $g_f(x_k) = 0$, то x_k является точкой минимума функции $f(x)$ и процесс (4) останавливается.

Наиболее общий результат о сходимости субградиентного метода связан с классическими условиями регулировки шага и содержится в следующей теореме [2].

Теорема 1. Пусть $f(x)$ — выпуклая функция с ограниченной областью минимумов X^*, $\{h_k\}_{k=0}^{\infty}$ — последовательность чисел, обладающая свойствами

$$h_k > 0; \ \lim_{k\to\infty} h_k = 0; \ \sum_{k=0}^{\infty} h_k = +\infty .$$

Тогда последовательность $\{x_k\}_{k=0}^{\infty}$, полученная по формуле (4), при произвольном $x_0 \in E^n$ обладает одним из следующих свойств: либо найдется такое $k = k^*$, что $x_{k^*} \in X^*$, либо $\lim_{k\to\infty} \min_{x\in X^*} \| x_k - x \| = 0$, $\lim_{k\to\infty} f(x_k) = \min_{x\in E^n} f(x) = f^*$. ■

Теорема 1 является одним из ярких результатов применения идеи Шора (1962). Не менее ярким результатом применения этой же идеи является предложенный Б.Т. Поляком субградиентный метод

$$x_{k+1} = x_k - h_k^* \frac{g_f(x_k)}{\| g_f(x_k) \|}, \ h_k^* = \frac{f(x_k) - f^*}{\| g_f(x_k) \|}, \ k = 0, 1, 2, \dots , \tag{5}$$

где регулировка шага использует априорное знание значения функции в точке минимума [5]. Здесь шаговый множитель h_k^* задает величину максимального сдвига в направлении нормированного антисубградиента, при котором угол между антисубградиентом и направлением из точки x_{k+1} в точку минимума будет нетупым. Это гарантирует уменьшение расстояния до множества X^* на каждой итерации метода (5).

Величину h_k^* называют шагом Поляка или шагом Агмона–Моцкина–Шенберга [6]. Этот шаг тесно связан с результатами И.И. Еремина о сходимости итерационных методов аппроксимации неподвижных точек с помощью операторов, обладающих свойством квазисжимаемости (фейеровости). Свойство фейеровских операторов можно считать аналогом уменьшения расстояния до множества X^* в субградиентном методе. Подробное изложение результатов по итерационным процессам фейеровского типа дано в [7].

Имеется несколько вариантов доказательства теоремы 1 или ее аналога для субградиентного процесса в форме (1). Все они основаны на изучении поведения последовательности $\{\rho_k\}_{k=0}^{\infty}$, где $\rho_k = \min_{x\in X^*} \| x_k - x \|$. Наиболее общий результат (для случая выпуклых функций, определенных в гильбертовом пространстве, когда минимизация производится при наличии ограничений) получен Б.Т. Поляком (Доклады АН СССР, 1967, № 1). Аналогичный результат для конечномерного случая получен Ю.М. Ермольевым (Кибернетика, 1966, № 4). Оба указанных результата используют принцип доказательства от противного и не содержат конструктивных механизмов их распространения на специальные классы выпуклых функций.

Конструктивный механизм содержит доказательство теоремы 1, принадлежащее Н.З. Шору (Труды I Зимней школы по математическому программированию, 1969, Дрогобыч). Оно использует вспомогательный результат о свойстве субградиентного процесса с постоянным шагом [2].

Лемма 1. Пусть в субградиентном методе (4) используется регулировка шага $h_k = h$, $h > 0$, для всех $k = 0, 1, \dots$ Тогда для любого $\varepsilon > 0$ и $x^* \in X^*$ найдутся такие $k = k^*$ и $\bar{x} \in E_n$, что будет выполняться свойство

$$f(\bar{x}) = f(x_{k^*}), \text{ причем } \| \bar{x} - x^* \| < \frac{h}{2}(1 + \varepsilon) . \quad ■$$

Лемма 1 гарантирует уменьшение расстояния до точек области минимумов только в случае достаточно большого расстояния от множества минимумов до точек x_k. Однако с помощью этой леммы можно обосновать сходимость субградиентного метода при дополнительных предположениях о свойствах минимизируемой функции. Так, например, если множество минимумов содержит сферу радиуса $r > h/2$, то субградиентный метод с постоянным шагом h гарантирует нахождение такого k^*, что $x_{k^*} \in X^*$.

При определенных дополнительных предположениях о свойствах функции $f(x)$ были получены субградиентные методы, сходящиеся со скоростью геометрической прогрессии [2].

Теорема 2. Пусть $f(x)$ — выпуклая функция, определенная на E^n, и для всех $x \in E^n$ при некотором $\varphi\,(0 \le \varphi < \pi/2)$ выполняется неравенство

$$(g_f(x), x - x^*(x)) \ge \cos\varphi \, \| g_f(x) \| \cdot \| x - x^*(x) \| , \tag{6}$$

где $x^*(x)$ — точка, принадлежащая множеству минимумов функции $f(x)$ и лежащая на кратчайшем расстоянии от x. Тогда если при заданном x_0 выбрать величину h_0, удовлетворяющую неравенству

$$h_0 \ge \begin{cases} \| x^*(x_0) - x_0 \| \cos\varphi, & \pi/4 \le \varphi < \pi/2, \\ \| x^*(x_0) - x_0 \| / (2\cos\varphi), & 0 \le \varphi < \pi/4, \end{cases}$$

определить $\{h_k\}_{k=0}^{\infty}$ в соответствии с рекуррентной формулой

$$h_{k+1} = h_k r(\varphi), \quad k = 0, 1, 2, \ldots,$$

где

$$r(\varphi) = \begin{cases} \sin\varphi, & \pi/4 \le \varphi < \pi/2, \\ 1/(2\cos\varphi), & 0 \le \varphi < \pi/4, \end{cases}$$

и вычислить $\{x_k\}_{k=1}^{\infty}$ по формуле (4), то либо при некотором k^* имеем $g_f(x_{k^*}) = 0$ и x_{k^*} принадлежит области минимумов, либо при всех $k = 0, 1, 2, \ldots$ выполняется неравенство

$$\| x_k - x^*(x_k) \| \le \begin{cases} h_k / \cos\varphi, & \pi/4 \le \varphi < \pi/2, \\ 2\cos\varphi \cdot h_k, & 0 \le \varphi < \pi/4. \end{cases}$$

■

Таким образом, если угол φ заранее известен, то, регулируя шаг по формулам теоремы 2, можно получить сходимость к минимуму со скоростью геометрической прогрессии со знаменателем $q = r(\varphi)$. В формуле (6) величина $\cos\varphi$ характеризует степень вытянутости поверхностей уровня функции $f(x)$. Если в некоторой окрестности минимума функции $f(x)$ не существует такого угла $\varphi < \pi/2$, что для любого x из этой окрестности выполняется (6), то такую функцию называют существенно овражной. При минимизации существенно овражных функций приведенный в теореме 2 способ регулировки шаговых множителей неприменим. В этом случае следует использовать универсальный способ выбора шаговых множителей, указанный в теореме 1.

Особо следует выделить сходящийся со скоростью геометрической прогрессии вариант субградиентного метода, где шаговый множитель остается в течение определенного числа итераций постоянным, а затем уменьшается в два раза [2].

Теорема 3. Пусть выпуклая функция $f(x)$ определена на E^n, x^* — единственная точка минимума $f(x)$ и заданы начальное приближение x_0 и числа σ и h_0, причем $\sigma \ge 2$, $h_0 \ge \| x_0 - x^* \| / \sigma$. Рассмотрим множество

$$Y=\{y:\|y-x^*\|\le\sigma h_0\}$$

и пусть для любой пары точек x, $z\in Y$ такой, что $f(x)=f(z)\ne f(x^*)$, выполняется условие

$$\|x-x^*\|\le\sigma\|z-x^*\|.$$

Рассмотрим субградиентный метод (4), где $h_k=h_0\cdot 2^{-\lfloor k/N\rfloor}$. Здесь $\lfloor a\rfloor$ — целая часть числа a. При достаточно большом h_0 и $N\ge 3\sigma^2+1$ выполняется неравенство

$$\|x_k-x^*\|\le 2\sigma h_k,\ k=0,1,2,\dots \qquad \blacksquare$$

Теорема 3 сформулирована в терминах, характеризующих степень «вытянутости» поверхностей уровня, и легко видеть, что чем больше σ, тем меньшей будет скорость сходимости субградиентного метода.

Ступенчатая регулировка шага (как в теореме 3) была тем первым способом регулировки шага в субградиентном методе, которую в 1962 г. Н.З. Шор применил при решении транспортной задачи в сетевой форме [1]. Выполнялось это на машине М-20, для которой С.В. Брановицкая составила программу. Программа имела 150 команд. Одна из задач объемом 23×467 решалась приблизительно 80 мин, задача объемом 90×243 решалась 35 мин. Анализ расчетов показал, что отклонения по стоимости от оптимального плана не превышают 0.2–03%. Эти расчеты на годы вперед определили модели блочных задач линейного и нелинейного программирования большой размерности (характеризуются блочной структурой матрицы ограничений и сравнительно небольшим числом связей между блоками), для которых Н.З. Шор вместе с учениками разрабатывал алгоритмы решения на основе субградиентных методов.

Ступенчатая регулировка шага, как и в теореме 3, была использована в 1967 г. Н.З. Шором и М.Б. Щепакиным при построении (на основе субградиентного метода) алгоритма решения двухэтапной задачи стохастического программирования. Вопросы, связанные со ступенчатой регулировкой шага для метода ОГС, остаются актуальными и сейчас. Так, например, в недавней работе Е.А. Нурминского [8] доказана сходимость фейеровских процессов с общей адаптивной схемой управления шаговыми множителями. Она связана со ступенчатой регулировкой шага, где шаговый множитель уменьшается в несколько раз не через фиксированное количество итераций, а определяется автоматически решением вспомогательной задачи нахождения вектора минимальной длины в выпуклой оболочке накопленных ранее векторов (субградиентов).

Наиболее интенсивно исследования по субградиентным методам в Институте кибернетики проводились в 60–70 годы прошлого столетия. Параллельно они проводились учеными из других научных центров СССР, например И.И. Ереминым (Свердловск) и Б.Т. Поляком (Москва) для решения задач выпуклого программирования с ограничениями. Следует отметить, что результаты по субградиентным методам, полученные в СССР, вызвали огромный интерес за рубежом [9], когда в них увидели ключ к решению задач большой размерности. Так, например, использование схем декомпозиции для решения блочных задач линейного и нелинейного программирования приводит к сравнительно небольшим задачам минимизации негладких функций от связывающих переменных или от множителей Лагранжа для связывающих ограничений. Центральную роль при решении блочных задач математического программирования сыграли ускоренные варианты субградиентных методов, которые будут рассмотрены ниже.

Отметим, что субградиентные методы остаются актуальными и в настоящее время. Так, несмотря на огромное быстродействие и значительную оперативную память современных компьютеров, всегда найдутся задачи с таким количеством переменных, для которых они являются единственным путем их решения. Кроме того, субградиентные методы можно рассматривать как способ ускорения методов по типу покоординатного спуска, если взамен одной координаты рассматривать группу координат, зафиксировав остальные. Методы покоординатного спуска, например, пропагандируются Ю.Е. Нестеровым, как эффективный подход к решению оптимизационных задач «сверхвысоких» размеров [10].

2. СУБГРАДИЕНТНЫЕ МЕТОДЫ С РАСТЯЖЕНИЕМ ПРОСТРАНСТВА

При решении практических задач субградиентные методы оказались медленно сходящимися, что привело к необходимости построения алгоритмов, которые были бы эффективными при минимизации овражных выпуклых функций. Здесь центральную роль сыграла вторая важная идея Н.З. Шора (1969). Она связана с применением линейных неортогональных преобразований пространства для улучшения свойств оптимизируемой функции в преобразованном пространстве переменных и восходит к работе Н.З. Шора и В.И. Билецкого [11]. Суть идеи Шора состоит в следующем.

Пусть на k-й итерации субградиентного метода проводится замена переменных $x = B_k y$, где B_k — неособенная $n \times n$-матрица (т.е. существует обратная матрица $A_k = B_k^{-1}$). Субградиент выпуклой функции $f(x)$ в точке x_k удовлетворяет неравенству

$$f(x) \geq f(x_k) + (g_f(x_k), x - x_k) \quad \forall x \in E^n,$$

откуда, осуществляя замену переменных $x = B_k y$, получаем

$$\varphi(y) \geq \varphi(y_k) + (B_k^T g_f(x_k),\ y - y_k) \quad \forall y \in E^n.$$

Вектор $g_\varphi(y_k) = B_k^T g_f(x_k)$ удовлетворяет неравенству

$$\varphi(y) \geq \varphi(y_k) + (g_\varphi(y_k),\ y - y_k) \quad \forall y \in E^n$$

и является субградиентом выпуклой функции $\varphi(y) = f(B_k y)$ в точке $y_k = A_k x_k$ преобразованного пространства переменных $y = A_k x$.

Пусть к функции $\varphi(y)$ применяется субградиентный метод, где h_k — шаговый множитель в направлении нормированного антисубградиента. В преобразованном пространстве переменных $y = A_k x$ этот метод имеет вид

$$y_{k+1} = y_k - h_k \frac{g_\varphi(y_k)}{\| g_\varphi(y_k) \|} = y_k - h_k \frac{B_k^T g_f(x_k)}{\| B_k^T g_f(x_k) \|} \tag{7}$$

и, следовательно, очередное приближение $x_{k+1} = B_k y_{k+1}$ будет получено по формуле

$$x_{k+1} = B_k y_k - h_k B_k \frac{B_k^T g_f(x_k)}{\| B_k^T g_f(x_k) \|} = x_k - h_k B_k \frac{B_k^T g_f(x_k)}{\| B_k^T g_f(x_k) \|}. \tag{8}$$

Формулы (7) и (8) являются центральными в идее Шора (1969). Очевидно, что если их дополнить правилом рекуррентного пересчета матрицы $B_{k+1} = B_k T_k$, то получим наглядную интерпретацию субградиентного метода с последовательным (от одной итерации к другой) преобразованием пространства переменных. Если матрицы T_k выбирать так, чтобы поверхность овражной функции в очередном преобразованном пространстве переменных становилась менее овражной, то такой метод окажется эффективнее, чем субградиентный метод.

Пусть x_0 — начальное приближение, B_0 — неособенная $n \times n$-матрица. Тогда субградиентный метод с последовательным преобразованием пространства переменных имеет вид

$$x_{k+1} = x_k - h_k B_k \frac{B_k^T g_f(x_k)}{\| B_k^T g_f(x_k) \|}, \quad B_{k+1} = B_k T_k, \ k = 0, 1, \ldots, \tag{9}$$

где h_k — шаговый множитель, T_k — $n \times n$-матрица, $g_f(x_k)$ — произвольный субградиент функции $f(x)$ в точке x_k.

Метод (9) принято называть B-формой субградиентного метода с преобразованием пространства; на каждой его итерации корректируется матрица, связанная с заменой переменных $y = Bx$. Его можно записать в H-форме (по типу методов переменной метрики) с помощью симметричной матрицы $H_k = B_k B_k^T$. Если матрицы $T_0, T_1, \ldots, T_{k-1}$ — неособенные и выпуклая функция $f(x)$ — непрерывно дифференцируема, $\nabla f(x_k)$ — ее градиент в точке x_k, то направление $-B_k B_k^T \nabla f(x_k) = -H_k \nabla f(x_k)$ всегда есть направлением спуска, т.е. направлением убывания функции $f(x)$. Метод (9) в B-форме или его аналог в H-форме является компактным изложением второй идеи Шора.

В 1969–1971 гг. под руководством Н.З. Шора построены два семейства субградиентных методов с растяжением пространства переменных, которые различаются выбором направления растяжения. В их основе лежит оператор растяжения пространства, который в матрично-векторной форме имеет вид

$$R_\alpha(\xi) = I_n + (\alpha - 1)\xi\xi^T, \ \xi \in E^n, \ \|\xi\| = 1, \ \alpha > 1,$$

где $(\cdot)^T$ означает транспонирование, I_n — единичная $n \times n$-матрица, α — коэффициент растяжения пространства, ξ — направление растяжения. Подробно свойства оператора $R_\alpha(\xi)$ изложены в монографии [2].

При описании алгоритмов в B-форме (на каждой итерации корректируется матрица B_k) используется оператор

$$R_\beta(\xi) = R_\alpha^{-1}(\xi) = I_n + (\beta - 1)\xi\xi^T, \ \beta = \frac{1}{\alpha} < 1.$$

Оператор $R_\beta(\xi)$ является обратным к оператору растяжения пространства $R_\alpha(\xi)$, и в методах с растяжением пространства переменных он обеспечивает пересчет матрицы B_{k+1} за $2n^2$ арифметических операций умножения. Действительно,

$$B_{k+1} = B_k R_\beta(\xi) = B_k(I_n + (\beta - 1)\xi\xi^T) = B_k + (\beta - 1)(B_k\xi)\xi^T,$$

откуда легко видеть, что вычисление вектора $\eta = B_k\xi$ требует n^2 умножений и столько же умножений требует построение одноранговой матрицы $\eta\xi^T$.

В первом семействе субградиентных методов используется операция растяжения пространства в направлении субградиента. B-форму этого семейства методов можно описать следующим образом.

Определение 3. Субградиентным методом с растяжением пространства в направлении субградиента называется процедура построения последовательностей $\{x_k\}_{k=0}^\infty$ и $\{B_k\}_{k=0}^\infty$ по следующему правилу:

$$x_{k+1} = x_k - h_k B_k \xi_k, \ B_{k+1} = B_k R_{\beta_k}(\xi_k), \ k = 0, 1, 2, \ldots, \tag{10}$$

где

$$\xi_k = \frac{B_k^T g_f(x_k)}{\| B_k^T g_f(x_k) \|}, \quad \beta_k = \frac{1}{\alpha_k} < 1. \tag{11}$$

Здесь x_0 — начальное приближение, $B_0 = I_n$ — единичная $n \times n$-матрица, h_k — шаговый множитель, α_k — коэффициент растяжения пространства, $g_f(x_k)$ — произвольный субградиент функции $f(x)$ в точке x_k. Если $g_f(x_k) = 0$, то x_k является точкой минимума функции $f(x)$ и процесс (10), (11) останавливается.

Первые эксперименты показали, что, выбирая $\alpha_k = 2$ и $h_k = \text{const}$, для многих примеров выпуклых овражных функций можно получить хорошие результаты [11]. К сожалению, такой простой способ не всегда приводит к цели. Теоретически удалось обосновать такие алгоритмы, где шаговый множитель h_k и коэффициенты растяжения пространства α_k выбирались таким образом, чтобы последовательность расстояний до точки минимума в соответствующих преобразованных пространствах не возрастала [12]. Для этого использовалась дополнительная информация о функции $f(x)$ — значение функции в точке минимума f^* и так называемые постоянные роста M и N.

Теорема 4. Пусть $f(x)$ — выпуклая функция, определенная в E^n, и в некоторой сферической окрестности S_d, $S_d = \{x : \| x - x^* \| \le d\}$, точки минимума x^* субградиент удовлетворяет двустороннему неравенству

$$N(f(x) - f(x^*)) \le (g_f(x), x - x^*) \le M(f(x) - f(x^*)), \tag{12}$$

где $M \ge N$ — положительные константы. Тогда если в методе (10), (11) принять

$$x_0 \in S_d, \; h_k = \frac{2MN}{M+N} \frac{f(x_k) - f(x^*)}{\| B_k^T g_f(x_k) \|}, \; \alpha_k = \frac{M+N}{M-N},$$

то последовательность $\{x_k\}_{k=0}^{\infty}$ удовлетворяет неравенству

$$\| A_k (x_k - x^*) \| \le d, \; A_k = B_k^{-1}, \; k = 0, 1, 2, \ldots \tag{13}$$

Из неравенства (13) следует локализация x^* в эллипсоиде $\Phi_k = \{x : \| A_k (x_k - x) \| \le d\}$ с центром в точке x_k. Отношение объемов эллипсоидов Φ_{k+1} и Φ_k задается следующим равенством:

$$\frac{\text{vol}(\Phi_{k+1})}{\text{vol}(\Phi_k)} = \beta_k = \frac{M-N}{M+N}. \qquad \blacksquare$$

Для выпуклых функций с постоянными роста M и N теорема 4 определяет вариант субградиентного метода с растяжением пространства в направлении субградиента, который сходится со скоростью геометрической прогрессии по отклонению наилучшего достигнутого значения $f(x)$ от оптимального $f^* = f(x^*)$. Это обеспечивает выполнение неравенства (13), в соответствии с которым объем эллипсоида, в котором локализуется точка x^*, убывает со скоростью геометрической прогрессии со знаменателем $\dfrac{M-N}{M+N}$.

Для квадратичной положительно-определенной функции в неравенстве (12) можно выбирать $M = N = 2$. Для кусочно-линейной функции, надграфик которой представляет собой конус с вершиной в точке (x^*, f^*), можно выбирать $M = N = 1$. Если в методе (10), (11) выбрать $\beta_{k+1} = \beta = 0$, то этим случаям соответствуют алгоритмы, которые сходятся за число шагов, не превышающее n. Решение невырожденной системы n линейных уравнений с n неизвестными $(a_i, x) + b_i = 0$, $i = 1, \ldots n$, можно заменить нахождением минимума $f(x) = \max\limits_{1 \le i \le n} |(a_i, x) + b_i|$. Если взять $f^* = 0$, $\beta_k = 0$ и применить метод (10), (11), то получим алгоритм, соответствующий известной конечной процедуре решения линейных алгебраических систем — методу ортогонализации градиентов.

Известный метод эллипсоидов[1] является частным случаем методов с растяжением пространства в направлении субградиента [13].

Теорема 5. Пусть $f(x)$ — выпуклая функция, определенная в E^n, и начальное приближение x_0 такое, что существует точка $x^* \in X^*$, для которой выполняется $\| x_0 - x^* \| \le d$. Тогда, если в методе (10), (11) принять

$$h_0 = \frac{d}{n+1}, \; h_{k+1} = h_k \frac{n}{\sqrt{n^2-1}}, \; \alpha_k = \alpha = \sqrt{\frac{n+1}{n-1}}, \; k = 0,1,2, \ldots,$$

то последовательность $\{x_k\}_{k=0}^{\infty}$ удовлетворяет неравенству

$$\| A_k (x_k - x^*) \| \le h_k (n+1), \; A_k = B_k^{-1}, \; k = 0,1,2, \ldots \tag{14}$$

■

Множество точек x, удовлетворяющих неравенству

$$\| A_k (x_k - x) \| \le (n+1) h_k = d \left(\frac{n}{\sqrt{n^2-1}} \right)^k,$$

представляет собой эллипсоид Φ_k. Его объем $\mathrm{vol}(\Phi_k)$ определяется по формуле

$$\mathrm{vol}(\Phi_k) = \frac{v_0 d^n \left(\frac{n}{\sqrt{n^2-1}} \right)^{nk}}{\det A_k},$$

где v_0 — объем единичного n-мерного шара. Отсюда получаем

$$\frac{\mathrm{vol}(\Phi_{k+1})}{\mathrm{vol}(\Phi_k)} = \frac{\left(\frac{n}{\sqrt{n^2-1}} \right)^n \cdot \det A_k}{\det A_{k+1}} = \frac{\left(\frac{n}{\sqrt{n^2-1}} \right)^n \cdot \det A_k}{\det R_\alpha(\xi_k) \cdot \det A_k} =$$

$$= \frac{1}{\alpha} \left(\frac{n}{\sqrt{n^2-1}} \right)^n = \sqrt{\frac{n-1}{n+1}} \left(\frac{n}{\sqrt{n^2-1}} \right)^n = q_n \approx 1 - \frac{1}{2n^2}.$$

Следовательно, объем эллипсоида, в котором локализуется точка x^*, в соответствии с неравенством (14) убывает со скоростью геометрической прогрессии со знаменателем q_n. Этот знаменатель зависит лишь от n — размерности пространства переменных и не зависит от свойств минимизируемой функции $f(x)$. Благодаря этому факту метод эллипсоидов сыграл важную роль в теории сложности задач математического программирования. На его основе в 1979 г. Л.Г. Хачиян построил и обосновал первый полиномиальный алгоритм решения задачи линейного программирования с рациональными коэффициентами. Метод эллипсоидов позволил обосновать полиномиальные алгоритмы для ряда комбинаторных задач [15].

Опыт применения алгоритмов с растяжением пространства в направлении субградиента показал существенное ускорение субградиентных процессов. Однако оказалось, что такие методы в принципе не могут быть монотонными. Это связано с простым геометрическим фактом: если находиться на границе двух «кусков» кусочно-гладкой поверхности уровня, а градиенты к этим гладким «кускам», вычисленные в данной точке, образуют тупой угол, то никакое растяжение про-

[1] Первыми метод эллипсоидов в H-форме предложили Д.Б. Юдин и А.С. Немировский [14], исходя из методов последовательных отсечений.

странства в направлении градиентов не может преобразовать этот угол в острый, он может лишь приближаться к $\pi / 2$, оставаясь тупым. Применяя растяжение пространства в направлении субградиента, невозможно получить направление убывания функции в виде антиградиента к одному из кусков в растянутом пространстве. В то же время растяжение пространства в направлении разности двух указанных градиентов с достаточным коэффициентом растяжения преобразует тупой угол между градиентами в острый, т.е. соответствующие образы этих антиградиентов в растянутом пространстве становятся направлениями убывания функции.

Это стимулировало разработку второго семейства субградиентных методов, в которых используется растяжение пространства в направлении разности двух последовательных субградиентов. Эти методы получили название *r*-алгоритмов (от русского слова «разность») и стали одним из центральных результатов докторской диссертации Н.З. Шора (1970). Выбор шагового множителя в *r*-алгоритмах связан с поиском минимума функции по направлению. При определенной регулировке шага и коэффициентов растяжения пространства они являются монотонными по минимизируемой функции.

Рассмотрим описание *r*-алгоритмов в *B*-форме для минимизации выпуклой функции $f(x)$, определенной в E^n, и будем предполагать, что $f(x)$ имеет ограниченную область минимумов X^*. Для этого требуется выполнение условия $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$, которое обеспечивает корректность регулировки шага.

Определение 4. *r*-алгоритмом называется процедура построения последовательностей $\{x_k\}_{k=0}^{\infty}$ и $\{B_k\}_{k=0}^{\infty}$ по следующему правилу:

$$x_{k+1} = x_k - h_k B_k \xi_k \,, \quad B_{k+1} = B_k R_{\beta_k}(\eta_k)\,, \quad k = 0, 1, 2, \ldots, \tag{15}$$

где

$$\xi_k = \frac{B_k^T g_f(x_k)}{\| B_k^T g_f(x_k) \|}\,, \quad h_k \ge h_k^* = \operatorname*{argmin}_{h \ge 0} f(x_k - h B_k \xi_k)\,, \tag{16}$$

$$\eta_k = \frac{B_k^T r_k}{\| B_k^T r_k \|}\,, \quad r_k = g_f(x_{k+1}) - g_f(x_k), \quad \beta_k = \frac{1}{\alpha_k} < 1\,. \tag{17}$$

Здесь x_0 — начальное приближение, $B_0 = I_n$ — единичная $n \times n$-матрица[2], h_k — шаговый множитель, α_k — коэффициент растяжения пространства, $g_f(x_k)$ и $g_f(x_{k+1})$ — произвольные субградиенты функции $f(x)$ в точках x_k и x_{k+1}. Если $g_f(x_k) = 0$, то x_k является точкой минимума функции $f(x)$ и процесс (15)–(17) останавливается.

Применительно к задачам минимизации гладких функций *r*-алгоритмы по своей формальной структуре близки к алгоритмам квазиньютоновского типа с переменной метрикой. Так, предельный вариант *r*-алгоритма с бесконечным коэффициентом растяжения (здесь $\beta_k = \beta = 0$, $h_k = h_k^*$) является проективным вариантом метода сопряженных градиентов [16]. Для задачи минимизации выпуклой непрерывно дифференцируемой функции $f(x)$ предельный вариант *r*-алгоритма с восстановлением матрицы B_k после каждых *n* итераций обладает квадратичной скоростью сходимости при обычных условиях гладкости и регулярности $f(x)$ [2].

Несмотря на то, что *r*-алгоритмы используются уже 40 лет, проблема обоснования их сходимости для всего класса выпуклых функций остается открытой и в настоящее время. Еще в 1982 г. Н.З. Шор и В.И. Гершович отметили: «Теория всего класса алгоритмов с растяжением пространства далека от совершенства. Нам кажется достаточно реалистичной целью — построение такого ал-

[2] В качестве матрицы B_0 часто выбирают диагональную матрицу D_n с положительными элементами на диагонали, с помощью которой осуществляется масштабирование переменных.

горитма, который по своей практической эффективности не уступал бы r-алгоритму и был столь же хорошо обоснован, как метод эллипсоидов». Шагом в этом направлении можно считать работу [17], где для преобразования специального эллипсоида в шар используется антиовражный прием, близкий к тому, который имеет место в r-алгоритмах. Однако здесь растяжение пространства реализуется в направлении разности двух нормированных субградиентов, и близким к направлению разности двух субградиентов оно будет только тогда, когда нормы субградиентов близки.

Замечательное свойство r-алгоритма заключается в том, что его конкретные реализации показывают хорошие результаты при минимизации овражных функций. Одним из эффективных зарекомендовал себя вариант $r(\alpha)$-алгоритма с постоянным коэффициентом растяжения пространства α и адаптивным способом регулировки шага. В нем величина h_k настраивается в процессе выполнения одномерного спуска в направлении нормированного антисубградиента в преобразованном пространстве переменных с помощью параметров h_0^0, q_1, n_h, q_2. Здесь h_0^0 — величина начального шага (используется на первой итерации, на каждой последующей итерации уточняется); q_1 — коэффициент уменьшения шага ($q_1 \le 1$), если условие завершения спуска по направлению ($h_k^0 > h_k^*$) выполняется всего за один шаг одномерного спуска; натуральное число n_h задает число шагов одномерного спуска ($n_h > 1$), через каждые n_h шагов h_k^0 будет увеличиваться в q_2 раз, где q_2 — коэффициент увеличения шага ($q_2 \ge 1$). Подробные рекомендации по выбору коэффициента растяжения пространства и параметров адаптивной регулировки шага даны в [18, с. 45–47]. Суть их выбора состоит в том, чтобы адаптивный способ регулировки шага позволял увеличивать точность поиска минимума функции по направлению в процессе счета и при этом число шагов по направлению не должно превышать в среднем двух-трех на одну итерацию.

За последние 40 лет в Институте кибернетики им. В.М. Глушкова НАН Украины накоплен значительный опыт решения оптимизационных задач с помощью r-алгоритма. К настоящему времени усилиями Н.З. Шора и его учеников Н.Г. Журбенко, Л.П. Шабашовой, В.И. Гершовича, А.В. Кунцевича, П.И. Стецюка, А.П. Лиховида и др. разработано несколько модификаций r-алгоритма применительно к решению различных задач оптимизации [19–21]. Эти модификации использовались в задачах линейного и нелинейного программирования, блочных задачах с различными схемами декомпозиции, при решении минимаксных и матричных задач оптимизации, для вычисления двойственных лагранжевых оценок в многоэкстремальных и комбинаторных задачах оптимизации. На практике они применялись для решения задач оптимального планирования, оптимального проектирования, синтеза и анализа сетей, восстановления изображений, эллипсоидальной аппроксимации и локализации и др. Модификации r-алгоритмов стали центральными методами в системах поддержки и принятия решений для планирования структурно-технологических преобразований на основе семейства оптимизационных межотраслевых моделей с переменными коэффициентами прямых затрат [22].

Следует подчеркнуть, что идея о применении линейных неортогональных преобразований пространства для улучшения свойств оптимизируемой функции в преобразованном пространстве переменных оказалась эффективной в выпуклой оптимизации. Модификации r-алгоритмов, разработанные на основе этой идеи Н.З. Шором и его школой, являются эффективными для минимизации негладких функций. Мировую известность получил метод эллипсоидов — частный случай субградиентных методов с растяжением пространства в направлении субградиента. Конгресс по математическому программированию, который проводился в Бонне в 1982 г., был посвящен методу эллипсоидов и его приложениям. В избранных трудах конгресса [23] опубликован обзорный доклад Н.З. Шора

«Generalized gradient methods of nondifferentiable optimization employing space dilatation operations» о методах негладкой оптимизации, разработанных в Институте кибернетики. Участие Н.З. Шора в работе этого конгресса во многом определило его третью идею, которая будет изложена в следующем разделе.

В рамках идеи Шора (1969) в форме метода (9) естественно ожидать новых семейств субградиентных методов с преобразованием пространства для минимизации негладких выпуклых функций. Так, например, в [24] построено и обосновано ряд эффективных субградиентних методов с преобразованием пространства для нахождения точки минимума выпуклой функции при известном оптимальном значении функции. Они используют шаг Поляка в преобразованном пространстве переменных и гарантируют монотонное уменьшение расстояния до множества минимумов в последовательно преобразованных пространствах переменных. Преобразования пространства реализуются с помощью одноранговогo эллипсоидального оператора и доортогонализирующего однорангового оператора, и цена одной итерации этих методов такая же, как и у r-алгоритмов.

3. ДВОЙСТВЕННЫЕ ОЦЕНКИ В ЭКСТРЕМАЛЬНЫХ КВАДРАТИЧНЫХ ЗАДАЧАХ

Многие задачи булевого линейного программирования могут быть переформулированы как экстремальные квадратичные задачи с булевыми переменными. Так, например, условие булевости переменной $x \in \{0,1\}$ представляется квадратичным равенством $x^2 - x = 0$. Если две булевы переменные x_i и x_j не могут одновременно принимать значение, равное единице, то это условие может быть записано в виде равенства $x_i x_j = 0$ (равносильно линейному неравенству $x_i + x_j \le 1$). Для экстремальных квадратичных задач Н.З. Шор предложил использовать двойственный подход к получению и уточнению оценок целевой функции (третья важная идея). В квадратичных задачах на минимум эти оценки будут границами снизу для минимального значения целевой функции, а в квадратичных задачах на максимум — границами сверху для максимального значения целевой функции.

Третья идея Н.З. Шора включает алгоритмы нахождения двойственных оценок на основе методов недифференцируемой оптимизации и использование функционально избыточных квадратичных ограничений (их добавление не изменяет множества допустимых решений исходной квадратичной задачи) для улучшения точности двойственных оценок. Впервые двойственный подход к получению оценок предложен в работе Н.З. Шора и А.С. Давыдова (1985) для задач булевого программирования [25]. В 1986–1987 годах он был дополнен процедурами уточнения оценок и применен к другим многоэкстремальным задачам, в том числе к задаче нахождения глобального минимума полинома [26–28]. Наиболее полное изложение результатов Н.З. Шора можно найти в его англоязычной монографии [3], а менее полное — в русскоязычной [18].

Опишем суть двойственного подхода для получения и уточнения оценок в квадратичных экстремальных задачах на примере экстремальной квадратичной задачи с ограничениями–равенствами: найти

$$Q^* = \max_{x \in E^n} Q_0(x) \tag{18}$$

при ограничениях

$$Q_i(x) = 0, \quad i = 1, \ldots, m. \tag{19}$$

Здесь квадратичные функции имеют вид

$$Q_i(x) = (K_i x, x) + (b_i, x) + c_i,$$

где K_i — симметричные вещественные $n \times n$-матрицы, b_i — n-мерные векторы из E^n, c_i — вещественные числа, $i = 0, \ldots, m$. Некоторые из функций $Q_i(x)$, $i = 0, 1, \ldots, m$, могут быть также линейными.

В общем случае задача (18), (19) многоэкстремальна и относится к классу NP-трудных задач. Оценки сверху для Q^* можно получить путем следующей лагранжевой релаксации. Пусть $u \in E^m$ — вектор множителей Лагранжа, соответствующий ограничениям (19). Функция Лагранжа для задачи (18), (19) имеет вид

$$L(x, u) = Q_0(x) + \sum_{i=1}^{m} u_i Q_i(x) \equiv (K(u)x, x) + (b(u), x) + c(u),$$

где

$$K(u) = K_0 + \sum_{i=1}^{m} u_i K_i, \; b(u) = b_0 + \sum_{i=1}^{m} u_i b_i, \; c(u) = c_0 + \sum_{i=1}^{m} u_i c_i.$$

Рассмотрим функцию

$$\Psi(u) = \max_{x \in R^n} L(x, u) \equiv \max_{x \in R^n} [(K(u)x, x) + (b(u), x) + c(u)].$$

Пусть $\lambda_{\max}(K)$ — максимальное собственное число симметричной $n \times n$-матрицы K.

Функция $\Psi(u)$ является выпуклой функцией от переменных u (как результат взятия операции максимума по переменным x для семейства линейных по переменным u функций). Область определения функции $\Psi(u)$ обозначим $\mathrm{dom}\,\Psi$. Она состоит из $\Omega^- = \{u \in E^m : \lambda_{\max}(K(u)) < 0\}$ (подмножество $u \in E^m$, для которых матрица $K(u)$ отрицательно определена) и подмножества тех точек $u \in \Omega^0 = \{u \in E^m : \lambda_{\max}(K(u)) = 0\}$, для которых система линейных уравнений

$$2K(u)x + b(u) = 0 \tag{20}$$

имеет решение. Для всех других точек $\Psi(u) = +\infty$.

Если $\mathrm{dom}\,\Psi \neq \varnothing$, то для любого $u \in \mathrm{dom}\,\Psi$ значение функции $\Psi(u)$ является нетривиальной (т.е. не равной $+\infty$) оценкой сверху для Q^* — оптимального значения целевой функции в задаче (18), (19). Наилучшая оценка сверху для Q^* в классе лагранжевых оценок вида $\Psi(u)$ связана с решением следующей задачи негладкой оптимизации:

$$\Psi^* = \min_{u \in \mathrm{dom}\,\Psi} \Psi(u). \tag{21}$$

Точками негладкости функции $\Psi(u)$ есть точки границы множества Ω^-, где система линейных уравнений (20) имеет неединственное решение.

Оценку Ψ^* с любой заданной точностью можно найти за полиномиальное время методом эллипсоидов; для одной его итерации требуется $O(m^2) + O(n^3)$ арифметических операций. Из них $O(m^2)$ операций связано с растяжением пространства двойственных переменных (множителей Лагранжа), а $O(n^3)$ операций требуются для вычисления вектора, определяющего полупространство локализации оптимальных множителей Лагранжа (при фиксированных значениях множителей Лагранжа решается система линейных уравнений с симметричной $n \times n$-матрицей, определяется максимальное собственное число симметричной $n \times n$-матрицы и соответствующий этому числу собственный вектор).

Оценка Ψ^* обладает следующими свойствами. Если минимум в (21) достигается на $u^* \in \Omega^-$, то $\Psi^* = Q^*$ (т.е. оценка точная, см. лемму 4.1 из [18, с. 90]. При этом находится и точка глобального минимума $x^* = x(u^*)$, где $x(u^*)$ — решение системы (20) при $u = u^*$. Если минимум в (21) достигается на границе области Ω^-, то может существовать так называемый «разрыв двойственности» $\Delta^* = \Psi^* - Q^* > 0$. Один из способов уменьшения Δ^* связан с введением функционально избыточных ограничений (при этом может увеличиться и количество переменных в задаче).

Функционально избыточные ограничения — это ограничения, добавление которых не изменяет множества допустимых решений начальной квадратичной задачи. Однако при этом изменяется функция Лагранжа, что в некоторых случаях позволяет уменьшить Δ^*. Если к исходной задаче (18), (19) прибавить r функционально избыточных квадратичных ограничений $Q_{m+1}(x)=0$, ..., $Q_{m+r}(x)=0$, $r \ge 1$, то новая квадратичная задача примет вид: найти

$$Q^* = \max_{x \in R^n} Q_0(x) \tag{22}$$

при ограничениях

$$Q_i(x)=0, \; i=1, \ldots, m+r. \tag{23}$$

Теорема 6. Если Ψ_1^* — оценка вида Ψ^* для задачи (22), (23), то $\Psi_1^* \le \Psi^*$.

Доказательство. Задаче (22), (23) соответствует вектор множителей Лагранжа $U \in E^{m+r}$, и функция Лагранжа для нее имеет вид

$$L_1(x,U) = Q_0(x) + \sum_{i=1}^{m+r} u_i Q_i(x) = L(x,u) + \sum_{i=m+1}^{m+r} u_i Q_i(x).$$

Поскольку $L_1(x, (\{u\}, 0, \ldots, 0)) = L(x,u)$ и $\Psi_1(\{u\}, 0, \ldots, 0) = \Psi(u)$, то

$$\Psi_1^* = \min_{U \in \mathrm{dom}\, \Psi_1} \Psi_1(U) \le \min_{u \in \mathrm{dom}\, \Psi} \Psi(u) = \Psi^*.$$

Теорема 6 не только констатирует, что функционально избыточные ограничения могут улучшить точность двойственной оценки, но и поясняет, что это улучшение есть следствием увеличения количества множителей Лагранжа [3]. Ограничения, которые являются линейными комбинациями уже существующих ограничений, не отражаются на точности двойственной оценки, т.е. $\Psi_1^* = \Psi^*$. Вклад таких ограничений в функцию Лагранжа эквивалентен лишь определенному изменению множителей Лагранжа при существующих ограничениях. Однако добавление функционально избыточных ограничений, которые являются нетривиальными следствиями из условий задачи, в ряде случаев приводит к тому, что двойственная оценка Ψ_1^* может стать точной для Q^*. Последнее означает, что оптимальное значение Q^* со сколь угодно большой точностью можно найти за полиномиальное время, которое зависит от числа переменных и числа ограничений в квадратичной задаче.

Функционально избыточными могут быть следующие ограничения:

а) квадратичные следствия линейных ограничений: например, квадратичное ограничение в форме $(b_i^T x + c_i)(b_j^T x + c_j) \ge 0$ является следствием из двух линейных ограничений-неравенств: $b_i^T x + c_i \ge 0$ и $b_j^T x + c_j \ge 0$;

б) квадратичные ограничения, которые характеризуют неоднозначность представления произведения трех либо большего числа переменных задачи. Как правило, они имеют место при сведении полиномиальной задачи к квадратичной. Например, имеются переменные x_1, $x_2 = x_1^2$ и $x_3 = x_1^3$. Тогда квадратичное ограничение $x_2^2 - x_1 x_3 = 0$ есть следствием неоднозначного представления x_1^4, а именно $x_1^4 = (x_1^2)^2 = (x_1^3)(x_1)$;

в) квадратичные ограничения, которые являются следствиями булевости или бинарности переменных задачи. Например, для бинарных переменных $x_i^2 = 1$, $x_j^2 = 1$, $x_k^2 = 1$ всегда справедливо квадратичное неравенство $x_i x_j + x_i x_k + x_j x_k \ge -1$.

Более детальную информацию о семействах фунционально избыточных ограничений, их использовании для нахождения глобального минимума полинома и в экстремальных задачах на графах (максимальное устойчивое множество вершин графа, максимальный разрез графа и др.) можно найти в монографиях [3, 18].

Чрезвычайно интересными оказались результаты, полученные Н.З. Шором для задачи нахождения взвешенного максимального независимого множества вершин графа [3]. Эта задача играет большую роль в многочисленных приложениях: теории информации и кодировании, проектировании различных устройств при определенных условиях несовместности; она тесно связана с известными задачами выбора, разбиения множеств, раскраски графов и другими комбинаторными задачами, имеющими массу приложений. В то же время она принадлежит к классу *NP*-полных задач (при целочисленных весах вершин). Двойственные оценки Шора для этой задачи тесно связаны с известными числами Ловаса $\theta(G, w)$ и $\theta'(G, w)$ [15], которые играют большую роль при обосновании результатов о полиномиальной разрешимости ряда задач в теории графов.

Пусть $G=(V, E)$ — взвешенный неориентированный граф с множеством вершин V и множеством ребер E, вес каждой вершины $i \in V$ задан положительным целым числом w_i. Подмножество вершин $S \subseteq V$ называется устойчивым (или независимым) множеством графа G, если для любых $i, j \in S$ ребро (i, j) не принадлежит E. Взвешенное число устойчивости графа G определяется как $\alpha(G, w) = \max \sum\limits_{i \in S} w_i$, где $S \subseteq V$ — устойчивое множество. Множество S^*, на котором достигается $\alpha(G, w)$, называется максимальным взвешенным устойчивым (или независимым) множеством. Задача нахождения $\alpha(G, w)$ является *NP*-трудной даже в частном случае, когда все веса равны единице [15]. Поэтому вычисление верхних оценок, достаточно хорошо аппроксимирующих сверху $\alpha(G, w)$, имеет как практический, так и теоретический интерес. Н.З. Шор предложил три верхних оценки для $\alpha(G, w)$.

Первую оценку обозначим $\Psi_1^*(G, w)$. Она связана с формулировкой задачи о максимальном взвешенном устойчивом множестве графа в виде следующей квадратичной булевой задачи:

$$\alpha(G, w) = \max \sum_{i \in V} w_i x_i \tag{24}$$

при ограничениях

$$x_i x_j = 0 \quad \forall (i, j) \in E, \tag{25}$$

$$x_i^2 - x_i = 0 \quad \forall i \in V. \tag{26}$$

Здесь булева переменная $x_i \in \{0,1\}$ равна единице, если вершина $i \in V$ включается в устойчивое множество, и равна нулю в противном случае. Булевы переменные для всех вершин описаны квадратичными ограничениями — равенствами (26). Квадратичные ограничения (25) означают, что если две вершины связаны ребром в графе G, то они обе не могут одновременно принадлежать устойчивому множеству.

Вторая оценка (обозначим ее $\Psi_2^*(G, w)$) соответствует квадратичной задаче, в которой к ограничениям (25), (26) добавлено следующее семейство функционально избыточных ограничений:

$$x_i x_j \ge 0 \quad \forall (i, j) \notin E. \tag{27}$$

Они являются следствиями неравенств $x_i = x_i^2 \ge 0$ для всех $i \in V$.

Третья оценка $\Psi_3^*(G, w)$ связана с квадратичной задачей, в которой к ограничениям (25)–(27) добавлено семейство функционально избыточных ограничений

$$x_i x_k + x_j x_k \le x_k \quad \forall (i, j) \in E, \quad k \ne i, j. \tag{28}$$

Квадратичные ограничения (28) в форме неравенств содержат все возможные квадратичные ограничения, построенные по следующему правилу. Для булевых переменных x_i и x_j, когда $(i, j) \in E$ и вершины i и j не могут одновременно включаться в устойчивое множество, всегда справедливо линейное неравенство $x_i + x_j \le 1$. Если его умножить на переменную x_k такую, что $k \ne i, j$, то знак неравенства не изменится в силу того, что $x_k = x_k^2 \ge 0$. В результате получаем квадратичное неравенство $x_i x_k + x_j x_k \le x_k$, справедливое для $(i, j) \in E$ и $k \ne i, j$.

Результаты о точности всех трех оценок Шора и их связь с числами Ловаса содержатся в следующей теореме [3].

Теорема 7. Для оценок $\Psi_1^*(G, w)$, $\Psi_2^*(G, w)$ и $\Psi_3^*(G, w)$ справедливо неравенство

$$\Psi_1^*(G, w) = \theta(G, w) \ge \Psi_2^*(G, w) = \theta'(G, w) \ge \Psi_3^*(G, w) \ge \alpha(G, w).$$

Следовательно, оценки Шора $\Psi_1^*(G, w)$ и $\Psi_2^*(G, w)$ обладают такими же свойствами, как и числа Ловаса $\theta(G, w)$ и $\theta'(G, w)$, например $\Psi_1^*(G, w) = \alpha(G, w)$, если граф G принадлежит семейству совершенных графов. Наиболее интересной оказалась третья оценка Шора $\Psi_3^*(G, w)$. Она всегда не хуже, чем оценка $\Psi_2^*(G, w)$, а наличие ограничений (28) придает ей ряд замечательных свойств, которые связаны со специальными семействами графов. Так, например, в [3] показано, что оценка $\Psi_3^*(G, w)$ является точной для $\alpha(G, w)$, когда граф G есть *t*-совершенным либо *h*-совершенным.

Функционально избыточные ограничения сыграли огромную роль при получении эффективных оценок снизу для целевой функции в задачах нахождения глобального минимума полиномиальной функции $P(x)$ от одной или нескольких переменных. Эти задачи специальным образом сводятся к многоэкстремальным квадратичным задачам (на минимум) с определенными семействами функционально избыточных квадратичных ограничений. В [3] доказано, что двойственная оценка для таких квадратичных задач совпадает со значением p^* полинома $P(x)$ в точке глобального минимума тогда и только тогда, когда полином $\bar{P}(x) = P(x) - p^*$ может быть представлен в виде суммы квадратов других полиномов. Эти результаты имеют отношение к классическим работам Д. Гильберта по разложению неотрицательных полиномиальных форм в сумму квадратов. Разработанный Шором метод дает возможность не только доказать существование такой декомпозиции (если она существует), но и найти одно из возможных представлений полинома $\bar{P}(x)$ в виде суммы квадратов других полиномов. Более того, этим методом можно определить значение глобального минимума полинома $P(x)$.

Итак, третья идея Шора (1985), связанная с двойственным подходом для получения и уточнения оценок в квадратичных экстремальных задачах, может быть использована при создании эффективных методов решения тех многоэкстремальных задач, которые можно описать с помощью квадратичных моделей (такие модели встречаются во многих приложениях). Эффективность таких методов обеспечивается двумя моментами. Во-первых, двойственные оценки для квадратичных моделей являются более точными, чем оценки в линеаризованных аналогах этих моделей, и целесообразно их использовать в сочетании с методом ветвей и границ. Во-вторых, использование функционально избыточных ограничений позволяет выделить среди многоэкстремальных задач такие их подклассы, которые разрешимы за полиномиальное время. При этом можно даже указать верхнюю границу сложности такого класса задач, которая зависит от количества квадратичных ограничений.

Особого внимания заслуживают и алгоритмы нахождения двойственных оценок в квадратичных моделях. Их можно считать альтернативой использованию методов внутренних точек для решения задач полуопределенного программирования (semidefinite programming), т.е. задач оптимизации, в которых в качестве ограничения фигурирует требование неотрицательной определенности некоторых матриц. Действительно, многие задачи полуопределенного программирования целесообразно рассматривать как задачи недифференцируемой оптимизации и применять для их решения эффективные методы минимизации негладких выпуклых функций. Условие неотрицательной определенности симметричной $(n \times n)$-матрицы X (принято обозначать $X \succeq 0$) эквивалентно тому, что минимальное собственное число матрицы X не отрицательное: $\lambda_{\min}(X) \geq 0$. Но $\lambda_{\min}(X)$ — вогнутая недифференцируемая функция элементов матрицы, т.е. если элементы матрицы $X(u) = \{x_{ij}\}_{i,\, j=1}^{n}$ являются линейными функциями от вектора варьируемых параметров $u \in R^m$, то условие $X(u) \succeq 0$ эквивалентно выпуклому негладкому ограничению $\varphi(u) = -\lambda_{\min}(X(u)) \leq 0$.

Исследования в данных направлениях активно продолжаются. Так, например, недавно выявлены новые свойства для оценки $\Psi_3^*(G, w)$ [29]. Они связаны с тем, что из квадратичных ограничений (25), (26) и (28) вытекают линейные неравенства, справедливые для ряда других известных подструктур в графе, в том числе и для p-колеса (английский термин p-wheel). В графе p-колесо состоит из клики с количеством вершин $p \geq 1$ и нечетного цикла с количеством вершин $(2k+1)$, где $k \geq 2$, причем вершины клики и нечетного цикла не пересекаются, а каждая вершина клики связана со всеми вершинами нечетного цикла. Это дает возможность существенно расширить семейство графов, для которых задача нахождения $\alpha(G, w)$ не является NP-трудной, а может быть решена за полиномиальное время. Как частный случай, сюда относится известное семейство W-совершенных графов; для него доказательство основано на использовании метода эллипсоидов [15]. W-совершенному графу соответствует случай, когда $p = 1$, т.е. в p-колесо входит клика только с одной вершиной.

ЗАКЛЮЧЕНИЕ

Научное наследие Наума Зуселевича Шора намного богаче, чем нам удалось изложить в данной статье. Здесь отмечены только центральные вехи в его творчестве, сделан акцент на субградиентные методы недифференцируемой оптимизации, разработанные в Институте кибернетики им. В.М. Глушкова НАН Украины. Эти методы оказали огромное влияние на развитие теории и практики для многих направлений математического программирования. Книги Н.З. Шора стали настольными для ведущих отечественных и зарубежных специалистов в области математического программирования. Яркой характеристикой этому может служить приведенное ниже письмо профессора С. Бойда (Stephen Boyd) из Стэнфордского университета.

Уважаемый профессор Шор!

Мы никогда не встречались, но ваши работы оказали на меня огромное влияние. Я начал с вашей небольшой книги по субградиентным методам (1985), я тогда был еще аспирантом. А сейчас я читаю вашу новую книгу о недифференцируемой оптимизации (1998) и просто наслаждаюсь ею.

Я высылаю Вам три написанных мною книги. Первая — о проектировании линейных контроллеров с помощью выпуклой оптимизации, вторая — о матричных неравенствах, а третья — учебник по выпуклой оптимизации. [...] Надеюсь, Вы увидите ваше сильное влияние во всех этих книгах.

С уважением,

Стефан П. Бойд, 15.04.2005

В письме упоминаются обе англоязычные монографии Н.З. Шора. Первая книга [2] была переведена польскими учеными К. Кивелем и А. Рущиньским и опубликована издательством Springer–Verlag в 1985 г. Вторая монография [3], опубликованная издательством Kluwer Academic Publishers, не имеет русскоязычного эквивалента. Ее основой стала русскоязычная монография Н.З. Шора и С.И. Стеценко [18], которая была дополнена рядом новых результатов по полиномиальным, матричным и булевым задачам. Центральные из этих результатов отображены в двух сборниках избранных трудов Н.З. Шора, изданных в 2008 и 2009 годах [30, 31]. Они подготовлены к опубликованию в тесном сотрудничестве с учеными Академии транспорта, информатики и коммуникаций (г. Кишинэу, Республика Молдова) и используются в материалах учебных курсов по методам недифференцируемой оптимизации и их приложениям в сложных задачах математического программирования.

Англоязычная статья [32], опубликованная в 2002 г., заканчивается словами: «We want to emphasize that methods of nondifferentiable optimization must become necessary part of the courses of applied mathematics for mathematical and technical education.» Это утвержение Н.З. Шор постоянно доказывал на практике, читая курсы по методам недифференцируемой оптимизации и их приложениям в ведущих вузах Украины: Киевском отделении Московского физико-технического института, Национальном техническом университете Украины «КПИ», Киевском национальном университете имени Тараса Шевченко, Соломоновом университете. Он всегда помогал молодым ученым внимательными и доброжелательными научными консультациями, рекомендациями.

Наума Зуселевича любили и уважали все, кто имел счастье работать с ним на протяжении сорока восьми лет его научной деятельности.

СПИСОК ЛИТЕРАТУРЫ

1. Шор Н.З. Применение метода градиентного спуска для решения сетевой транспортной задачи. Материалы науч. семинара по теорет. и прикл. вопр. кибернетики и исследования операций: Науч. совет по кибернетики АН УССР. — Киев. — 1962. — Вып. 1. — С. 9–17.
2. Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. — Киев: Наук. думка, 1979. — 199 с.
3. Shor N.Z. Nondifferentiable optimization and polynomial problems. — Boston; Dordrecht; London: Kluwer Academic Publishers, 1998. — 412 p.
4. Сергієнко І.В. Методи оптимізації та системного аналізу для задач трансобчислювальної складності. — К.: Академперіодика, 2010. — 296 с.
5. Поляк Б.Т. Введение в оптимизацию. — М.: Наука, 1983. — 384 с.
6. Еремин И.И. Обобщение релаксационного метода Агмона-Моцкина // УМН. — 1965. — т. XX. — Вып. 2 (122). — С. 183–187.
7. Васин В.В., Еремин И.И. Операторы и итерационные процессы фейеровского типа (теория и приложения). — Москва; Ижевск: Институт компьютерных исследований, НИЦ «Регулярная и хаотическая динамика», 2005. — 200 с.
8. Нурминский Е.А. Фейеровские алгоритмы с адаптивным шагом // Журн. вычисл. математики и мат. физики. — 2011. — **51.** — Вып. 5. — С. 791–801.
9. Balinski M.L., Wolfe P. (eds.) Nondifferentiable optimization. Math. Programming Study, 3. — Amsterdam: North-Holland, 1975. — 178 p.
10. Nesterov Y. Efficiency of coordinate descent methods on huge-scale optimization problems // CORE Discussion Paper #2010/2, 2010. — 23 p.
11. Шор Н.З., Билецкий В.И. Метод растяжения пространства для ускорения сходимости в задачах овражного типа // Тр. семинара Науч. совета АН УССР по кибернетике «Теория оптимальных решений». — Киев. — 1969. — № 2. — С. 3–18.
12. Шор Н.З. Использование операций растяжения пространства в задачах минимизации выпуклых функций // Кибернетика. — 1970. — № 1. — С. 6–12.
13. Шор Н.З. Метод отсечения с растяжением пространства для решения задач выпуклого программирования // Там же. — 1977. — № 1. — С. 94–95.
14. Юдин Д.Б., Немировский А.С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач // Экономика и мат. методы. — 1976. — Вып. 2. — С. 357–359.

15. Grötschel M., Lovàsz L., Schrijver A. Geometric algorithms and combinatorial optimization. — Berlin: Springer–Verlag, 1988. — 362 p.

16. Шор Н.З., Журбенко Н.Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. — 1971. — № 3. — С. 51–59.

17. Стецюк П.И. *r*-алгоритмы и эллипсоиды // Кибернетика и системеный анализ. — 1996. — № 1. — С. 113–134.

18. Шор Н.З., Стеценко С.И. Квадратичные экстремальные задачи и недифференцируемая оптимизация. — Киев: Наук. думка, 1989. — 208 с.

19. Шор Н.З., Стецюк П.И. Использование модификации *r*-алгоритма для нахождения глобального минимума полиномиальных функций // Кибернетика и системный анализ. — 1997. — № 4. — С. 28–49.

20. Kappel F., Kuntsevich A.V. An implementation of Shor's *r*-algorithm // Computational Optimization and Applications. — 2000. — **15.** — P. 193–205.

21. Шор Н.З., Журбенко Н.Г., Лиховид А.П., Стецюк П.И. Развитие алгоритмов недифференцируемой оптимизации и их приложения // Кибернетика и системный анализ. — 2003. — № 4. — С. 80–94.

22. Сергиенко И.В., Михалевич М.В., Стецюк П.И., Кошлай Л.Б. Модели и информационные технологии для поддержки принятия решений при проведении структурно-технологических преобразований // Там же. — 2009. — № 2. — С. 26–49.

23. Bachem A., Grötschel M., Korte B. (eds.) Mathematical programming: the state of art, Bonn, 1982. — Berlin: Springer–Verlag, 1983. — 655 p.

24. Стецюк П.И. Ортогонализующие линейные операторы в выпуклом программировании // Кибернетика и системный анализ. — 1997. — № 3. — С. 97–119 (Ч. I). — 1997. — № 5. — С. 111–124 (Ч. II).

25. Шор Н.З., Давыдов А.С. О методе получения оценок в квадратичных экстремальных задачах с булевыми переменными // Там же. — 1985. — № 2. — С. 48–50.

26. Шор Н.З. Квадратичные оптимизационные задачи // Известия АН СССР. «Техн. кибернетика». — 1987. — № 1. — С. 128–139.

27. Шор Н.З. Об одном подходе к получению глобальных экстремумов в полиномиальных задачах математического программирования // Кибернетика. — 1987. — № 5. — С. 102–106.

28. Шор Н.З. Об одном классе оценок глобального минимума полиномиальных функций // Там же. — 1987. — № 6. — С. 9–11.

29. Стецюк П.И. О новых свойствах оценок Шора для взвешенного числа устойчивости графа // Праці міжнарод. конф. «50 років Інституту кібернетики ім. В.М. Глушкова НАН України». — К.: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2008. — С. 164–173.

30. Шор Н.З. Методы недифференцируемой оптимизации и сложные экстремальные задачи: Сб. избр. тр. — Кишинэу: Эврика, 2008. — 270 с.

31. Шор Н.З. Методы минимизации негладких функций и матричные задачи оптимизации: Сб. избр. тр. — Кишинэу: Эврика, 2009. — 240 с.

32. Shor N.Z., Stetsyuk P.I. Lagrangian bounds in multiextremal polynomial and discrete optimization problems // J. of Global Optimization. — 2002. — **23**. — P. 1– 41.

Поступила 22.08.2011