Теорія оптимальних рішень

УДК 519.8

Н.З. ШОР, П.И. СТЕЦЮК, О.А. БЕРЕЗОВСКИЙ

ДВОЙСТВЕННЫЕ ОЦЕНКИ ДЛЯ СПЕЦИАЛЬНОЙ ОПТИМИЗАЦИОННОЙ ЗАДАЧИ КВАДРАТИЧНОГО ТИПА НА МНОГООБРАЗИИ ШТИФФЕЛЯ

Рассматривается задача глобальной оптимизации, связанная с нахождением минимума квадратичной однородной функции специального вида на многообразии Штиффеля. Рассмотрены лагранжевые двойственные оценки для трех вариантов этой задачи, которые различаются ограничениями на компоненты ортонормированных векторов. На тестовых примерах показано, что их можно эффективно использовать при нахождении глобального минимума.

Многобразие Штиффеля (Stiefel manifolds) состоит из всех ортонормированных систем векторов $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k \in R^n$, где R^n – n-мерное евклидово пространство и $k \le n$. В работе [1] рассмотрена следующая оптимизационная задача на многообразии Штиффеля:

$$\min\left\{ f(\vec{x}_1, \vec{x}_2, ..., \vec{x}_k) = \sum_{i=1}^{k} \vec{x}_i^T A_i \vec{x}_i \right\} \qquad (1)$$

при ограничениях

$$\vec{x}_i^T \vec{x}_i = \delta_{ij}, \; 1 \le i, j \le k, \qquad (2)$$

где A_i, $i = \overline{1,k}$ – заданные $n \times n$ вещественные симметричные матрицы, δ_{ij} – символ Кронекера. Ограничения (2) задают многообразие Штиффеля, которое принято обозначать $M_{n,k}$.

В зависимости от условий на компоненты векторов из $M_{n,k}$ выделим следующие варианты задачи (1)–(2):

(R) компоненты векторов $\vec{x}_i$, $i = \overline{1,k}$, принадлежат множеству действительных чисел. Этот вариант совпадает с постановкой задачи (1)–(3);

(Z) компоненты векторов $\vec{x}_i$, $i = \overline{1,k}$, принадлежат множеству целых чисел. Для этого варианта векторы $\vec{x}_i$, $i = \overline{1,k}$, содержат только компоненты либо 0, либо +1, либо –1;

(N) компоненты векторов $\vec{x}_i$, $i=\overline{1,k}$, принадлежат множеству натуральных чисел, т. е. векторы $\vec{x}_i$, $i=\overline{1,k}$, могут содержать только компоненты 0 или 1.

Варианты (Z) и (N) определяются только диагональными коэффициентами матриц A_i и не зависят от коэффициентов вне диагонали, что обусловлено спецификой целевой функции вида (1). Кроме того варианты (Z) и (N) близки в том смысле, что им соответствует одно и то же значение целевой функции в точках глобальных экстремумов. Отличие состоит лишь в числе глобальных экстремумов, т. е. если у варианта (N) их количество равно L, то у варианта (Z) их будет $2^k L$.

В общем случае задача (1)–(2) для каждого из трех вариантов является многоэкстремальной задачей нелинейного программирования (пусть даже и квадратичного типа), и ее решение представляет проблемы даже при небольших размерностях $n=2$ и $n=3$. Так, например, в [2] сообщается, что простой пример на $M_{2,2}$ с диагональными коэффициентами матриц A_i, $i=\overline{1,2}$ требует трех миллионов вычислений значений функции при использовании программы GlobSol [3–4]. Однако, используя технику двойственных лагранжевых квадратичных оценок [5] нахождение глобального минимума задачи (1)–(2) можно значительно ускорить в ряде случаев.

Рассмотрим оптимизационные задачи квадратичного типа, соответствующие каждому из указанных трех вариантов этой задачи, и лагранжевые двойственные оценки для каждой из задач.

Для каждого $i=\overline{1,k}$ обозначим коэффициенты симметричной $n\times n$ матрицы A_i – a_{ijl}, $j=\overline{1,n}$, $l=\overline{1,n}$, а компоненты n-мерного вектора $\vec{x}_i$ – x_{ij}, $j=\overline{1,n}$.

Варианту (R) соответствует оптимизационная задача квадратичного типа:

$$Q_1^* = \min \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{l=1}^{n} a_{ijl}x_{ij}x_{il} \tag{3}$$

при ограничениях

$$\sum_{j=1}^{n} x_{ij}^2 = 1,\ i=\overline{1,k}, \tag{4}$$

$$\sum_{l=1}^{n} x_{il}x_{jl} = 0,\ i=\overline{1,k-1},\ j=\overline{i+1,k}. \tag{5}$$

Задача (3)–(5) – переписанный с учетом расшифровки символа Кронекера δ_{ij} аналог задачи (1)–(2). Здесь ограничения (4) задают условие нормированности векторов $\vec{x}_i$, $i=\overline{1,k}$, а ограничения (5) – условие их взаимной ортогональности. Указанных ограничений достаточно для задания многообразия Штиффеля

$M_{n,k}$, состоящего из всех ортонормированных систем векторов $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k \in R^n$ ($k \le n$).

Пусть $u \in R^m$, где $m = k(k+1)/2$ – вектор множителей Лагранжа, состоящий из двух частей: $\{u_{ii}, i = \overline{1,k}\}$ – множители Лагранжа, соответствующие ограничениям (4) и $\{u_{ij}, i = \overline{1, k-1}, j = \overline{i+1, k}\}$ – множители Лагранжа, соответствующие ограничениям (5). Согласно [5] задаче (3)–(5) соответствует следующая двойственная лагранжева оценка:

$$\psi_1^* = \psi_1(u^*) = \sup_{u: K(u) \ge 0} \psi_1(u), \tag{6}$$

где функция $\psi_1(u) = -\sum_{i=1}^{k} u_{ii}$, а $K(u) \ge 0$ обозначает неотрицательную определенность семейства симметричных $(kn \times kn)$-матриц вида:

$$K(u) = K_0 + \sum_{i=1}^{k} u_{ii} K_{ii} + \frac{1}{2} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} u_{ij} (K_{ij} + K_{ji}).$$

Здесь матрицы K_0 и K_{ij} состоят из k^2 блоков размером $n \times n$ и задаются следующим образом. По диагонали матрицы K_0 размещены блоки, заданные симметричными $(n \times n)$-матрицами A_i, $i = \overline{1,k}$, а все внедиагональные блоки матрицы K_0 равны нулевым $(n \times n)$-матрицам. K_{ij} – матрица, (i, j)-й блок которой равен единичной $(n \times n)$-матрице, а все остальные блоки равны нулевым $(n \times n)$-матрицам.

Варианту (Z) поставим в соответствие следующую оптимизационную задачу квадратичного типа:

$$Q_2^* = \min \sum_{i=1}^{k} \sum_{j=1}^{n} \sum_{l=1}^{n} a_{ijl} x_{ij} x_{il} \tag{7}$$

при ограничениях

$$\sum_{j=1}^{n} x_{ij}^2 = 1, \ i = \overline{1,k}, \tag{8}$$

$$\sum_{l=1}^{n} x_{il} x_{jl} = 0, \ i = \overline{1, k-1}, \ j = \overline{i+1, k}, \tag{9}$$

$$x_{ij} x_{il} = 0, \ i = \overline{1,k}, \ j = \overline{1, n-1}, \ l = \overline{j+1, n}, \tag{10}$$

$$x_{ij} x_{lj} = 0, \ j = \overline{1,k}, \ i = \overline{1, k-1}, \ l = \overline{i+1, k}, \tag{11}$$

Здесь ограничения (8) и (9) такие же, как и в предыдущей задаче, т. е. задают многообразие Штиффеля $M_{n,k}$ на множестве вещественных чисел. Тот факт, что компоненты векторов должны принадлежать множеству целых чисел, уточняется посредством двух групп дополнительных ограничений. Так группа ограничений (10) обозначает, что для любой пары компонент вектора их произведение должно равняться нулю. Это связано с тем фактом, что для варианта (Z) каждый из векторов $\vec{x}_i$ содержит только одну ненулевую компоненту, равную +1, либо –1. Последнее следует из условий нормировки векторов $\vec{x}_i$, $i=\overline{1,k}$, заданных ограничениями (8), и того, что компоненты векторов должны принадлежать множеству целых чисел. Ограничения (11) означают, что для двух различных векторов компоненты с одним и тем же номером не могут одновременно быть ненулевыми, т. е. ± 1 для разных векторов должны находиться на разных местах. Последнее обусловлено уже не только ортонормированностью, но и взаимной ортогональностью векторов.

Указанные группы ограничений идентифицируют вариант (Z) в том смысле, что множество допустимых решений системы (8)–(11) будет состоять только из целых чисел: 0, +1 и –1. Отметим, что последнее было бы справедливо и тогда, если рассмотреть для варианта (Z) задачу только с одной группой ограничений, т. е. либо (10), либо (11). Однако, согласно [5], включение в систему ограничений функционально избыточных ограничений способствует улучшению нижней двойственной лагранжевой оценки.

Задаче (7)–(11) соответствует двойственная лагранжева оценка:

$$\psi_2^* = \psi_2(u^*, v^*, V^*) = \sup_{u,v,V:K(u,v,V)\geq 0} \psi_2(u,v,V), \qquad (12)$$

где $\psi_2(u,v,V) = \psi_1(u) = -\sum_{i=1}^{k} u_{ii}$.

Здесь $u \in R^m$ – вектор множителей Лагранжа, соответствующих ограничениям (8) и (9); $v \in R^{m_1}$, $m_1 = k\dfrac{n(n-1)}{2}$ – вектор множителей Лагранжа, соответствующих ограничениям (10); $V \in R^{m_2}$, $m_2 = n\dfrac{k(k-1)}{2}$ – вектор множителей Лагранжа, соответствующих ограничениям (11); $K(u,v,V) \geq 0$ обозначает неотрицательную определенность семейства симметричных матриц

$$K(u,v,V) = K(u) + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n-1}\sum_{l=j+1}^{n} v_{ijl}(K_{ijl} + K_{ilj}) + \frac{1}{2}\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\sum_{j=1}^{n} V_{ilj}(\tilde{K}_{ilj} + \tilde{K}_{lij}),$$

где K_{ijl} – матрица размера $nk \times nk$, в i-м диагональном блоке которой содержится $(n \times n)$-матрица с (j,l)-м коэффициентом, равным единице, и всеми остальными коэффициентами, равными нулю; $\tilde{K}_{ilj}$ – матрица размера $nk \times nk$, для которой блок (i,l) содержит $(n\times n)$-матрицу с j-м диагональным коэффициентом, равным единице, и всеми остальными коэффициентами, равными нулю. Все остальные блоки матрицы $\tilde{K}_{ilj}$ равны нулевым $(n\times n)$-матрицам.

Как видим, функция $\psi_2(u,v,V)$ имеет такой же вид как и функция $\psi_1(u)$ для первой задачи. Это обусловлено однородностью добавленных ограничений (10)–(11) и тем, что правая часть в этих ограничениях тождественно равна нулю.

Далее рассмотрим вариант (N). Ему соответствует оптимизационная задача квадратичного типа:

$$Q_3^* = \min \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{l=1}^{n} a_{ijl}x_{ij}x_{il} \tag{13}$$

при ограничениях

$$\sum_{j=1}^{n} x_{ij}^2 = 1,\ i=\overline{1,k}, \tag{14}$$

$$\sum_{l=1}^{n} x_{il}x_{jl} = 0,\ i=\overline{1,k-1},\ j=\overline{i+1,k}, \tag{15}$$

$$x_{ij}x_{il} = 0,\ i=\overline{1,k},\ j=\overline{1,n-1},\ l=\overline{j+1,n}, \tag{16}$$

$$x_{ij}x_{lj} = 0,\ j=\overline{1,k},\ i=\overline{1,k-1},\ l=\overline{i+1,k}, \tag{17}$$

$$x_{ij}^2 - x_{ij} = 0,\ i=\overline{1,k},\ j=\overline{1,n}. \tag{18}$$

Задача (13)–(18) отличается от задачи (7)–(11) тем, что к последней добавлены булевые ограничения (18) на неизвестные компоненты векторов $\vec{x}_i$, $i=\overline{1,k}$. Эти ограничения исключают из множества допустимых точек $M_{n,k}$ векторы с компонентами, равными –1, и, соответственно, компоненты векторов могут быть равны либо нулю, либо единице. Следовательно, среди допустимых решений системы ограничений (14)–(18) будут векторы, содержащие только одну компоненту равную единице, а уже какая конкретно компонента будет определяться глобальным экстремумом задачи (13)–(18).

Задача (13)–(18) существенно отличается от двух предыдущих. В обеих предыдущих задачах квадратичные функции, задающие как функцию цели так и ограничения, были однородными и не содержали линейных членов. В последней же задаче булевые ограничения приводят к линейным членам. В силу этого нахождение двойственной лагранжевой оценки для задачи (13)–(18) будет сложнее и потребует решения системы линейных уравнений.

Пусть $w\in R^{m_3}$, ($m_3 = nk$) – вектор множителей Лагранжа, соответствующий ограничениям (18), а $u\in R^{m}$, $v\in R^{m_1}$, $V\in R^{m_2}$ – векторы множителей Лагранжа для ограничений (14)–(15), (16) и (17), соответственно.

Задаче (18)–(23) соответствует такая двойственная лагранжева оценка:

$$\psi_3^* = \sup_{u,v,V,w:K(u,v,V,w)\geq 0} \psi_3(u,v,V,w),$$

где

$$\psi_3(u,v,V,w) = \inf_{x\in R^{kn}} \left\{ \big(K(u,v,V,w)x,x\big) - \sum_{i=1}^{k}\sum_{j=1}^{n} w_{ij}x_{ij} - \sum_{i=1}^{k} u_{ii} \right\}.$$

Здесь $K(u,v,V,w)\geq 0$ обозначает неотрицательную определенность семейства симметричных матриц

$$K(u,v,V,w) = K(u,v,V) + diag\{w_{11},...,w_{1n},...,w_{k1},...,w_{kn}\}.$$

Когда ψ_3^* достигается внутри области положительной определенности семейства матриц $K(u,v,V,w)$ (т. е. $K(u^*,v^*,V^*,w^*)>0$), точку глобального экстремума x^* для задачи (13)–(18) можно найти по следующей формуле

$$x^* = (x_{11}^*,...,x_{1n}^*,...,x_{k1}^*,...,x_{kn}^*)^T = \frac{1}{2}K^{-1}(u^*,v^*,V^*,w^*)w^*.$$

В данном случае оценка ψ_3^* будет точной нижней оценкой для Q_3^* (т. е. $\psi_3^* = Q_3^*$).

Эффективность двойственных оценок ψ_1^*, ψ_2^*, ψ_3^* исследовалась на ряде тестовых примеров (для их нахождения использовалась программа DSQTPR [6] с использованием модификации r-алгоритма [7]). Ниже приведем два из них.

Пример 1. Рассмотрим задачу (1)–(2), где $A_1 = diag(-1,2,3)$, $A_2 = diag(4,-5,6)$, $A_3 = diag(7,8,-9)$. Целевая функция имеет вид

$$f(x) = -x_{11}^2 + 2x_{12}^2 + 3x_{13}^2 + 4x_{21}^2 - 5x_{22}^2 + 6x_{23}^2 + 7x_{31}^2 + 8x_{32}^2 - 9x_{33}^2,$$

множество допустимых решений состоит из $8 = 2^3$ точек следующего вида: $x^* = ((\pm1,0,0)^T,(0,\pm1,0)^T,(0,0,\pm1)^T)$. Оптимальное значение целевой функции равно сумме отрицательных компонент матриц A_i, т. е. $f^* = f(x^*) = -15$. Следовательно, для вариантов (R) и (Z) имеем $Q_1^* = Q_2^* = -15$. Для варианта (N) также $Q_3^* = -15$, но решением будет единственная точка $x^* = ((1,0,0)^T,(0,1,0)^T,(0,0,1)^T)$.

Для первого примера все лагранжевые двойственные оценки являются точными, т. е. $\psi_1^* = -15 = Q_1^*$, $\psi_2^* = -15 = Q_2^*$, $\psi_3^* = -15 = Q_3^*$. Кроме того, для варианта (N) программа DSQTPR обеспечивает нахождение точного оптимального решения задачи (18)–(23) по переменным x^*. Для этого ей потребовалось 208 итераций r-алгоритма, и, следовательно, затраты на нахождение и строгое обоснование точки глобального минимума для варианта (N) небольшие (время счета $\approx$ 1 сек на компьютере IBM PC/AT/486DX/66MHz).

Недостаток первого примера состоит в том, что глобальный минимум для варианта (N) однозначен и значение целевой функции в нем сильно отличается от значений целевой функции в других допустимых точках. Поэтому ниже опишем второй, более сложный пример, когда глобальный минимум для варианта (N) неоднозначен.

Пример 2. Рассмотрим задачу (1)–(2) с матрицами $A_1 = diag(1,2,3)$, $A_2 = diag(4,5,6)$, $A_3 = diag(7,8,9)$, т. е. целевая функция имеет вид

$$f(x) = x_{11}^2 + 2x_{12}^2 + 3x_{13}^2 + 4x_{21}^2 + 5x_{22}^2 + 6x_{23}^2 + 7x_{31}^2 + 8x_{32}^2 + 9x_{33}^2.$$

Для данного примера $Q_1^* = Q_2^* = Q_3^* = 15$, а множество допустимых решений будет иметь следующий вид. Для варианта (N) имеем шесть допустимых решений:

$$x_1^* = ((1,0,0)^T, (0,1,0)^T, (0,0,1)^T);\ x_2^* = ((1,0,0)^T, (0,0,1)^T, (0,1,0)^T);$$

$$x_3^* = ((0,1,0)^T, (1,0,0)^T, (0,0,1)^T);\ x_4^* = ((0,1,0)^T, (0,0,1)^T, (1,0,0)^T);$$

$$x_5^* = ((0,0,1)^T, (1,0,0)^T, (0,1,0)^T);\ x_6^* = ((0,0,1)^T, (0,1,0)^T, (1,0,0)^T).$$

Для варианта (Z) количество допустимых решений будет равно $6 \cdot 2^3 = 48$, где решения будут отличаться тем, что на соответствующих единице местах могут находиться как единица, так и минус единица. Множество допустимых решений для варианта (R) содержит бесконечное число точек.

Для второго примера лагранжевые двойственные оценки $\psi_1^* = 12$ и $\psi_2^* = 12$, и ни одна из них не является точной нижней оценкой для Q_1^* и Q_2^*. Заметим, что разрыв двойственности $\Delta_i = Q_i^* - \psi_i^*$, $i = 1,2$, равен трем и есть достаточно большим ($\approx 20\%$). Оценка $\psi_3^* = 15$ является точной. Из-за неоднозначности решений программа DSQTPR не гарантирует нахождения оптимального решения задачи (18)–(23) по переменным x^*. Для того, чтобы получить одно из решений требуется соответственно "возмутить" коэффициенты матриц A_i, $i = 1,2,3$. Так, например, для "возмущенных" матриц $\overline{A}_1 = diag(0.9999,2,3)$, $\overline{A}_2 = diag(4,4.9999,6)$, $\overline{A}_3 = diag(7,8,9)$ получим решение x_1^*, а для "возмущенных" матриц $\overline{A}_1 = diag(1,2,3)$, $\overline{A}_2 = diag(4,4.9999,6)$, $\overline{A}_3 = diag(6.9999,8,9)$, получим решение x_6^*. Для этого программе DSQTPR потребовалось 499 итераций r-алгоритма в первом случае и 487 итераций во втором.

В заключение отметим, что описанный в работе аппарат нахождения двойственных оценок с использованием функционально избыточных ограничений можно перенести на случай, когда квадратичная функция цели имеет общий вид (при этом ограничения остаются неизменными). То есть предлагаемый подход для получения нижних оценок функционала (а в некоторых случаях и оптимальных значений компонент вектора переменных) возможно использовать для оценок NP-трудных задач, например, в квадратичной задаче о назначениях.

Н.З. Шор, П.І. Стецюк, О.А. Березовський

ДВОЇСТІ ОЦІНКИ ДЛЯ СПЕЦІАЛЬНОЇ ОПТИМІЗАЦІЙНОЇ ЗАДАЧІ КВАДРАТИЧНОГО ТИПУ НА МНОЖИНІ ШТИФФЕЛЯ

Розглядається задача глобальної оптимізації, пов'язана з знаходженням мінімуму квадратичної однорідної функції спеціального вигляду на множині Штиффеля. Розглянуто лагранжеві двоїсті оцінки для трьох варіантів цієї задачі, які розрізняються обмеженнями на компоненти ортонормованих векторів. На тестових прикладах показано, що їх можна ефективно використовувати при знаходженні глобального мінімуму.

N.Z. Shor, P.I. Stetsyuk, O.A. Berezovskyi

DUAL BOUNDS FOR SPECIAL OPTIMIZATION QUADRATIC TYPE PROBLEM ON STIEFEL MANIFOLDS

We study the global optimization problem connected to finding of a minimum of homogeneous quadratic function of a special kind on Stiefel manifold. It's considered Lagrange dual bounds for three variants of this problem which differ with restrictions on components orthonormal vectors. It's shown on tests that they can effectively be used for finding of a global minimum.

1. *Rapcsak T.* On minimization on Stiefel manifold // European J. of Operational Research.– 2002. – 143. – P. 365–376.
2. *Balogh J.,Csendes T.,Rapcsak T.* Global optimization problems on Stiefel manifold // NMCM–2002 Book of Abstracts, Miscolc, Hungary, 2002. – P. 19–21.
3. *Corliss G.F., Kearfott R.B.* Rigorous global search: Industrial applications // In T. Csendes (editor): Development in Reliable Computing, Kluwer, Dordrecht, 1999. – P. 1–16.
4. *Kearfott R.B.* Rigorous Global Search: Continuous Problems, Kluwer, Dordrecht, 1996. – 98 p.
5. *Shor N.Z.* Nondifferentiable Optimization and Polynomial Problems. – Dordrecht, Kluwer, 1998. – 394 p.
6. *Shor N.Z., Stetsyuk P.I.* Dual Solution of Quadratic-Type Problems by r-algorithm (subroutine DSQTPr) // Abstracts of Second International Workshop "Recent Advances in Non-Differentiable Optimization", (October, 1-4, 2001, Kyiv, Ukraine). – P. 36.
7. *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.

Получено 06.08.2004