УДК 518.9

ПРОЕКЦИОННЫЕ МЕТОДЫ РЕШЕНИЯ НЕЛИНЕЙНЫХ ВАРИАЦИОННЫХ НЕРАВЕНСТВ

В.М. ПАНИН, Т.В. ЛАВРИНА

Рассматриваются конечномерные вариационные неравенства с сильно монотонным оператором и численные методы их решения. Изучаются проекционные методы первого порядка с линеаризацией ограничений, основанные на аппарате квадратичного программирования. Установлена нелокальная сходимость к решению и линейная скорость сходимости в его окрестности.

ПОСТАНОВКА ЗАДАЧИ

Вариационные неравенства (в.н.) включают в себя задачи оптимизации, дополнительности, отыскания седловых точек, решение нелинейных уравнений. Возросшее к ним внимание объясняется множеством моделей экономического равновесия, описываемых с помощью в.н. Наиболее полно численные методы решения конечномерных в.н. представлены в $[1-3]$, описание равновесных моделей приведено в $[3]$. Настоящая статья является обобщением работ $[4, 5]$ по численному решению в.н.

Решением в.н. $\langle F(y), x-y)\rangle \ge 0$ на множестве $\Omega \subset R^n$ называется точка $x_* \in \Omega$ такая, что

$$\langle F(x_*),\, x-x_*\rangle \ge 0, \quad \forall x \in \Omega\,, \tag{1}$$

где $\langle , \rangle$ — скалярное произведение векторов в R^n; $F(x)$: $R^n \to R^n$.

Далее $\Omega = \left\{ x \in R^n \mid g_i(x) \le 0,\ i=1,\ldots,l \right\}$ — ограниченное множество, содержащее внутреннюю точку $\bar{x}$: $g_i(\bar{x}) \le -\delta < 0$, $i=1,\ldots,l$; функции $g_i(x)$ и $F(x)$ достаточно гладкие, при этом $g_i(x)$ выпуклы, а $F(x)$ — сильно монотонный оператор: для любых $x, y \in R^n$

$$m\|x\|^2 \le \langle F_x(y)x, x\rangle \le M\|x\|^2, \quad M \ge m > 0\,. \tag{2}$$

Сформулированную задачу назовем в.н. (1), (2). Всюду векторы x, λ, $F(x)$, $g(x) = \{g_i(x)\}_{i=1}^{l}$, $g_i^{'}(x)$ рассматриваются как вектор-столбцы; I_n — единичная $(n \times n)$-матрица; верхний индекс T означает транспонирование; C, C_i, $\bar{\alpha}$, $\bar{\alpha}_i$ — стандартные обозначения положительных констант. Производная по аргументу u вектор-функции $f(u)$: $R^s \to R^t$ обозначается $f_u(u)$, что согласуется с общепринятым обозначением якобиана вектор-функции $f(u)$ как матрицы $(t \times s)$ со строками $f_{iu}(u)$. Для скалярной функции $f(u): R^s \to R^1$ запись $f_u(u)$ означает вектор-строку, а $f'(u)$ — вектор-столбец: $f'(u) = f_u^T(u)$.

Решение x_* в.н. (1), (2) единственно, а необходимые условия первого порядка являются также достаточными и имеют вид [1–3]:

$$F(x_*) + g_x^T(x_*)\lambda_* = 0, \quad \lambda_*^i g_i(x_*) = 0, \quad i = 1, \ldots, l, \tag{3}$$

где $\lambda_* \in \Lambda^+ = \left\{ \lambda \in R^l \mid \lambda^i \geq 0, \; i = 1, \ldots, l \right\}$; Λ_* — множество векторов Лагранжа λ_*.

В существующих методах численного решения в.н. (1), (2) – так называемых методах оптимизационного подхода – приближения к решению определяются как $x_{k+1} = \bar{x}_k$, где $\bar{x}_k$ — решение вспомогательной задачи оптимизации

$$\min \left\langle F(x_k) + \frac{1}{2} H_k (x - x_k), x - x_k \right\rangle, \quad x \in \Omega \tag{4}$$

с \`симметричной матрицей $H_k = H_k^T$ такой, что для любых $p \in R^n$

$$a \|p\|^2 \leq \langle H_k p, p \rangle \leq A \|p\|^2, \quad A \geq a > 0, \quad k = 0, 1, \ldots, \tag{5}$$

A и a — константы.

К методам оптимизационного подхода с различным выбором матрицы $H_k = H_k^T$ можно отнести следующиех [1–3]:

$H_k = H$ — проекционные методы, H — постоянная матрица ;

$H_k = D_k$ — линеаризованный метод Якоби, D_k — диагональная часть матрицы $F_x(x_k)$;

$H_k = (F_x(x_k) + F_x^T(x_k))/2$ — симметричный метод Ньютона.

Нелокально сходящимися среди них являются только проекционные методы, в которых $\bar{x}_k$ — решение вспомогательной задачи

$$\min \left\langle \alpha F(x_k) + \frac{1}{2} H (x - x_k), \, x - x_k \right\rangle, \quad x \in \Omega, \tag{6}$$

а значение $\alpha > 0$ постоянно и должно удовлетворять требованию $\left\| I_n - \alpha H^{-1/2} F_x(x_k) H^{-1/2} \right\| < 1$ [3]. Скорость сходимости к решению проекционных методов является линейной, т.е. не ниже убывающей геометрической прогрессии [3].В частности, таким алгоритмом является известный метод проекции градиента $x_{k+1} = \Pi_\Omega(x_k - \alpha F(x_k))$, в котором $H_k = I_n$, а выбор α подчинен условию $\| I_n - \alpha F_x(x_k) \| < 1$, Π_Ω — оператор проектирования на Ω.

С точки зрения реализации такие проекционные методы обладают двумя существенными недостатками.

1. Если функции $g_i(x)$ нелинейны, то вспомогательную задачу (6) нельзя решить за конечное число вычислений, и линейная скорость сходимости является нереализуемой.

2. Для определения α необходимо вычислять не только матрицу $H^{-1/2}$, но и матрицу $F_x(x_k)$ во всех точках $x_k \in \Omega$ (или оценить константу Липшица для $F(x)$ на Ω), что, как правило, нереально.

Эти недостатки отсутствуют в алгоритме проекционного типа с матрицей $H_k = I_n$, предложенном Пшеничным в [4] .Задачи (4) или (6) заменены в нем на задачу квадратичного программирования (з.к.п.)

$$\min_x \left\langle F(x_k) + \frac{1}{2} H_k (x - x_k), x - x_k \right\rangle, \quad x \in \Omega_k, \tag{7}$$

$$\Omega_k = \left\{ x \mid \quad g_i(x_k) + \left\langle g_i'(x_k), x - x_k \right\rangle \le 0, \ \ i = 1, \ldots, l \right\},$$

которая, в отличие от (4), (6) может быть точно решена за конечное число вычислений стандартными средствами. Алгоритм [4] строит последовательность $x_{k+1} = x_k + \alpha_k (\bar{x}_k - x_k)$. Значение α_k определяется в нем, исходя из условия релаксации (убывания) некоторой штрафной функции за конечное число вычислений $F(x)$, $g_i(x)$, $g_i'(x)$, не прибегая к вычислению матриц $F_x(x)$, $g_i''(x)$ и к их оценке в области Ω (или, что то же, не требуя задания константы Липшица для $F(x)$, $g_i'(x)$). Однако скорость сходимости $x_k \to x_*$ в [4] не установлена.

Приведем общий подход к построению эффективных численных методов проекционного типа со вспомогательной задачей (7). Отметим, что требование симметричности H_k в (7) не является ограничительным, поскольку при несимметричной H_k вместо задачи (7) можно рассматривать эквивалентную ей с матрицей $(H_k + H_k^T)/2$, которая симметрична.

ОПТИМИЗАЦИОННЫЙ ПОДХОД И ЭКВИВАЛЕНТНАЯ ЗАДАЧА ОПТИМИЗАЦИИ

Эквивалентной оптимизационной задачей (э.о.з.) для в.н. (1), (2) назовем такую задачу оптимизации

$$\min \Theta(x), \quad x \in \Omega, \tag{8}$$

которая среди множества своих решений (стационарных точек) содержит решения (стационарные точки) исходного в.н.

В качестве целевой функции $\Theta(x)$ в [6] предложена следующая:

$$\Theta(x) = \max_{y \in \Omega} \left\langle -F(x) - \frac{1}{2} H(x)(y-x), y-x \right\rangle . \tag{9}$$

Очевидны трудности численного решения э.о.з. (8), (9) , когда ограничения в (1) нелинейны, поскольку функцию $\Theta(x)$ нельзя вычислить точно, не говоря уже об отыскании ее производных по направлению в точках итерационного процесса.

Построим э.о.з., свободную от недостатков э.о.з. (8), (9). Заменим з.к.п. (7) в текущей точке $x_k = x$ эквивалентной ей:

$$\max_p \left[-\langle F(x), p \rangle - \frac{1}{2} \langle H(x)p, p \rangle \mid g(x) + g_x(x)p \le 0 \right]. \tag{10}$$

Ее решение $p(x)$ единственно, поскольку $H(x)$ удовлетворяет аналогу (5). Пусть $\lambda(x)$ — двойственное решение задачи (10) и

$$\Theta(x) = -\langle F(x), p(x) \rangle - \frac{1}{2} \langle H(x)p(x), p(x) \rangle \text{ —} \tag{11}$$

ее оптимальное решение.

Покажем., что задача (8), (11) является э.о.з. для в.н. (1), (2). Зафиксируем точку x и рассмотрим функцию Лагранжа для (10):

$$W(x,p,\lambda) = -\left\langle F(x) + \frac{1}{2} H(x)p,\ p \right\rangle - \langle \lambda,\ g(x) + g_x(x)p \rangle .$$

По теореме Куна-Таккера

$$\max_p \min_{\lambda \in \Lambda^+} W(x,p,\lambda) = \Theta(x) = W(x, p(x), \lambda(x)) = \min_{\lambda \in \Lambda^+} \max_p W(x,p,\lambda) . \tag{12}$$

Для текущего $\lambda \in \Lambda^+$ обозначим $\varphi(x,\lambda) = \max\limits_p W(x,p,\lambda) = W(x, p(x,\lambda), \lambda)$, где $p(x,\lambda) = \arg\max\limits_p W(x,p,\lambda)$. Согласно правому неравенству (12), производная $W_p(x, p(x,\lambda), \lambda) = 0$, т.е.

$$F(x) + H(x)p(x,\lambda) + g_x^T(x)\lambda = 0 . \tag{13}$$

Использование (13) в выражении для $W(x,p,\lambda)$ дает

$$\varphi(x,\lambda) = \frac{1}{2} \langle H(x)p(x,\lambda), p(x,\lambda) \rangle - \langle \lambda, g(x) \rangle . \tag{14}$$

Получим другое представление для $\varphi(x,\lambda)$. Введем обозначение $L_x(x,\lambda) \equiv F(x) + g_x^T(x)\lambda$. Оно является исключением из принятого выше правила обозначений, поскольку, во-первых, $L_x(x,\lambda)$ — вектор-столбец, а, во-вторых, нижний индекс x здесь не означает, вообще говоря, дифференцирования по x. На основании (13) $p(x,\lambda) = -H^{-1}(x)L_x(x,\lambda)$, и вместо (14) можно записать

$$\varphi(x,\lambda) = \frac{1}{2}\left\langle H^{-1}(x)L_x(x,\lambda), L_x(x,\lambda)\right\rangle - \left\langle \lambda, g(x)\right\rangle . \qquad (15)$$

Согласно правому неравенству (12), $\Theta(x) = \min\limits_{\lambda} \varphi(x,\lambda), \quad \lambda \in \Lambda^+$. Этот минимум достигается в точке $\lambda = \lambda(x)$, поскольку в соответствии с (12) по определению седловой точки $\{p(x),\ \lambda(x)\}$ имеет место $\Theta(x) = \max\limits_{p} W(x,p,\lambda(x)) \equiv \varphi(x,\ \lambda(x))$. Таким образом,

$$\Theta(x) = \min_{\lambda\in\Lambda^+} \varphi(x,\ \lambda) = \varphi(x,\ \lambda(x)), \qquad (16)$$

и задача оптимизации (8), (11) превращается в следующую: найти

$$\min_{x\in\Omega}\ \min_{\lambda\in\Lambda^+} \varphi(x,\lambda) = \min_{x\in\Omega}\ \min_{\lambda\in\Lambda^+}\left[\frac{1}{2}\left\langle H^{-1}(x)L_x(x,\lambda), L_x(x,\lambda)\right\rangle - \left\langle \lambda,\ g(x)\right\rangle\right], \qquad (17)$$

в которой внутренний минимум по λ достигается, согласно (16), на решениях $\lambda(x)$ двойственной к (10) задачи.

Задача (17) эквивалентна в.н. (1), (2). Действительно, в решениях $\hat{x} \in \Omega$ задачи (17) все слагаемые функции $\varphi(x,\lambda)$ (15) неотрицательны и равны нулю, что означает выполнение условий (3), т.е. $\hat{x} = x_*$. Обратное очевидно.

Э.о.з. (17) отличается от э.о.з. (8), (9) возможностью ее эффективного численного решения. Для того, чтобы убедиться в этом, изучим ее свойства.

Зафиксируем x_k, $H_k = H_k^T$ в (7) и пусть $\bar{x}_k$, $\lambda_k = \lambda(x_k)$ — ее прямое и двойственное решения, $p_k = \bar{x}_k - x_k$. Поскольку $\min \varphi\ (x_k,\ \lambda)$ по $\lambda \in \Lambda^+$ достигается в точке λ_k, то для решения э.о.з. (17) можно применить схему декомпозиции, минимизируя $\varphi(x,\lambda)$ сначала по $\lambda \in \Lambda^+$ при фиксированном x_k, а затем — по x при фиксированном λ_k. Считая λ_k найденным, рассмотрим задачу $\min \varphi\ (x,\lambda_k),\ \ x \in \Omega$. Для нее недифференцируемой штрафной функцией является, как известно, функция

$$\Phi_N(x,\lambda_k) = \varphi(x,\lambda_k) + N\, g^+(x), \qquad (18)$$

где $g^+(x) = \max\{0,\ g_1(x),\ldots, g_l(x)\}$, $N > 0$ — коэффициент штрафа. Обозначим $\partial\Phi_N(x_k,\ \lambda_k)\,/\,\partial p_k$ ее производную по направлению p_k в точке x_k; $z_k^T = \left[x_k^T, \lambda_k^T\right]$, $-(k) = \{i \mid\ g_i(x_k) \le 0\}$, $+(k) = \{i \mid\ g_i(x_k) > 0\}$.

Лемма 1. Имеет место оценка

$$\partial\Phi_N(z_k)\,/\,\partial p_k \le -\psi_k\,, \qquad (19)$$

где

$$\psi_k = \psi_k^{(1)} + \psi_k^{(2)}, \quad \psi_k^{(1)} = \langle L_{xx}(z_k) p_k, p_k \rangle,$$

$$\psi_k^{(2)} = -\sum_{-(k)} \lambda_k^i g_i(x_k) + \left[N g^+(x_k) - \sum_{+(k)} \lambda_k^i g_i(x_k) \right].$$

Доказательство. Из необходимых условий экстремума для з.к.п. (7) следует

$$p_k = -H_k^{-1} L_x(z_k), \quad \langle \lambda_k, g(x_k) \rangle = -\langle g_x(x_k) p_k, \lambda_k \rangle. \tag{20}$$

Тогда, с учетом симметричности матрицы H_k^{-1},

$$\frac{1}{2} \frac{\partial \left\langle H_k^{-1} L_x(z_k), L_x(z_k) \right\rangle}{\partial p_k} = \left\langle L_{xx}^T(z_k) H_k^{-1} L_x(z_k), p_k \right\rangle = -\langle L_{xx}(z_k) p_k, p_k \rangle,$$

$$\frac{\partial \langle -\lambda_k, g(x_k) \rangle}{\partial p_k} = \langle \lambda_k, g(x_k) \rangle = \sum_{-(k)} \lambda_k^i g_i(x_k) + \sum_{+(k)} \lambda_k^i g_i(x_k),$$

где слева фигурируют производные по направлению p_k.

Оценим $N \partial g^+(x_k) / \partial p_k$. Пусть $R_k = \left\{ i \in 0, 1, \ldots, l \mid g_i(x_k) = g^+(x_k) \right\}$, где $g_0(x) \equiv 0$. Так как $\langle g_i'(x_k), p_k \rangle \le -g_i(x_k), \quad i = 0, 1, \ldots, l$, то

$$N \partial g^+(x_k) / \partial p_k = N \max_{i \in R_k} \langle g_i'(x_k), p_k \rangle \le N \max_{i \in R_k} (-g_i(x_k)) = -N g^+(x_k). \tag{21}$$

Суммируя (21) с предыдущими производными по направлению, приходим к (19). Лемма доказана.

Связь между $\Phi_N(z_k)$ и ψ_k устанавливает

Лемма 2. Пусть симметричная матрица H_k в (7) удовлетворяет (5). Тогда

$$\psi_k \ge \gamma^{-1} \Phi_N(z_k), \quad \gamma = \max \left\{ 1; A(2m)^{-1} \right\}. \tag{22}$$

Доказательство. Из (2) вытекает

$$m \|p_k\|^2 \le \langle L_{xx}(z_k) p_k, p_k \rangle \le \overline{M} \|p_k\|^2, \quad \overline{M} \ge M, \tag{23}$$

причем величина $\overline{M}$ ограничена, если ограничена последовательность $\{z_k\} \subset R^n \times \Lambda^+$. На основании первого равенства (20) $\left\langle H_k^{-1} L_x(z_k), L_x(z_k) \right\rangle =$ $= \langle H_k p_k, p_k \rangle$ и, используя последовательно правое неравенство из (5) и левое — из (23), получим

$$\left\langle H_k^{-1} L_x(z_k), L_x(z_k) \right\rangle \le A\|p_k\|^2 \le Am^{-1}\left\langle L_{xx}(z_k)\,p_k, p_k \right\rangle = Am^{-1}\psi_k^{(1)} .$$

Поэтому из выражений для ψ_k, $\Phi_N(z_k)$ следует

$$\Phi_N(z_k) = \frac{1}{2}\left\langle H_k^{-1} L_x(z_k), L_x(z_k) \right\rangle + \psi_k^{(2)} \le A(2m)^{-1}\psi_k^{(1)} + \psi_k^{(2)} \le$$

$$\le \left[\psi_k^{(1)} + \psi_k^{(2)}\right] \max\left\{1;\, A(2m)^{-1}\right\} = \gamma\psi_k .$$

Лемма доказана.

Из нее и из леммы 1 вытекает

Следствие. Если N удовлетворяет условию

$$N \ge \sum_{+(k)} \lambda_k^i , \tag{24}$$

то

$$\partial\Phi_N(z_k)/\partial p_k \le -\gamma^{-1}\Phi_N(z_k) \le 0 . \tag{25}$$

Действительно, условия (23), (24) влекут $\psi_k \ge 0$, и (25) следует из (19), (22).

ОБЩИЕ СВОЙСТВА МЕТОДОВ ОПТИМИЗАЦИОННОГО ПОДХОДА

Предварительно приведем общие свойства любого метода, использующего следующее аппроксимирующее в.н.:

$$\left\langle F(x_k) + H_k(\bar{x}_k - x_k),\, x - \bar{x}_k \right\rangle \ge 0 , \quad \forall x \in \Omega_k , \tag{26}$$

в котором матрица H_k удовлетворяет (5), но необязательно симметрична (напомним, что $\bar{x}_k$ — решение в.н. (26)). Необходимые условия первого порядка для него имеют вид

$$H_k\, p_k + L_x(z_k) = 0 , \quad \lambda_k^i\left[g_i(x_k) + \left\langle g_i'(x_k), p_k \right\rangle\right] = 0, \quad i = 1,\ldots,l, \tag{27}$$

где $\lambda_k^i \ge 0$, $p_k = \bar{x}_k - x_k$. Отметим, что условия оптимальности для з.к.п. (7) являются частным случаем в.н. (26) (системы (27)) при $H_k = H_k^T$.

Лемма 3. Пусть $\{x_k\}$ — ограниченная последовательность, матрица H_k удовлетворяет (5). Тогда решения в.н. (26) ограничены: $\|p_k\| \le C_1$ $\|\lambda_k\| \le C_2$.

Доказательство. Так как функции $g_i(x)$ выпуклы, то $\bar{x} \in \Omega \in \Omega_k$. Полагая в (26) $x = \bar{x}$ и учитывая $\bar{x}_k = x_k + p_k$, получаем

$$\left\langle H_k\, p_k, p_k \right\rangle \le \left\langle F(x_k),\, \bar{x} - x_k - p_k \right\rangle + \left\langle H_k\, p_k,\, \bar{x} - x_k \right\rangle .$$

Вместе с (5) это влечет

$$a\|p_k\|^2 \le \left[\|F(x_k)\| + A\|\bar{x}-x_k\|\right]\|p_k\| + C ,$$

откуда $\|p_k\| \le C_1$, $\|\bar{x}_k\| \le C$.

Скалярно умножая первое равенство (27) на $\bar{x}_k - \bar{x}$ и учитывая остальные равенства (27), получаем

$$\left\langle F(x_k) + H_k p_k , \bar{x}_k - \bar{x} \right\rangle = -\left\langle g_x^T(x_k)\lambda_k , p_k + x_k - \bar{x} \right\rangle =$$

$$= \left\langle \lambda_k , g(x_k) + g_x(x_k)(\bar{x} - x_k) \right\rangle \le \sum_{i=1}^{l} \lambda_k^i g_i(\bar{x}) \le -\delta \sum_{i=1}^{l} \lambda_k^i . \qquad (28)$$

Отсюда следует $\|\lambda_k\| \le C_0 \|\bar{x}_k - \bar{x}\| \|F(x_k) + H_k p_k\| \le C_2$.

Лемма доказана.

Лемма 3 не использует симметричности матрицы H_k и относится к аппроксимирующему в.н. (26). В частном случае, когда в качестве (26) выступают условия оптимальности з.к.п. (7) с матрицей $H_k = I_n$, ограниченность λ_k установлена в [4]. Результат [4] несложно обобщить на случай $H_k \ne I_n$. Для полноты изложения приведем это обобщение.

Лемма 4. Пусть $\{x_k\}$ — ограниченная последовательность и матрица $H_k = H_k^T$ в (7) удовлетворяет (5). Тогда $\|\lambda_k\| \le C_1$, $\|p_k\| \le C_2$.

Доказательство. При фиксированном $x = x_k$ на основании (16), (12)

$$\Theta(x_k) = \varphi(z_k) = \max_p W(x_k, p, \lambda_k) \ge W(x_k, \bar{x} - x_k, \lambda_k) =$$

$$= \left\langle -F(x_k) - \frac{1}{2} H_k(\bar{x} - x_k), \bar{x} - x_k \right\rangle - \sum_{i=1}^{l} \lambda_k^i \left[g_i(x_k) + \left\langle g_i'(x_k), \bar{x} - x_k \right\rangle \right]. \qquad (29)$$

По аналогии с (28) в последней сумме $g_i(x_k) + \left\langle g_i'(x_k), \bar{x} - x_k \right\rangle \le -\delta$. Очевидно также, что для величины $\Theta(x)$, определяемой (11), справедливо

$$\Theta(x_k) \le \max_p \left[-\left\langle F(x_k), p \right\rangle - \frac{1}{2}\left\langle H_k p, p \right\rangle \right] = \frac{1}{2}\left\langle H_k^{-1} F(x_k), F(x_k) \right\rangle .$$

Так как матрица H_k^{-1} , как и H_k , симметрична и удовлетворяет аналогу (5), то отсюда следует $\Theta(x_k) \le C$. Теперь из (29) легко вытекает $\|\lambda_k\| \le C_1$, тогда ограниченность $\|p_k\| \le C_2$ является следствием первого равенства (20).

Лемма доказана.

Сформулируем критерий сходимости алгоритмов с аппроксимирующим в.н. (26) и отдельно — для его частного случая, когда используется з.к.п.(7) с матрицей $H_k = H_k^T$.

Предложение 1. Если $\{x_k\}$ — ограниченная последовательнось и матрица H_k в (26) удовлетворяет (5), то условие $p_k \to 0$ $(p_k = 0)$ необходимо и достаточно для того, чтобы $x_k \to x_*$ $(x_k = x_*)$.

Оно вытекает из леммы 3 и из условий (3), являющихся необходимыми и достаточными условиями первого порядка.

Предложение 2. Если $\{x_k\}$ — ограниченная последовательность, матрица $H_k = H_k^T$ в з.к.п. (7) удовлетворяет (5), а коэффициент N — требованию (24), то каждое из условий $\psi_k \to 0$, $\Phi_N(z_k) \to 0$ необходимо и достаточно для того, чтобы $x_k \to x_*$.

Оно легко следует из условий (3) и того факта, что все составляющие величин $\Phi_N(z_k)$, ψ_k неотрицательны вследствие (23), (24) и (5).

ПРОЕКЦИОННЫЙ МЕТОД: ФОРМУЛИРОВКА, СВОЙСТВА

Пусть x_0 — произвольное начальное приближение из R^n; $\varepsilon \in (0;1)$ и C_g — любые константы, $C_g > g^+(x_0)$, $H_k = H$ — постоянная симметричная матрица, удовлетворяющая (5). Опишем k-ю итерацию метода $x_{k+1} = x_k + \alpha_k p_k$, $k = 0,1,\ldots$, считая точку x_k и значение N_{k-1} найденными.

Шаг 1. Решить задачу (7) при $H_k = H$ и найти $p_k = \bar{x}_k - x_k$ и λ_k. Если $p_k = 0$, то $x_k = x_*$ — процесс закончен; если $p_k \neq 0$ — переход к шагу 2.

Шаг 2. Найти коэффициент штрафа N_k:

$$N_k = \max\left\{ N_{k-1}\,,\ 2\sum_{i\in+(k)} \lambda_k^i \right\},\quad N_0 = 2\sum_{i\in+(0)} \lambda_0^i\,.$$

Шаг 3. Определить α_k как наибольшее из чисел $\alpha = 2^{-s}$, $s = 0,1,\ldots$, при котором одновременно начинают выполняться неравенства

$$g^+(x_k + \alpha p_k) \le C_g\,, \tag{30}$$

$$\Phi_N(x_k + \alpha p_k, \lambda_k) \le (1 - \varepsilon\alpha^2)\,\Phi_N(x_k, \lambda_k)\,, \tag{31}$$

где N — текущее N_k. Найти $x_{k+1} = x_k + \alpha_k p_k$; конец итерации.

Свойства алгоритма устанавливает

Теорема 1. Сформулированный метод сходится к решению в.н. (1), (2) по прямым и двойственным переменным, $x_k \to x_*$ $\lambda_k \to \Lambda_*$. Каждая итерация метода осуществляется за равномерно ограниченное (по k) число вычислений и $\alpha_k \ge \bar{\alpha} > 0$. При $k \ge k_0$ значение $N_k = N$ постоянно и справедливы оценки линейной скорости сходимости:

$$\Phi_N(x_k, \lambda_k) \le C_0 q^k, \quad \|p_k\| \le C_1 q_1^k, \quad \|x_k - x_*\| \le C_2 q_1^k, \tag{32}$$

где $q \in (0;1)$ — константа, $q_1 = q^{1/2}$.

Доказательство. Сначала проведем обоснование правила выбора α_k в алгоритме. Имеем

$$g^+(x_k + \alpha p_k) = \max_{i=0,\dots,l} \left[g_i(x_k) + \alpha \left\langle g_i'(x_k), p_k \right\rangle + \frac{\alpha^2}{2} \left\langle g_i''(x_{ki}) p_k, p_k \right\rangle \right], \tag{33}$$

где x_{ki} — промежуточные точки разложения в ряд Тейлора. Обозначим $g_k(\alpha) = \max_i \left[g_i(x_k) + \alpha \left\langle g_i'(x_k), p_k \right\rangle \right]$, $i = 0,\dots,l$. Из выпуклости $g_k(\alpha)$ следует

$$g_k(\alpha) \le g_k(0) + \alpha\,(g_k(1) - g_k(0)) \le g^+(x_k) - \alpha g^+(x_k),$$

где учитывалось, что $g_k(1) \le 0$ вследствие выполнения ограничений в (7). Теперь из (33) имеем при любом $\alpha \in (0;1]$

$$g^+(x_k + \alpha p_k) \le g^+(x_k) - \alpha g^+(x_k) + \alpha^2 T_g \|p_k\|^2 / 2, \tag{34}$$

где T_g — мажоранта для $\|g_i''(x)\|$ на отрезке $[x_k; \bar{x}_k]$, $i = 1,\dots,l$. Поэтому неравенство (30) справедливо, по крайней мере, для достаточно малых $\alpha > 0$.

Используя доказательство леммы 1, получаем

$$\varphi(x_k + \alpha p_k, \lambda_k) \le$$

$$\le \varphi(z_k) - \alpha \left[\left\langle L_{xx}(z_k) p_k, p_k \right\rangle + \left\langle \lambda_k, g(x_k) \right\rangle \right] + \alpha^2 T_\varphi \|p_k\|^2 / 2, \tag{35}$$

где учитывалась отрицательная определенность матриц $-g_i''(x)$; T_φ — мажоранта для $\|\varphi_{xx}(x, \lambda_k)\|$ на $[x_k, \bar{x}_k]$. В качестве T_φ может выступать удвоенная константа Липшица для $\varphi_x(x, \lambda_k)$, если в задаче (1), (2) непрерывность $F_{xx}(x)$, $g_i'''(x)$ ослабить требованием существования константы Липшица для $F_x(x)$, $g_i''(x)$.

На основании (34), (35) и (22)

$$\Phi_N(x_k + \alpha p_k, \lambda_k) - \Phi_N(z_k) \le -\alpha \psi_k + \frac{\alpha^2}{2} T \|p_k\|^2 \le$$

$$\le -\alpha \gamma^{-1} \Phi_N(z_k) + \frac{\alpha^2}{2} T \|p_k\|^2, \tag{36}$$

где $T = T_\varphi + N\,T_g$. Так как $\Phi_N(z_k) > 0$ в точках $x_k \neq x_*$, то отсюда вытекает (31), по крайней мере, для достаточно малых $\alpha > 0$. Итак, процедура определения α_k в алгоритме является обоснованной.

Из $g^+(x_k) \le C_g$ вытекает ограниченность $\{x_k\}$, поскольку множество $\{x \mid g^+(x) \le C_g\}$ ограничено. Поэтому λ_k и p_k ограничены (лемма 4)), величина $N_k = N$ постоянна при $k \ge k_0$ и удовлетворяет (24); T, T_g, T_φ — константы, не зависящие от k.

Получим оценку снизу для α_k. На основании (34)

$$g^+(x_k + \alpha p_k) \le (1-\alpha)\,C_g + \alpha^2\,T_g\,\|p_k\|^2 / 2 \le C_g\,,$$

если $\alpha \le 2C_g\,(T\,\|p_k\|^2)^{-1}$. Это означает, что (30) выполняется при указанных α.

Установим, при каких α выполняется (31). Из (36) вытекает

$$\Phi_N(x_k + \alpha p_k, \lambda_k) - \Phi_N(z_k) \le$$

$$\le -\alpha^2\,\Phi_N(z_k)\left[(\alpha\gamma)^{-1} - \frac{T\,\|p_k\|^2}{2\Phi_N(z_k)}\right] \le -\varepsilon\alpha^2\,\Phi_N(z_k)$$

при условии, что величина в квадратных скобках не меньше, чем ε. Последнее имеет место для $\alpha \le 2\Phi_N(z_k)\left[\gamma(2\varepsilon\Phi_N(z_k) + T\,\|p_k\|^2)\right]^{-1}$. Таким образом, процедура определения α_k гарантирует получение значений

$$\alpha_k \ge \min\left\{1\,;\,\frac{C_g}{T\,\|p_k\|^2}\,;\,\frac{\Phi_N(z_k)}{\gamma\,(\,2\varepsilon\Phi_N(z_k) + T\|p_k\|^2\,)}\right\}.$$

Здесь $C_g\left[T\|p_k\|^2\right]^{-1} \ge \overline{\alpha}_1 > 0$ вследствие ограниченности p_k. Далее, из $\Phi_N(z_k) \ge \langle H\,p_k, p_k\rangle / 2$ и из (5) следует

$$\|p_k\|^2 \le 2\Phi_N(z_k)\,a^{-1}\,. \tag{37}$$

Поэтому

$$\frac{\Phi_N(z_k)}{\gamma(\,2\varepsilon\Phi_N(z_k) + T\|p_k\|^2\,)} \ge \frac{a}{2\gamma(\varepsilon a + T\,)} = \overline{\alpha}_2 > 0\,. \tag{38}$$

В итоге $\alpha_k \ge \overline{\alpha}$, где $\overline{\alpha} > 0$ — константа.. Теперь из неравенства (31), рассматриваемого при $\alpha = \alpha_k$, вытекает $\Phi_N(z_k) \to 0$, откуда $x_k \to x_*$, $\psi_k \to 0$ (предложение 2) и $p_k \to 0$, $\lambda_k \to \Lambda_*$, что очевидно.

Установим оценки скорости сходимости (32). На основании (16) $\Phi_N(x_{k+1},\lambda_{k+1}) \le \Phi_N(x_{k+1,}\lambda_k)$. Вместе с (31) это дает $\Phi_N(z_{k+1}) \le$ $\le q\,\Phi_N(z_k)$, $q = 1-\varepsilon\,(\overline{\alpha})^2 < 1$. При этом, как легко заметить, $\overline{\alpha} = \overline{\alpha}_2$, когда $p_k \to 0$. Если k_0 — наименьший номер итерации с установившимся $N_k = N$, то при $k \ge k_0$ приходим к первой из оценок (32):

$$\Phi_N(z_k) \le \Phi_N(z_{k_0})\,q^{\,k-k_0} = C_0\,q^k\,,\quad C_0 = \Phi_N(z_{k_0})\,q^{\,-k_0}\,.$$

Вследствие (37) отсюда вытекает вторая оценка (32) со значениями $C_1 = (2C_0\,/\,a)^{1/2}$, $q_1 = q^{1/2} < 1$.

Наконец, при $\rho > k > k_0$ с учетом предыдущей оценки

$$\left\| x_\rho - x_k \right\| \le \sum_{i=k}^{\rho-1} \| p_i \| \le C_1 \sum_{i=k}^{\rho-1} q_1^i \le C_1 \sum_{i=k}^{\infty} q_1^i = C_1 \frac{q_1^k}{1-q_1}\,.$$

Переходя здесь к пределу $x_\rho \to x_*$ при $\rho \to \infty$, получаем последнюю оценку (32) со значением $C_2 = C_1(1-q_1)^{-1}$.

Теорема доказана.

Сравнивая алгоритм с методом Пшеничного [4], отметим, что в последнем вместо (31), (24) фигурируют соответственно

$$\Phi_N(x_k + \alpha p_k\,,\lambda_k) \le \Phi_N(z_k) - \overline{\varepsilon}\alpha^2 \| p_k \|^2\,,$$

$$N = \max\left\{ N_{k-1}\,,\; 2\sum_{i=1}^{l} \lambda_k^i \right\}, \tag{39}$$

где $\overline{\varepsilon} > 0$ — произвольная константа и $H_k = I_n$ в (7). Понятно, что неравенство (24) выполняется при меньших, чем в (39) значениях N. Это влечет уменьшение T и увеличение $\overline{\alpha}_2$ в (38), чем обеспечивается снижение величины q в оценках (32). Если некоторые ограничения $g_j(x) \le 0$ линейны, причем $g_j(x_0) \le 0$, то коэффициент штрафа N «не накапливает», в отличие от [4], значений λ_k^j по процессу, поскольку $g_j(x_k) \le 0$ при всех $k = 0, 1, \ldots$ (если $j = 1,\ldots,l$, то $N = 0$), что также ведет к убыстрению сходимости процесса.

Покажем теперь, используя теорему 1, что алгоритм Пшеничного [4] обладает линейной скоростью сходимости. Заменим первое неравенство (39) более общим

$$\Phi_N(x_k + \alpha p_k, \lambda_k) \le \Phi_N(z_k) - \overline{\varepsilon}\alpha^2 \langle H p_k, p_k \rangle / 2\,,\quad H = H^T\,, \tag{40}$$

совпадающим с (39) при $H = I_n$. Этот алгоритм отличается от описанного выше (шаг 1 – шаг 3) лишь выбором α_k из условия проверки неравенства

(40) вместо (31). Для краткости назовем его проекционным методом (40), а рассмотренный выше — соответственно проекционным методом (31).

Теорема 2. Если $\bar{\varepsilon} < 1$, то для проекционного метода (40) справедливы все утверждения теоремы 1.

Доказательство. Обозначим β_k шаговые множители проекционного метода (40) $x_{k+1} = x_k + \beta_k p_k$ и примем $\varepsilon = \bar{\varepsilon}$ в (31). Поскольку $\Phi_N(z_k) \geq \geq \langle H p_k, p_k \rangle / 2$, то из (31) вытекает (40), так что процедура определения β_k в методе (40) является обоснованной. Из сравнения (31) и (40) следует $1 \geq \beta_k \geq \alpha_k \geq \bar{\alpha}$. На основании (40) $p_k \to 0,\ \ x_k \to x_*$ (предложение 1).

Установим оценки скорости сходимости. В точке x_k, достаточно близкой к x_*, возможны две ситуации:

1. $\langle H p_k, p_k \rangle = o(\psi_k^{(2)})$.

Оценим сначала величину α_k. Вследствие (23) имеем $\| p_k \|^2 = o(\psi_k^{(2)})$, $\| L_x(z_k) \|^2 = \| H p_k \|^2 = o(\psi_k^{(2)})$ и из выражений для $\Phi_N(z_k), \psi_k$ следует $\Phi_N(z_k) = \psi_k^{(2)} + o(\psi_k^{(2)})$, $\psi_k = \psi_k^{(2)} + o(\psi_k^{(2)})$. Поэтому $\psi_k = \Phi_N(z_k) + + o(\Phi_N(z_k))$, $\| p_k \|^2 = o(\Phi_N(z_k))$ и для любого $\alpha \leq 1$

$$-\alpha\psi_k + \frac{\alpha^2}{2} T \| p_k \|^2 = -\alpha\Phi_N(z_k) + o(\Phi_N(z_k)) <$$

$$< -\bar{\varepsilon}\alpha\Phi_N(z_k) \leq -\bar{\varepsilon}\alpha^2\Phi_N(z_k).$$

Используя это соотношение в первом неравенстве (36), приходим к (31) при любом $\alpha \in (0;1]$, что гарантирует в алгоритме получение $\alpha_k = 1$ вблизи точки x_*.

Поскольку $1 \geq \beta_k \geq \alpha_k$, то $\beta_k = \alpha_k = 1$. Тогда из доказательства теоремы 1 вытекает $\Phi_N(z_{k+1}) \leq q_1 \Phi_N(z_k)$, $q_1 = 1 - \bar{\varepsilon} < 1$.

2. $\langle H p_k, p_k \rangle \geq C\psi_k^{(2)},\ C > 0$.

С учетом выражения для $\Phi_N(z_k)$ имеем

$$\langle H p_k, p_k \rangle \geq \frac{1}{2} \langle H p_k, p_k \rangle + \frac{C}{2} \psi_k^{(2)} \geq \bar{C} \Phi_N(z_k),$$

где $\bar{C} = \min\{1; C/2\}$. Тогда из (40) следует

$$\Phi_N(x_k + \beta_k p_k, \lambda_k) \leq \Phi_N(z_k) - \bar{\varepsilon}\beta_k^2 \bar{C} \Phi_N(z_k)/2 \leq q_2 \Phi_N(z_k),$$

где $q_2 = 1 - \bar{\varepsilon}(\bar{\alpha})^2 \bar{C}/2 < 1$.

В итоге в любой точке x_k алгоритма (40) при $k \geq k_0$ выполняется $\Phi_N(z_{k+1}) \leq q\Phi_N(z_k)$, $q = \max\{q_1; q_2\} < 1$, откуда следуют все оценки (32).

Теорема доказана.

ЗАМЕЧАНИЯ

1. Опираясь на лемму 1, можно вместо (31) или (40) использовать следующее неравенство, обычно применяемое в условной оптимизации:

$$\Phi_N(x_k + \alpha p_k, \lambda_k) \le \Phi_N(z_k) - \varepsilon\alpha\psi_k, \quad \varepsilon \in (0;1) \qquad (41)$$

и по приведенной методике доказать все утверждения теоремы 1. Однако такой алгоритм безперспективен, поскольку, в отличие от приведенных выше, требует для определения α_k вычисления ψ_k, т.е. вычисления $L_{xx}(z_k)$, и является весьма трудоемким алгоритмом второго порядка. В то же время замена ψ_k в (41) на $\langle H p_k, p_k \rangle$ или на $\Phi_N(z_k)$, а также одновременное (вместе с α) дробление ε означают переход от (41) к менее трудоемкому неравенству (40) или, соответственно, к (31). Для задачи отыскания седловой точки выпукло-вогнутой функции, являющейся частным случаем в.н., неравенство (39) при отыскании α_k было предложено в [7].

2. Условие (2), а точнее — (23), для проекционных методов существенно не только при установлении оценок (32), но и для сходимости к решению. Они неприменимы, вообще говоря, для решения вырожденных задач ($m = 0$ в (2)). В частном варианте вариационных неравенств — в задачах оптимизации — проекционные методы сохраняют сходимость за счет использования самой специфики оптимизационной задачи, а именно, за счет потенциональности оператора $F(x)$, т.е существования функции $f_0(x)$, определяющей оператор $F(x) = f_0'(x)$. Так, в методе линеаризации Пшеничного [8] сходимость к решению (вырожденных) задач оптимизации устанавливается за счет определения α_k из условия релаксации штрафной функции $\widetilde{\Phi}_N(x) = f_0(x) + N g^+(x)$, использующей $f_0(x)$. Модификация правила выбора α_k метода линеаризации [8] (аналогичная использованию (41)) позволила установить [9] оценки скорости сходимости процесса: линейную — при условии (2) и оценку $o(k^{-1})$ по функционалу $\widetilde{\Phi}_N(x)$ — в случае вырождения выпуклой задачи оптимизации. Для численного решения монотонных (с вырождением) в.н. разрабатываются специальные подходы (методы итеративной регуляризации, экстраградиентный метод и другие) [2, 10].

ЛИТЕРАТУРА

1. *Harker P.T., Pang J.S.* Finite-dimensional variational inequality and nonlinear complementarity problems:a survey of theory, algorithms and application // Math. Progr. — 1990. — **48**, N 2. — P. 161–220.
2. *Панин В.М., Скопецкий В.В., Лаврина Т.В.* Модели и методы конечномерных вариационных неравенств // Кибернетика и системный анализ.— 2000. — № 6. — С. 47–61.

3. *Nagurney A.* Network economics : a variational inequality approach. — Norvell : Kluver akad. publ., 1993. — 326 p.
4. *Пшеничный Б.Н, Калжанов М.У.* Метод решения вариационных неравенств // Кибернетика и системный анализ. — 1992. — № 6. — С. 48–55.
5. *Панин В.М., Александрова В.М.* Линейная сходимость метода решения вариационных неравенств // Кибернетика и системный анализ. — 1994. — № 3. — С. 176–180.
6. *Jia Hao Wu, Florian M., Marcotte P.* A general descent framework for the monotone variational inequality problem // Math. Progr. — 1993. — **61**, N 3. — P. 281–300.
7. *Данилин Ю.М., Панин В.М.* О некоторых методах поиска седловых точек // Кибернетика. — 1974. — № 3. — С. 119–124.
8. *Пшеничный Б.Н.* Метод линеаризации. — М.: Наука, 1983. — 136 с.
9. *Панин В.М.* Оценки сходимости двух методов линеаризации // Докл. АН УССР. Сер. А. — 1984. — № 5. — С. 77–79.
10. *Гольштейн Е.Г., Третьяков В.М.* Модифицированные функции Лагранжа. — М.: Наука, 1989. — 400 с.

Поступила 20.03.2002