4.11.* Экстраградиентный метод

Покажем, что, несколько усложнив схему градиентного метода, можно получить сходящийся метод решения задач выпуклого программирования, используя классическую нормальную функцию Лагранжа. Этот метод будем называть *экстраполяционным градиентным* или, короче, *экстраградиентным методом*. В нем наряду с градиентными шагами будут присутствовать так называемые прогнозные шаги, что оправдывает присутствие в названии этого метода слова «экстраполяционный».

4.11.1. Описание процесса

Рассмотрим задачу условной минимизации

$$f(x) \to \min, \quad x \in X = \{x \in X_0 \colon g_1(x) \le 0, \ldots, g_m(x) \le 0\}, \tag{4.155}$$

где X_0 – выпуклое замкнутое множество из E^n, а $g(x) = (g_1(x), \ldots, g_m(x))^\top$ – функция из $C^1(X_0)$, $i = 1, \ldots, m$. Нормальная функция Лагранжа этой задачи

$$L(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle, \quad x \in X_0, \ \lambda \in \Lambda_0 \overset{\text{def}}{=} E^m_+. \tag{4.156}$$

Она имеет производные

$$\frac{\partial L(x,\lambda)}{\partial x} = L'_x(x,\lambda) = f'(x) + \lambda^\top g'(x); \quad \frac{\partial L(x,\lambda)}{\partial \lambda} = L'_\lambda(x,\lambda) = g(x), \tag{4.157}$$

где $g'(x) = \begin{pmatrix} \dfrac{\partial g_1(x)}{\partial x^1} & \ldots & \dfrac{\partial g_1(x)}{\partial x^n} \\ \ldots & \ldots & \ldots \\ \dfrac{\partial g_m(x)}{\partial x^1} & \ldots & \dfrac{\partial g_m(x)}{\partial x^n} \end{pmatrix}$, $x \in X_0$, – матрица Якоби.

Опишем один шаг экстраградиентного метода. Пусть начальное приближение $(x_0, \lambda_0) \in X_0 \times \Lambda_0$ задано. Для перехода от k-го приближения (x_k, λ_k), $k \ge 0$, к $(k+1)$-му сначала делаем прогнозный шаг

$$\begin{aligned} \overline{x}_k &= P_{X_0}(x_k - \beta L_x(x_k, \lambda_k)) = P_{X_0}(x_k - \beta(f'(x_k) + \lambda_k^\top g'(x_k)); \\ \overline{\lambda}_k &= P_{\Lambda_0}(\lambda_k + \beta L_\lambda(x_k, \lambda_k)) = P_{\Lambda_0}(\lambda_k + \beta g(x_k)), \end{aligned} \tag{4.158}$$

затем основной шаг

$$\begin{aligned} x_{k+1} &= P_{X_0}(x_k - \beta L_x(\overline{x}_k, \overline{\lambda}_k)) = P_{X_0}(x_k - \beta(f'(\overline{x}_k) + \overline{\lambda}_k^\top g'(\overline{x}_k)); \\ \lambda_{k+1} &= P_{\Lambda_0}(\lambda_k + \beta L_\lambda(\overline{x}_k, \overline{\lambda}_k)) = P_{\Lambda_0}(\lambda_k + \beta g(\overline{x}_k)), \end{aligned} \tag{4.159}$$

где $\beta > 0$ – параметр метода. Запишем метод (4.158), (4.159) в кратком виде. Для этого введем обозначения

$$z = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad Az \overset{\text{def}}{=} \begin{pmatrix} L_x(x,\lambda) \\ -L_\lambda(x,\lambda) \end{pmatrix} = \begin{pmatrix} f'(x) + \lambda^\top g'(x) \\ -g(x) \end{pmatrix}, \tag{4.160}$$

$$Z = X_0 \times \Lambda_0 \subset E^{n+m}.$$

Заметим, что оператор проектирования точки $z = \begin{pmatrix} x \\ \lambda \end{pmatrix}$ на множество Z вычисляется по формуле (см. упражнение 3.18)

$$P_Z z = \begin{pmatrix} P_{X_0} x \\ P_{\Lambda_0} \lambda \end{pmatrix}.$$

Поэтому метод (4.158), (4.159) можем кратко записать так:

$$z_0 = (x_0, \lambda_0); \quad \overline{z}_k = P_Z(z_k - \beta A z_k), \; z_{k+1} = P_Z(z_k - \beta A \overline{z}_k), \quad k = 0, 1, \dots . \tag{4.161}$$

4.11.2. Сходимость экстраградиентного метода

Сходимость экстраградиентного метода (4.161) будем доказывать, предполагая, что выполнены следующие условия I—V:

I. Множество $X_0 \subseteq E^n$ выпукло, замкнуто.

II. Функции $f(x)$, $g_1(x)$, ..., $g_m(x)$ выпуклы и непрерывно дифференцируемы на X_0.

III. Для любого ограниченного подмножества X_1 множества X_0 существует постоянная $M = M(X_1) > 0$ такая, что

$$\|g(x) - g(y)\| \le M\|x - y\|; \; \|f'(x) - f'(y)\| \le M\|x - y\|;$$
$$\|g'(x) - g'(y)\| \le M\|x - y\| \quad \forall x, y \in X_1. \tag{4.162}$$

IV. Нормальная функция Лагранжа $L(x, \lambda)$ (4.156) задачи (4.155) имеет седловую точку $s_* = \begin{pmatrix} x_* \\ \lambda_* \end{pmatrix} \in Z$, т.е.

$$L(x_*, \lambda) \le L(x_*, \lambda_*) \le L(x, \lambda_*) \quad \forall x \in X_0, \; \lambda \in \Lambda_0 = E^m_+. \tag{4.163}$$

Множество всех седловых точек функции $L(x, \lambda)$ будем обозначать через S_*.

Для формулировки условия V нам понадобится множество

$$S_0 = \left\{ z = \begin{pmatrix} x \\ \lambda \end{pmatrix} \in Z : \|z - z_0\| \le 4\left(\|z_* - z_0\| + \|Az_0\|\right) \overset{\text{def}}{=} R_0 \right\}, \tag{4.164}$$

где $z_0 = \begin{pmatrix} x_0 \\ \lambda_0 \end{pmatrix} \in Z$ — начальная точка процесса (4.161), z_* — какая-либо седловая точка функции $L(x, \lambda)$.

Множество S_0 (4.164) выпукло, замкнуто, ограничено, и на нем отображение A из (4.160) удовлетворяет условию Липшица с некоторой константой L_0:

$$\|Az - Ay\| \le L_0\|z - y\| \quad \forall z, y \in S_0 \tag{4.165}$$

— это вытекает из условия (4.162) при $X_1 = \{x \in X_0 : \|x - x_0\| \le R_0\}$. Подчеркнем, что множество $Z = X_0 \times E^m_+$ не ограничено и, как очевидно из формул (4.160), отображение A не будет удовлетворять условию Липшица на Z.

Теперь можем сформулировать условие V.

V. Будем предполагать, что параметр β в методе (4.161) выбран из условия

$$0<\beta<\min\left\{\frac{1}{L_0};1\right\}. \tag{4.166}$$

Теорема 4.24. *Пусть выполнены условия I—V. Тогда метод (4.161), (4.166) при любом начальном приближении $z_0 \in Z$ порождает последовательности $\{z_k\}$, $\{\overline{z}_k\}$, которые принадлежат множеству S_0 из (4.164) и монотонно сходятся к некоторой седловой точке $\tilde{s}_* = \begin{pmatrix}\tilde{x}_*\\ \tilde{\lambda}^*\end{pmatrix} \in S_0$ функции Лагранжа (4.156) с первой компонентой $\tilde{x}_* \in X$, являющейся решением задачи (4.155).*

Для доказательства этой теоремы нам понадобятся следующие две леммы.

Лемма 4.2. Пусть выполнены условия I, II. Тогда точка $s_* = \begin{pmatrix}x_*\\ \lambda_*\end{pmatrix}$ из Z будет седловой точкой функции (4.156) тогда и только тогда, когда выполняется неравенство

$$\langle As_*, z-s_*\rangle \ge 0 \quad \forall z\in Z. \tag{4.167}$$

Доказательство

Функция $L(x, \lambda)$ выпукла по переменной x на X_0 при каждом $\lambda \in E_+^m$, вогнута (линейна) по переменной $\lambda \in E_+^m$ при каждом $x \in X_0$. Согласно теореме 3.4 тогда правое и левое неравенства (4.163) соответственно выполняются тогда и только тогда, когда

$$\langle L_x'(x_*, \lambda^*), x-x_*\rangle \ge 0 \quad \forall x\in X_0; \qquad \langle -L_\lambda'(x_*, \lambda^*), \lambda-\lambda^*\rangle \ge 0 \quad \forall \lambda \in E_+^m.$$

Сложив эти неравенства, придем к формуле (4.167). ■

Лемма 4.3. Пусть выполнены условия I, II. Тогда отображение A из (4.160) является монотонным, т.е.

$$\langle Az-Ay, z-y\rangle \ge 0 \quad \forall z, y\in Z. \tag{4.168}$$

Доказательство

Так как функция $L(x, \lambda)$ дифференцируема по x, λ, выпукла по x, вогнута по λ, то в силу теоремы 3.1 имеем

$$\langle L_x'(x, \lambda), x-y\rangle \ge L(x,\lambda)-L(y,\lambda); \quad \langle L_x'(y,p), y-x\rangle \ge L(y,p)-L(x,p);$$
$$\langle -L_\lambda'(x,\lambda), \lambda-p\rangle \ge L(x,p)-L(x,\lambda); \quad \langle -L_\lambda'(y,p), p-\lambda\rangle \ge L(y,\lambda)-L(y,p),$$

при всех $x, y \in X_0$, $\lambda, p \in E_+^m$. Сложив эти неравенства, получим

$$\langle L_x'(x,\lambda)-L_x'(y,p), x-y\rangle + \langle -L_\lambda'(x,\lambda)-(-L_\lambda'(y,p)), \lambda-p\rangle \ge 0,$$

что равносильно (4.168). ■

Замечание 4.7. Выше мы уже встречались с монотонными отображениями — это градиент выпуклой функции (теорема 3.1). Однако здесь отображение A из (4.160) в отличие от упомянутых случаев не является градиентом какой-либо функции.

Перейдем к доказательству теоремы.