ROZDZIAŁ 4

Metody minimalizacji różniczkowalnej z ograniczeniami

4.1 Programowanie kwadratowe

Definicja 4.1.1 Pod pojęciem *zadania programowania kwadratowego* rozumie się zadanie maksymalizacji lub minimalizacji funkcji kwadratowej przy ograniczeniach liniowych równościowych lub nierównościowych.

4.1.1 Programowanie kwadratowe przy ograniczeniach równościowych

W ustępie tym będziemy się zajmować zadaniem programowania kwadratowego postaci

$$\begin{array}{ll} \text{minimalizować} & q(x) = \frac{1}{2}x^T G x + g^T x \\ \text{względem} & x \in \mathbb{R}^n \\ \text{przy ograniczeniach} & Ax = b, \end{array} \tag{4.1}$$

gdzie G jest macierzą symetryczną typu $n \times n$, A jest macierzą typu $m \times n$, $g \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. Zakładamy, że układ $Ax = b$ posiada rozwiązanie, czyli $b \in \operatorname{im} A$. Bez szkody dla ogólności rozważań zakładamy również, że macierz A jest rzędu m. W razie potrzeby możemy bowiem usunąć liniowo zależne wiersze. Twierdzenie 2.3.42 określa warunki wystarczające istnienia w punkcie $x^* \in \mathbb{R}^n$ rozwiązania dla zadania minimalizacji z ograniczeniami równościowymi, natomiast celem tego ustępu będzie podanie metod wyznaczenia takiego rozwiązania.

Metoda bezpośredniej eliminacji zmiennych

Metoda ta polega na wyznaczeniu m zmiennych z ograniczeń i wstawieniu ich do funkcji celu redukując w ten sposób wyjściowe zadanie do zadania minimalizacji bez ograniczeń funkcji $n-m$ zmiennych.

Ćwiczenie 4.1.2 Wyznaczyć rozwiązanie następującego zadania programowania kwadratowego

$$\begin{array}{ll} \text{minimalizować} & q(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 4x_1x_2 + x_3^2 \\ \text{względem} & (x_1, x_2, x_3) \in \mathbb{R}^3 \\ \text{przy ograniczeniach} & x_1 + x_2 + 2x_3 = 2 \\ & x_1 - x_2 = 2 \end{array}$$

(a) wyznaczając punkt KT i sprawdzając warunki rzędu drugiego dla zadania z ograniczeniami;

(b) eliminując z układu równań zmienne x_2 i x_3 i sprawdzając warunki wystarczające dla zadania bez ograniczeń.

Wróćmy teraz do przypadku ogólnego, czyli do zadania (4.1). Wydzielmy z macierzy A podmacierz nieosobliwą A_1 typu $m \times m$ (istnienie takiej macierzy jest zagwarantowane założeniem $r(A) = m$). Dla uproszczenia zapisu zapiszmy (po ewentualnym przenumerowaniu zmiennych) macierze A i G w postaci blokowej:

$$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}, \qquad G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix},$$

w których wydzielono bloki A_1 i G_{11} typu $m \times m$, gdzie A jest macierzą nieosobliwą, i podobnie wektory $x, g \in \mathbb{R}^n$ w postaci blokowej

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \in \mathbb{R}^m \times \mathbb{R}^{n-m}$$

z wydzielonymi wektorami $x_1, g_1 \in \mathbb{R}^m$ (oczywiście $G_{21} = G_{12}^T$, bo G jest macierzą symetryczną). Funkcję celu możemy więc zapisać w postaci

$$q(x_1, x_2) = \frac{1}{2} x_1^T G_{11} x_1 + x_1^T G_{12} x_2 + \frac{1}{2} x_2^T G_{22} x_2 + g_1^T x_1 + g_2^T x_2,$$

zaś ograniczenia – w postaci

$$A_1 x_1 + A_2 x_2 = b.$$

Wyznaczając x_1 z tych ograniczeń otrzymamy

$$x_1 = A_1^{-1}(b - A_2 x_2). \tag{4.2}$$

Wstawiając tak wyznacznone x_1 do funkcji celu $q(x_1, x_2)$ otrzymamy po prostych rachunkach, że wyjściowe zadanie (4.1) zredukuje się do następującego zadania bez ograniczeń:

$$\begin{array}{ll} \text{minimalizować} & r(x_2) = \frac{1}{2} x_2^T H x_2 + h^T x_2 + c \\ \text{względem} & x_2 \in \mathbb{R}^{n-m}, \end{array}$$

gdzie

$$H = G_{22} - 2(A_1^{-1}A_2)^T G_{12} + (A_1^{-1}A_2)^T G_{11} A_1^{-1} A_2,$$
$$h = G_{12}^T A_1^{-1} b^T - (A_1^{-1}A_2)^T G_{11} A_1^{-1} b - (A_1^{-1}A_2)^T g_1 + g_2$$

i

$$c = \frac{1}{2} b^T A_1^{-T} G_{11} A_1^{-1} b + g_1^T A_1^{-1} b.$$

Jeśli macierz H jest dodatnio określona, to zgodnie z twierdzeniem 2.2.2 zadanie powyższe posiada dokładnie jedno rozwiązanie $x_2^* = H^{-1}h$ (można je otrzymać rozwiązując układ równań liniowych $\nabla r(x_2) = 0$). Wstawiając wyznaczone x_2^* do równości (4.2) dostajemy x_1^* i rozwiązanie $x^* = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}$ wyjściowego zadania. Wektor mnożników Lagrange'a $y^* \in \mathbb{R}^m$ dla zadania (4.1) otrzymujemy rozwiązując układ równań $\nabla_x L(x^*, \lambda) = 0$, czyli

$$Gx^* + g + A^T y = 0.$$

Z twierdzenia Kroneckera-Capellego wynika, że istnieje dokładnie jedno rozwiązanie y^* tego układu n równań z m niewiadomymi, bo $r(A) = m$. Nietrudno zauważyć, że rozwiązanie to wyraża się wzorem

$$y^* = -A_1^{-1}(G_{11}x_1^* + G_{12}x_2^* + g_1).$$

Uogólniona metoda eliminacji zmiennych

Metoda ta polega w istocie na tym, aby w eliminacji części zmiennych z ograniczeń równościowych $Ax = b$ zadania (4.1) wykorzystać rozkład spełniającego te ograniczenia wektora x na składową leżącą w $\ker A$ i składową leżącą w $\operatorname{Im} A^T$. Rozkład taki jest możliwy, ponieważ

$$\mathbb{R}^n = \ker A \oplus \operatorname{im} A^T.$$

Niech S i Z będą macierzami odpowiednio typu $n \times m$ i $n \times (n-m)$ takimi, że $[S, Z]$ jest macierzą nieosobliwą i

$$A[S, Z] = [I, 0], \tag{4.3}$$

gdzie I jest macierzą jednostkową typu $m \times m$, zaś 0 jest macierzą zerową typu $n \times (n-m)$. Można przyjąć $[S, Z] = \begin{bmatrix} A \\ B \end{bmatrix}^{-1}$, gdzie B jest macierzą o $n - m$ wierszach dobraną tak, aby macierz $\begin{bmatrix} A \\ B \end{bmatrix}$ była nieosobliwa. Mamy bowiem

$$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} [S, Z] = \begin{bmatrix} AS & AZ \\ BS & BZ \end{bmatrix}, \tag{4.4}$$

czyli w szczególności spełniona jest równość (4.3). Możemy również przyjąć $S = A^T(AA^T)^{-1}$ i $Z = I - A^T(AA^T)^{-1}A$. Zauważmy bowiem, że dla takich macierzy S i Z zachodzi równość (4.3). Jeśli S i Z spełniają (4.3), to nietrudno zauważyć, że $x = Sb + u$, gdzie $u \in \ker A$, jest rozwiązaniem układu $Ax = b$. Zachodzi również implikacja odwrotna, czyli dowolne rozwiązanie układu $Ax = b$ jest dane równością $x = Sb + u$, gdzie $u \in \ker A$. Wynika to z faktu, że dowolne rozwiązanie układu niejednorodnego $Ax = b$ jest sumą rozwiązania szczególnego równania niejednorodnego (tutaj Sb) i rozwiązania ogólnego równania jednorodnego (tutaj $u \in \ker A$). Zauważmy, że kolumny macierzy Z rozpinają bazę w przestrzeni $\ker A$ (wszystkie należą do $\ker A$, jest ich $n - m$ i są liniowo niezależne), wobec tego $u = Zv$ dla pewnego $v \in \mathbb{R}^{n-m}$. Zatem każdy punkt x spełniający równanie $Ax = b$ można przedstawić w postaci $x = Sb + Zv$. Wstawiając tak wyrażony x do funkcji celu q otrzymujemy *zredukowaną funkcję celu*

$$r(v) = \frac{1}{2} v^T Z^T G Z v + (g + GSb)^T Z v + \frac{1}{2}(g + GSb)^T Sb + c, \tag{4.5}$$

gdzie $c = \frac{1}{2}(Sb)^T GSb + g^T Sb$. Zmienne $v_i, i = 1, ..., n - m$, nazywają się *zmiennymi zredukowanymi*. Wyjściowe zadanie programowania kwadratowego zostało więc zredukowane do zadania minimalizacji bez ograniczeń z kwadratową funkcją celu r. Macierz $Z^T GZ$ nazywa się *zredukowanym hesjanem* zaś wektor $Z^T(g+GSb)$ – *zredukowanym gradientem* (por. ustęp 2.3.3). Przypuśćmy, że zredukowany hesjan jest dodatnio określony. Wówczas zredukowane zadanie

$$\begin{array}{ll} \text{minimalizować} & r(v) \\ \text{względem} & v \in \mathbb{R}^{n-m} \end{array}$$

posiada dokładnie jedno rozwiązanie v^* (por. twierdzenie 2.3.44), które można wyznaczyć z równania $\nabla r(v) = 0$, czyli

$$Z^T GZv + Z^T(g + GSb) = 0.$$

Ponieważ macierz $Z^T GZ$ jest nieosobliwa (jako dodatnio określona), więc

$$v^* = -(Z^T GZ)^{-1} Z^T (g + GSb), \tag{4.6}$$

natomiast rozwiązanie x^* wyjściowego zadania otrzymuje się z równości

$$x^* = Sb + Zv^*. \tag{4.7}$$

Wektor mnożników Lagrange'a y^* wyznacza się z równania $\nabla_x L(x^*, y) = 0$, czyli

$$Gx^* + g + A^T y = 0.$$

Po lewostronnym pomnożeniu ostatniego równania przez S^T otrzymamy

$$y^* = -S^T(Gx^* + g). \tag{4.8}$$

Macierze S i Z można wybierać na różne sposoby. Niektóre z nich były podane wyżej. Jeden z innych sposobów jest wynikiem faktoryzacji ortogonalnej macierzy A^T. Jeśli bowiem macierz A^T przedstawimy w postaci

$$A^T = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R,$$

gdzie $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ jest macierzą ortogonalną typu $n \times n$, R – górną macierzą trójkątną typu $m \times m$, a pozostałe macierze są odpowiednich typów (muszą być zachowane reguły blokowego mnożenia macierzy), to przyjmując $S = Q_1 R^{-T}$ i $Z = Q_2$ zajdzie żądana równość $A[S, Z] = [I, 0]$.

Metoda Lagrange'a

Przedstawimy teraz inną metodę rozwiązania zadania (4.1) używając warunków rzędu drugiego dla zadania minimalizacji z ograniczeniami równościowymi. Funkcja Lagrange'a dla zadania (4.1) ma postać

$$L(x, y) = \frac{1}{2} x^T G x + g^T x + y^T (Ax - b).$$

Jest jasne, że spełniony jest podstawowy warunek regularności, gdyż spełniony jest warunek Karlina albo warunek Fiacco–McCormicka. Punkt Kuhna–Tuckera (x^*, y^*) jest rozwiązaniem układu równań

$$\begin{aligned} \nabla_x L(x, y) &= Gx + g + A^T y = 0, \\ \nabla_y L(x, y) &= Ax - b = 0, \end{aligned}$$

który możemy zapisać w postaci blokowej

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -g \\ b \end{bmatrix}. \tag{4.9}$$

Macierz

$$H := \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \tag{4.10}$$

nazywa się *macierzą Lagrange'a* (ang. *Lagrangian matrix*) albo hesjanem obrzeżonym (por. ustęp 2.3.3). Jeśli H jest macierzą nieosobliwą, to jedynym punktem Kuhna–Tuckera jest

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = H^{-1} \begin{bmatrix} -g \\ b \end{bmatrix}. \tag{4.11}$$

Metoda wyznaczenia rozwiązania optymalnego zadania (4.1) poprzez rozwiązanie równania (4.9) nazywa się *metodą Lagrange'a*. Podamy dwa sposoby rozwiązania tego równania. Pierwszy z nich polega na odwróceniu macierzy Lagrange'a. Pamiętajmy jednak przy tym, że w ten sposób otrzymamy jedynie punkt Kuhna–Tuckera (x^*, y^*), który jest jedynie kandydatem na rozwiązanie optymalne zadania (4.1). Przy dodatkowym założeniu nieujemnej określoności macierzy G zadanie (4.1) jest wypukłe i zgodnie z twierdzeniem 2.3.24 punkt x^* jest rozwiązaniem optymalnym. Skoro mamy odwrócić macierz Lagrange'a H, powstaje pytanie, przy jakich założeniach macierz ta jest nieosobliwa. Odpowiedź daje poniższy lemat.

Lemat 4.1.3 *Jeśli A ma pełny rząd wierszowy i macierz G jest dodatnio określona, to macierz H jest nieosobliwa i zachodzi równość*

$$H^{-1} = \begin{bmatrix} G^{-1} - G^{-1}A^T(AG^{-1}A^T)^{-1}AG^{-1} & G^{-1}A^T(AG^{-1}A^T)^{-1} \\ (AG^{-1}A^T)^{-1}AG^{-1} & -(AG^{-1}A^T)^{-1} \end{bmatrix}. \tag{4.12}$$

Dowód. Zauważmy najpierw, że przy założeniach lematu macierz $AG^{-1}A^T$ jest również dodatnio określona. Istotnie. Jeśli $u \in \mathbb{R}^m$, $u \neq 0$, to $A^T u \neq 0$, bo wiersze macierzy A są liniowo niezależne. Skoro macierz G jest dodatnio określona, więc nietrudno zauważyć, że macierz G^{-1} jest również dodatnio określona. Wobec tego $u^T AG^{-1}A^T u = (A^T u)G^{-1}(A^T u) > 0$, co oznacza, że macierz $AG^{-1}A^T$ jest również dodatnio określona. Przemnażając przez siebie prawe strony równości (4.10) i (4.12) otrzymamy, że macierz stojąca po prawej stronie równości (4.12) jest odwrotnością macierzy H. ■

Często jednak macierz G nie jest dodatnio określona, a nawet nie jest nieosobliwa, a mimo to zadanie (4.1) może posiadać dokładnie jedno rozwiązanie. Dlatego, jeśli nie wiemy, czy macierz G jest dodatnio określona lepiej jest zastosować nastepujące twierdzenie, z którego wynika drugi sposób rozwiązania równania (4.9).

Twierdzenie 4.1.4 *Jeśli A ma pełny rząd wierszowy i hesjan zredukowany $Z^T GZ$ jest macierzą dodatnio określoną, to macierz Lagrange'a H jest nieosobliwa. W konsekwencji, istnieje dokładnie jeden punkt Kuhna–Tuckera (x^*, y^*) dany równością (4.11). Ponadto x^* jest jedynym rozwiązaniem optymalnym zadania (4.1).*

Dowód. Przypuśćmy, że dla $z = (x, y)$ zachodzi $Hz = 0$, czyli

$$\begin{bmatrix} Gx + A^T y \\ Ay \end{bmatrix} = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{4.13}$$

a więc $Ax = 0$ i w konsekwencji

$$x^T Gx = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

Ponieważ $x \in \ker A$ i, zgodnie z tym, co powiedzieliśmy wyżej, kolumny macierzy Z rozpinają bazę w przestrzeni $\ker A$, a więc $x = Zu$ dla pewnego $u \in \mathbb{R}^{n-m}$. Mamy więc

$$0 = x^T Gx = u^T Z^T GZu.$$

Ponieważ założyliśmy, że macierz $Z^T GZ$ jest dodatnio określona, więc $u = 0$, a zatem i $x = Zu = 0$. Z równości (4.13) wynika więc, że $A^T y = 0$. Ponieważ, zgodnie z założeniem, A ma pełny rząd wierszowy, więc $y = 0$. Pokazaliśmy więc, że $z = (x, y) = 0$, co dowodzi, że H jest nieosobliwa. Równanie (4.9) posiada więc dokładnie jedno rozwiązanie (x^*, y^*), które jest jedynym punktem Kuhna–Tuckera zadania (4.1). Ponieważ zredukowany hesjan jest macierzą dodatnio określoną, więc zredukowana funkcja celu r dana równością (4.5) posiada dokładnie jeden minimizer v^*, w konsekwencji zadanie (4.1) posiada jedyne rozwiązanie optymalne x^* dane równością (4.7). ■

Warunki wystarczające na to, aby punkt x^* był rozwiązaniem zadania (4.1) są podane w twierdzeniu 2.3.47. Twierdzenie to jest jednak mało wygodne w sytuacji, gdy n jest duże, bowiem należy wówczas policzyć $n - m$ wyznaczników macierzy wysokich stopni. Dlatego często łatwiej jest sprawdzić, czy hesjan zredukowany $Z^T GZ$ jest macierzą dodatnio określoną i skorzystać z twierdzenia 4.1.4. Zwróćmy przy tym uwagę na to, że rozwiązanie optymalne i wektor mnożników Lagrange'a można wyznaczyć albo z równości (4.11) albo z równości (4.7) i (4.8).

Uwaga 4.1.5 W ustępie 2.3.3 podaliśmy warunki wystarczające na to, aby dany punkt x^* był rozwiązaniem optymalnym zadania minimalizacji z ograniczeniami równościowymi (twierdzenia 2.3.42 i 2.3.47). Zwróciliśmy przy tym uwagę na niewygodę przy stosowaniu twierdzenia 2.3.47, gdy wymiar zadania n jest duży i $m \ll n$, gdyż trzeba wówczas liczyć wiele wyznaczników wysokich stopni. Uwaga ta dotyczy oczywiście również zadania programowania kwadratowego jako szczególnego przypadku zadania minimalizacji z ograniczeniami. W tej sytuacji lepiej jest korzystać z twierdzeń 2.3.43 i 2.3.44, gdzie wystarczy zbadać określoność hesjanu zredukowanego wymiaru $n - m$.

Ćwiczenie 4.1.6 Rozwiązać zadanie podane w ćwiczeniu 4.1.2 metodą Lagrange'a. Porównać otrzymany wynik z rozwiązaniem otrzymanym w ćwiczeniu 4.1.2.

4.1.2 Programowanie kwadratowe przy ograniczeniach nierównościowych

Zajmiemy się teraz zadaniem programowania kwadratowego postaci

$$\begin{array}{ll} \text{minimalizować} & q(x) = \frac{1}{2}x^T G x + g^T x \\ \text{względem} & x \in \mathbb{R}^n \\ \text{przy ograniczeniach} & Ax \leq b \end{array} \tag{4.14}$$

gdzie G jest macierzą symetryczną typu $n \times n$, $A = [a_1, ..., a_m]$ jest macierzą typu $m \times n$ o wierszach $a_i \in \mathbb{R}^n$, $g \in \mathbb{R}^n$, $b = (b_1, ..., b_m) \in \mathbb{R}^m$. Dla uproszczenia założymy ponadto, że G jest macierzą dodatnio określoną.

Metoda ograniczeń aktywnych

Przedstawimy postać pojedynczej k-tej iteracji. Przypuśćmy, że znane jest pewne rozwiązanie dopuszczalne x_k otrzymane w poprzedniej iteracji, które jest pewnym przybliżeniem rozwiązania optymalnego rozważanego zadania (4.14). Podamy jak w metodzie ograniczeń aktywnym otrzymać kolejne przybliżenie x_{k+1}.

Oznaczmy przez I zbiór wszystkich ograniczeń, czyli $I = \{1, ..., m\}$, przez I_k – zbiór pewnych ograniczeń aktywnych w punkcie x_k. W opisie algorytmu podamy, jakie ograniczenia aktywne włączamy do I_k. Na początku można przyjąć

$$I_1 = I(x_1) = \{i \in I : a_i^T x_1 = b_i\}.$$

Zakładamy przy tym dla uproszczenia, że w trakcie realizacji metody w każdym punkcie x_k wektory $a_i, i \in I_k$ są liniowo niezależne (warunek regularności ograniczeń Fiacco–McCormicka). Odpowiada to mniej więcej założeniu o niewystępowaniu degeneracji w zadaniu programowania liniowego. W razie potrzeby można zaburzyć prawą stronę ograniczeń, podobnie jak w przypadku wystąpienia degeneracji w zadaniu programowania liniowego. W każdej iteracji metody ograniczeń aktywnych wyznaczane jest rozwiązanie zadania minimalizacji kwadratowej funkcji celu q przy ograniczeniach równościowych $a_i^T x = b_i, i \in I_k$. Metody takiej minimalizacji zostały podane w poprzednim ustępie. Oznaczając $x = x_k + d$, $q_k = q(x_k) = \frac{1}{2}x_k^T G x_k + g^T x_k$ i $g_k = \nabla q(x_k) = Gx_k + g$ możemy funkcję q zapisać w postaci

$$\begin{aligned} q(x) &= \frac{1}{2}(x_k + d)^T G (x_k + d) + g^T (x_k + d) \\ &= \frac{1}{2} d^T G d + g_k^T d + q_k, \end{aligned}$$

zaś ograniczenia w postaci $a_i^T d = 0, i \in I_k$. Zatem zadanie to przyjmie postać

$$\begin{array}{ll} \text{minimalizować} & \frac{1}{2}d^T G d + g_k^T d \\ \text{względem} & d \in \mathbb{R}^n \\ \text{przy ograniczeniach} & a_i^T d = 0, i \in I_k \end{array} \tag{4.15}$$

i może być rozwiązane uprzednio podanymi metodami.

Jeśli rozwiązanie d_k tego zadania jest niezerowe i punkt $x_k + d_k$ jest dopuszczalny również dla ograniczeń $i \in I \setminus I_k$, to w omawianej metodzie ograniczeń aktywnych przyjmujemy $x_{k+1} = x_k + d_k$. Może się jednak zdarzyć, że $x_k + d_k$ nie będzie spełniać niektórych ograniczeń $i \in I \setminus I_k$. W tym przypadku poszukujemy metodą minimalizacji kierunkowej największego $\alpha \in [0, 1]$ takiego, że punkt $x_k + \alpha d_k$ spełnia wszystkie ograniczenia $i \in I \setminus I_k$, czyli

$$\alpha_k = \max\{\alpha \in [0, 1] : a_i^T (x_k + \alpha d_k) \leq b_i, i \in I \setminus I_k\}.$$

Zapiszemy α_k w wygodniejszej postaci. Niech d_k będzie rozwiązaniem zadania (4.15) i niech $a_i^T(x_k + d_k) > b_i$, czyli $a_i^T d_k > 0$ dla pewnego $i \in I \setminus I_k$. Ponieważ $a_i^T(x_k + \alpha d_k) = b_i$ dla $\alpha = \frac{b_i - a_i^T x_k}{a_i^T d_k}$, więc nietrudno zauważyć, że

$$\alpha_k = \min\{1, \min_{i \notin I_k, a_i^T d_k > 0} \frac{b_i - a_i^T x_k}{a_i^T d_k}\}. \tag{4.16}$$

Przyjmujemy teraz $x_{k+1} = x_k + \alpha_k d_k$. Jeśli $\alpha_k < 1$, to oczywiście pewne nowe ograniczenie, powiedzmy $i_k \in I \setminus I_k$ staje się aktywne i dodajemy je do zbioru I_k, czyli przyjmujemy $I_{k+1} = I_k \cup \{i_k\}$ (oczywiście więcej niż jedno nowe ograniczenie może stać się aktywne, ale do zbioru I_k dołączamy tylko jedno). W przeciwnym wypadku ($\alpha_k = 1$) przyjmujemy $I_{k+1} = I_k$.

Przypuśćmy teraz, że $d_k = 0$ jest rozwiązaniem zadania (4.15). Fakt ten można stwierdzić sprawdzając na przykład, czy dla $d_k = 0$ istnieje wektor mnożników Lagrange'a dla zadania (4.15), bo jest to zadanie programowania wypukłego i spełnienie układu KT jest wówczas warunkiem wystarczającym na to, aby $d_k = 0$ było rozwiązaniem zadania (4.15). Przypadek, w którym $d_k = 0$ jest rozwiązaniem zadania (4.15) może się zdarzyć na przykład wówczas, gdy $\alpha_{k-1} = 1$, czyli gdy $x_k = x_{k-1} + d_{k-1}$. Wówczas wyznaczamy odpowiadające temu rozwiązaniu mnożniki Lagrange'a $\lambda_i, i \in I_k$ (w poprzednim ustępie podano jak je wyznaczyć). Jeśli wszystkie $\lambda_i \geq 0, i \in I_k$, to przyjmując $\lambda_i = 0$ dla $i \in I \setminus I_k$ otrzymamy, że (x_k, λ) jest punktem Kuhna–Tuckera dla zadania (4.14). Ponieważ – jak założyliśmy – G jest macierzą dodatnio określoną, więc zadanie (4.14) jest zadaniem programowania wypukłego i punkt x_k jest w tym przypadku rozwiązaniem globalnym tego zadania. Przypuśćmy teraz, że $\lambda_{j_k} < 0$ dla pewnego $j_k \in I_k$. Zgodnie z interpretacją mnożników Lagrange'a podaną w ustępie 2.3.4 usunięcie ograniczenia j_k ze zbioru I_k prowadzi do redukcji optymalnej wartości funkcji celu zadania (4.15). Zatem w przypadku $\lambda_{j_k} < 0$ przyjmujemy $I_k := I_k \setminus \{j_k\}$ i rozwiązujemy zadanie postaci (4.15) ale przy nowych ograniczeniach równościowych $i \in I_k$. Jeśli $\lambda_j < 0$ dla większej liczby ograniczeń $j \in I_k$, to zwykle przyjmuje się

$$j_k = \operatorname{argmin}\{\lambda_j : j \in I_k\}. \tag{4.17}$$

Przedstawimy teraz dokładniej postać algorytmu wyznaczającego rozwiązanie zadania (4.14)

Algorytm 4.1.7 (metoda ograniczeń aktywnych)

Krok 0 (*inicjalizacja*)

(a) Wybrać punkt startowy x_1.

(b) Położyć $I_1 = I(x_1) = \{i \in I : a_i^T x_1 = b_i\}$.

(c) Położyć $k = 1$.

Krok 1 (*minimalizacja przy ograniczeniach równościowych*)

(a) Sprawdzić, czy $d = 0$ jest rozwiązaniem zadania (4.15); jeśli nie, to przejść do kroku 2.

(b) Wyznaczyć mnożniki Lagrange'a $\lambda_i, i \in I_k$ dla zadania (4.15) i wyznaczyć j_k takie, że $\lambda_{j_k} = \min\{\lambda_j : j \in I_k\}$

(c) Jeśli $\lambda_{j_k} \geq 0$, to algorytm zatrzymuje się i $x^* = x_k$ jest rozwiązaniem optymalnym zadania (4.14); w przeciwnym przypadku przyjąć $I_k := I_k \setminus \{j_k\}$

Krok 2 (*wyznaczenie kierunku poszukiwań*)

(a) Wyznaczyć rozwiązanie d_k zadania (4.15).

(b) Wyznaczyć α_k z równości (4.16).

(c) Położyć $x_{k+1} = x_k + \alpha_k d_k$.

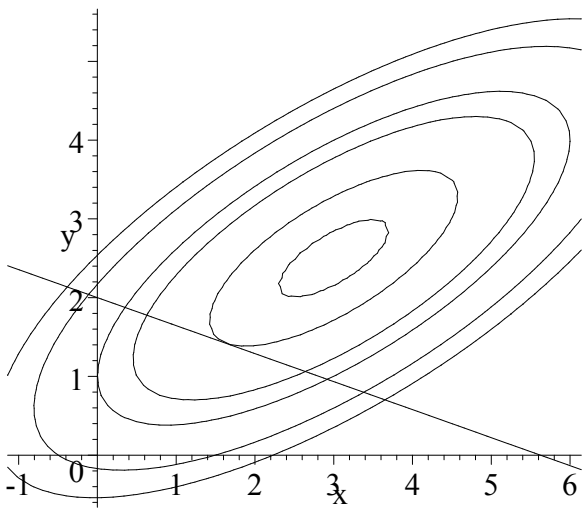
Krok 3 (*aktualizacja zbioru ograniczeń równościowych*)

(a) Jeśli $\alpha_k < 1$, to położyć $I_{k+1} = I_k \cup \{i_k\}$, gdzie i_k realizuje minimum w równości (4.16). W przeciwnym wypadku położyć $I_{k+1} = I_k$.

(b) Położyć $k := k + 1$ i przejść do kroku 1.

Interpretacja geometryczna metody ograniczeń aktywnych. Zilustrujemy opisaną metodę dla zadania

$$\begin{array}{ll} \text{minimalizować} & q(x,y) = (x-0.5)^2 + 2(y-1)^2 - 2xy \\ \text{przy ograniczeniach} & 16x + 45y \leq 90 \\ & x \geq 0 \\ & y \geq 0 \end{array}$$

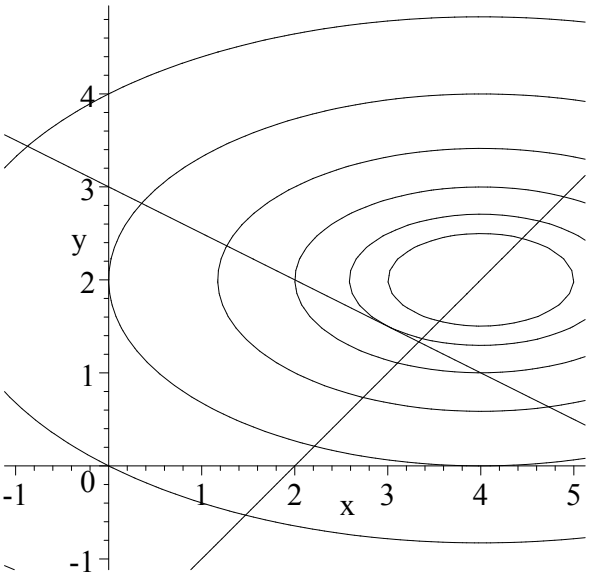


W punkcie startowym $z_1 = (x_1, y_1) = (0,0)$ drugie i trzecie ograniczenie jest aktywne, a więc $I_1 = \{2,3\}$. Wobec tego $d = 0$ jest rozwiązaniem zadania (4.15) dla wyznaczonego zbioru I. Wyznaczając mnożniki Lagrange'a otrzymujemy $\lambda_2 = -1, \lambda_3 = -4$, a więc $j_1 = 3$ i $\lambda_{j_1} = -4$ w kroku 1(b). Zgodnie z krokiem 1(c) przyjmujemy $I_1 := I_1 \setminus \{3\} = \{2\}$. W kroku 2(a) rozwiązujemy zadanie (4.15) dla nowo wyznaczonego zbioru I_1 i otrzymujemy jego rozwiązanie $d_1 = (0,1)$. Następnie w kroku 2(c) wyznaczamy α_1 z równości (4.16) i otrzymujemy $\alpha_1 = 1$. W kroku 2(c) kładziemy więc $z_2 = z_1 + d_1 = (0,1)$ i w kroku 3(a) kładziemy $I_2 = I_1 = \{2\}$ i po zwiększeniu k do 2 wracamy do kroku 1(a). Stwierdzamy, że $d = 0$ jest rozwiązaniem zadania (4.15) dla zbioru $I_2 = \{2\}$ i w kroku 1(b) wyznaczamy mnożniki Lagrange'a. Otrzymujemy $\lambda_2 = -3$, a więc w kroku 1(c) przyjmujemy $I_2 := I_2 \setminus \{2\} = \emptyset$. W kroku 2(a) rozwiązujemy zadanie (4.15) dla $I_2 = \emptyset$ i otrzymujemy jego rozwiązanie $d_2 = (3, 1.5)$ i odpowiadające mu rozwiązanie $z = (3, 2.5)$. Wyznaczenie α_2 z równości (4.16) daje $\alpha_2 \simeq 0.43 < 1$, więc w punkcie

$z_3 = z_2 + \alpha_2 d_2 \simeq (1.28, 1.64)$ pierwsze ograniczenie staje się aktywne i w kroku 3(a) kładziemy $I_3 = I_2 \cup \{1\} = \{1\}$. Po zwiększeniu k do 3 wracamy do kroku 1(a). Ponieważ $d = 0$ nie jest rozwiązaniem zadania (4.15) dla $I_3 = \{1\}$, więc przechodzimy do kroku 2(a) i rozwiązujemy zadanie (4.15) tym razem dla $I_3 = \{1\}$ i otrzymujemy $d_3 \simeq (0.36, -0.22)$. Wyznaczone w kroku 2(b) $\alpha_3 = 1$. Wobec tego w kroku 2(c) otrzymujemy kolejne przybliżenie rozwiązania $z_4 = z_2 + d_3 \simeq (1.66, 1.42)$. W kroku 3(a) kładziemy $I_4 = I_3 = \{1\}$, następnie zwiększamy k do 4 i wracamy do kroku 1(a). Widzimy, że wektor $d = 0$ jest rozwiązaniem zadania (4.15) dla $I_4 = \{1\}$. Wobec tego wyznaczamy odpowiadający temu rozwiązaniu wektor mnożników Lagrange'a. Otrzymujemy $\lambda_1 \simeq 0.035 > 0$. W wyjściowym zadaniu możemy przyjąć pozostałe mnożniki Lagrange'a równe zero, bo pozostałe ograniczenia są nieaktywne. Algorytm zatrzymuje się i stwierdzamy, że $z_4 \simeq (1.64, 1.42)$ jest rozwiązaniem optymalnym rozważanego zadania.

Ćwiczenie 4.1.8 Prześledzić metodę ograniczeń aktywnych dla zadania

$$\begin{array}{ll} \text{minimalizować} & q(x,y) = (x-4)^2 + 4(y-2)^2 \\ \text{przy ograniczeniach} & 3x + 6y \leq 18 \\ & x - y \leq 2 \\ & x \geq 0 \\ & y \geq 0 \end{array}$$



4.2 Sekwencyjne programowanie kwadratowe

Dane jest zadania minimalizacji różniczkowalnej z ograniczeniami równościowymi

$$\begin{array}{ll} \text{minimalizować} & f(x) \\ \text{względem} & x \in \mathbb{R}^n \\ \text{przy ograniczeniach} & c_i(x) = 0,\ i = 1, 2, ..., m, \end{array} \tag{4.18}$$

gdzie funkcja celu f i ograniczenia c_i, $i \in I := \{1, 2, ..., m\}$ są dwukrotnie różniczkowalne w sposób ciągły. Niech $x_k \in \mathbb{R}^n$ będzie pewnym przybliżeniem rozwiązania tego zadania. Pojedyncza iteracja sekwencyjnego programowania kwadratowego polega na zastąpieniu zadania (4.18) przez pewien model kwadratowy i wyznaczeniu kolejnego punktu x_{k+1} jedną z metod programowania kwadratowego przedstawionych w ustępie 4.1. Oznaczmy $c(x) = (c_1(x), ...c_m(x))$. Wówczas macierz Jacobiego odwzorowania c ma postać $\nabla c(x) = [\nabla c_1(x), ..., \nabla c_m(x)]$. Warunki Kuhna–Tuckera dla zadania (4.18) mają postać układu $n + m$ równań z $n + m$ niewiadomymi:

$$\begin{aligned} \nabla_x L(x, y) = \nabla f(x) + y^T \nabla c(x) = 0, \\ \nabla_y L(x, y) = c(x) = 0, \end{aligned} \tag{4.19}$$

który zapiszemy skrótowo $F(z) = 0$, gdzie $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, zaś $F(z) = (\nabla f(x) + y^T \nabla c(x), c(x)) \in \mathbb{R}^n \times \mathbb{R}^m$. Dla układu tego zastosujemy metodę Newtona $z_{k+1} = z_k - (\nabla F(z_k))^{-1} F(z_k)$, przy założeniu, że macierz $\nabla F(z_k)$ jest nieosobliwa. Zauważmy, że

$$\nabla F(z) = \begin{bmatrix} \nabla_{xx} L(x, y) & \nabla_{xy} L(x, y) \\ \nabla_{yx} L(x, y) & \nabla_{yy} L(x, y) \end{bmatrix} = \begin{bmatrix} \nabla^2 f(x) + y^T \nabla^2 c(x) & \nabla c(x) \\ (\nabla c(x))^T & 0 \end{bmatrix}$$

Zatem metoda Newtona zastosowana do układu (4.19) ma postać

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} \nabla^2 f(x_k) + y_k^T \nabla^2 c(x_k) & (\nabla c(x_k))^T \\ \nabla c(x_k) & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla f(x_k) + y_k^T \nabla c(x_k) \\ c(x_k) \end{bmatrix}$$

Zwróćmy przy okazji uwagę na to, że w przypadku, gdy funkcja celu jest kwadratowa, a ograniczenia liniowe, macierz $\nabla F(z_k)$ pokrywa się z macierzą Lagrange'a (4.10). Zatem opisana iteracja pokrywa się z opisaną w ustępie 4.1.1 metodą Lagrange'a dla funkcji kwadratowej bedącej przybliżeniem kwadratowym funkcji Lagrange'a w punkcie (x_k, y_k) (patrz definicja 1.5.24) i ograniczeń będących linearyzacjami funkcji c_i, $i \in I$, w punkcie x_k (patrz definicja 1.5.23).