ΜΙΚΤΕΣ ΜΕΘΟΔΟΙ ΚΛΙΣΗΣ-ΠΟΙΝΩΝ

Έστω συναρτήσεις $f_i \in C^1(\mathbb{R}^n)$, $i = 0,...,m$, και $U \subset \mathbb{R}^n$ ένα ***κυρτό κλειστό*** σύνολο. Θέτοντας

$$W := \{x \in U \mid f_i(x) \le 0,\ i = 1,...,l,\ \ f_i(x) = 0,\ i = l+1,...,m\},$$

θεωρούμε το ακόλουθο ***πρόβλημα ελαχιστοποίησης με περιορισμό και δεσμεύσεις***

(1) Να βρεθεί $\bar{x} \in W$ τέτοιο ώστε $f(\bar{x}) = \min_{x \in W} f(x)$.

Ένα σημείο $x \in W$ καλείται ***αποδεκτό***. Από το Θεώρημα Kuhn-Tucker-Lagrange (K-T-L), αν το $\bar{x}$ είναι λύση του προβλήματος (1), τότε υπάρχουν πολλαπλασιαστές λ_i, όχι όλοι μηδέν, με $\lambda_i \ge 0$, $i = 0,...,l$, τέτοιοι ώστε

(2) $\displaystyle\sum_{i=0}^{m} \lambda_i \nabla f_i(\bar{x})(y - \bar{x}) \ge 0$, για κάθε $y \in U$,

(3) $\lambda_i f_i(\bar{x}) = 0$, $i = 1,...,l$.

Έστω $\gamma > 0$, $b, c \in (0,1)$, (s_k) μία ακολουθία στο $(0,1]$, (M_i^j), $i = 1,...,l$, θετικές αύξουσες ακολουθίες που τείνουν στο $+\infty$, και (β^j) μια θετική φθίνουσα ακολουθία που τείνει στο 0. Οι ***μεικτές μέθοδοι κλίσης-ποινών, προβεβλημένης κλίσης-ποινών και Frank-Wolfe-ποινών, με βέλτιστο βήμα, ή με βήμα Armijo***, είναι επαναληπτικές μέθοδοι οι οποίες βρίσκουν στο όριο σημεία $\bar{x}$ που ικανοποιούν τις αναγκαίες συνθήκες K-T-L (2) και (3).

Ορίζουμε πρώτα τις ***ποινικοποιημένες συναρτήσεις***

$$f^j(x) := f_0(x) + \frac{1}{2}\left(\sum_{i=1}^{l} \{M_i^j [\max(0, f_i(x))]^2\} + \sum_{i=l+1}^{m} \{M_i^j [f_i(x)]^2\} \right).$$

Θεωρώντας τις σύνθετες συναρτήσεις $\phi(f^j(x))$, όπου $\phi(t) = [\max(0,t)]^2/2$ (με $\phi'(t) = \max(0,t)$), και $[f^j(x)]^2$, υπολογίζουμε την κλίση της f^j

$$\nabla f^j(x) := \nabla f_0(x) + \sum_{i=1}^{l} \{M_i^j [\max(0, f_i(x))] \nabla f_i(x)\} + \sum_{i=l+1}^{m} \{M_i^j f_i(x) \nabla f_i(x)\}.$$

Έχουμε επίσης $f^j \in C^1(\mathbb{R}^n)$, αφού τα ∇f^j είναι συνεχή, δηλ. $\nabla f^j \in [C(\mathbb{R}^n)]^n$.

Αλγόριθμος 3

Βήμα 1. Θέτουμε $k := 0$, $j := 1$, και επιλέγουμε ένα αρχικό $x_0 \in U$.

Βήμα 2. Βρίσκουμε ένα ***σημείο κατεύθυνσης*** y_k (δηλ. μια ***κατεύθυνση*** $y_k - x_k$) με έναν από τους τρεις παρακάτω τρόπους:

(Κ) Μέθοδος της ***Κλίσης*** ($U := \mathbb{R}^n$):

Θέτουμε $y_k = x_k - \nabla f^j(x_k)$, δηλ. $y_k - x_k = -\nabla f^j(x_k)$.

(ΠΚ) Μέθοδος ***Προβεβλημένης Κλίσης***: Βρίσκουμε $y_k \in U$ τέτοιο ώστε

(3) $\displaystyle\zeta_k := \nabla f^j(x_k)(y_k - x_k) + \frac{\gamma}{2}\|y_k - x_k\|_2^2 = \min_{y \in U}[\nabla f^j(x_k)(y - x_k) + \frac{\gamma}{2}\|y - x_k\|_2^2]$,

[Το y_k είναι η ***προβολή*** του διανύσματος $z_k := x_k - \dfrac{1}{\gamma}\nabla f^j(x_k)$ πάνω στο U.]

(FW) Μέθοδος ***Frank-Wolfe*** ($\gamma = 0$): Βρίσκουμε ένα $y_k \in U$ τέτοιο ώστε

(4) $\nabla f^j(x_k)(y_k - x_k) = \min_{y\in U} \nabla f^j(x_k)(y - x_k)$.

Και στις τρεις περιπτώσεις, θέτουμε $\delta_k := \nabla f\ (x_k)(y_k - x_k)$.

Βήμα 3. Αν $|\delta_k| \le \beta^j$, θέτουμε $x^j := x_k$, $y^j := y_k$, $\delta^j := \delta_k$, [$\zeta^j := \zeta_k$, ΠΚ,] $j := j+1$, και πηγαίνουμε στο Βήμα 2. Αλλιώς, πηγαίνουμε στο Βήμα 4.

Βήμα 4. Βρίσκουμε ένα ***βήμα*** α_k με έναν από τους δύο παρακάτω τρόπους:

(α) ***Βέλτιστο Βήμα***: Βρίσκουμε ένα $\alpha_k \in [0,1]$ τέτοιο ώστε

$$f^j(x_k + \alpha_k(y_k - x_k)) - f^j(x_k) = \min_{\substack{\alpha\ge 0\ (\mathrm{K}) \\ \alpha\in[0,1]\ (\Pi\mathrm{K,FW})}} [f^j(x_k + \alpha(y_k - x_k)) - f^j(x_k)].$$

(β) ***Βήμα Armijo***: Αρχικά θέτουμε $\alpha := s_k$. Αν το α ***δεν*** ικανοποιεί την ανισότητα

$$f^j(x_k + \alpha(y_k - x_k)) - f^j(x_k) \le \alpha b \delta_k,$$

θέτουμε επαναληπτικά $\alpha := c\alpha$ και επιλέγουμε για α_k το ***πρώτο*** α που την ικανοποιεί. Αν την ικανοποιεί, θέτουμε επαναληπτικά $\alpha := \alpha / c$ και επιλέγουμε για α_k το ***τελευταίο*** $\alpha \ge 0$ (K), ή $\alpha \in (0,1]$ (ΠΚ,FW, ώστε να ισχύει $x_{k+1} \in U$ στο Βήμα 5) που την ικανοποιεί.

Βήμα 5. Θέτουμε $x_{k+1} := x_k + \alpha_k(y_k - x_k)$, $k := k+1$, και πηγαίνουμε στο Βήμα 2.

Με τα x^j ορισμένα στο Βήμα 3, ορίζουμε τις ***ακολουθίες πολλαπλασιαστών***
$\lambda_i^j := M_i^j[\max(0, f_i(x^j))]$, $i = 1,...,l$, $\lambda_i^j := M_i^j f_i(x^j)$, $i = l+1,...,m$.

Θεώρημα 3. Υποθέτουμε ότι η f_0 (και άρα η κάθε f^j) είναι πιεστική αν το U είναι μη φραγμένο, και στη μέθοδο Frank-Wolfe-ποινών ότι το U είναι φραγμένο, άρα συμπαγές. Αν οι ακολουθίες (λ_i^j) είναι φραγμένες, τότε κάθε όριο $\overline{x}$ (αν υπάρχει, και υπάρχει αν το U είναι συμπαγές) συγκλίνουσας υπακολουθίας της ακολουθίας (x^j) που κατασκευάζεται στο Βήμα 3 είναι αποδεκτό και ικανοποιεί τις συνθήκες K-T-L (2) και (3).

Απόδειξη. Θα δείξουμε πρώτα ότι $j \to \infty$ στον Αλγόριθμο 3. Πράγματι, αν αυτό δεν ίσχυε, τότε το j θα παρέμενε σταθερό μετά από ένα πεπερασμένο αριθμό επαναλήψεων ως προς k. Από το Θεώρημα 1 ή 2, όμως, θα είχαμε τότε $\delta_k \to 0$, και από το Βήμα 3, $j \to \infty$ (αφού $\beta^j \to 0$), που είναι άτοπο. Συνεπώς, $j \to \infty$ στον Αλγόριθμο 3, και άρα $\delta^j \to 0$, και $\zeta^j \to 0$ στη μέθοδο ΠΚ, αφού $\delta^j \le \zeta^j \le 0$.

Έστω $(x^j)_{j\in J}$ μια υπακολουθία της (x^j) (βλ. Βήμα 3) που συγκλίνει σε κάποιο $\overline{x} \in U$. Επειδή οι ακολουθίες (λ_i^j) είναι φραγμένες (ανήκουν σε μια κλειστή μπάλα), υπάρχουν υπακολουθίες $(\lambda_i^j)_{j\in L\subset J}$ τέτοιες ώστε $\lambda_i^j \to \lambda_i$. Έχουμε τότε

$$0 = \lim_{\substack{j\to\infty \\ j\in L}} \frac{\lambda_i^j}{M_i^j} = \lim_{\substack{j\to\infty \\ j\in L}} [\max(0, f_i(x^j))] = \max(0, f_i(\overline{x})), \quad i = 1,...,l,$$

$$0 = \lim_{\substack{j\to\infty \\ j\in L}} \frac{\lambda_i^j}{M_i^j} = \lim_{\substack{j\to\infty \\ j\in L}} f_i(x^j) = f_i(\overline{x}), \quad i = l+1,...,m,$$

που δείχνουν ότι το $\bar{x}$ είναι αποδεκτό. Από το Βήμα 2, για τυχόν $y \in U$, έχουμε

(K) $\quad -\left\| \sum_{i=0}^{m} \lambda_i^j \nabla f_i(x^j) \right\|_2^2 = \delta^j$,

(ΠΚ) $\quad \sum_{i=0}^{m} \lambda_i^j \nabla f_i(x^j)(y - x^j) + \frac{1}{\gamma} \left\| y - x^j \right\|_2^2 \geq \zeta^j$,

(FW) $\quad \sum_{i=0}^{m} \lambda_i^j \nabla f_i(x^j)(y - x^j) \geq \delta^j$,

όπου θέσαμε $\lambda_0^j := 1$, και παίρνοντας το όριο όταν $j \to \infty$, $j \in L$, βρίσκουμε

(K) $\quad -\left\| \sum_{i=0}^{m} \lambda_i \nabla f_i(\bar{x}) \right\|_2^2 = 0$, δηλ. $\sum_{i=0}^{m} \lambda_i \nabla f_i(\bar{x}) = 0$,

(ΠΚ) $\quad \sum_{i=0}^{m} \lambda_i \nabla f_i(\bar{x})(y - \bar{x}) + \frac{1}{\gamma} \left\| y - \bar{x} \right\|_2^2 \geq 0$,

(FW) $\quad \sum_{i=0}^{m} \lambda_i \nabla f_i(\bar{x})(y - \bar{x}) \geq 0$,

με $\lambda_0 := 1$, και αυτά ισχύουν για κάθε $y \in U$. Με επιχείρημα παρόμοιο με του Θεωρήματος 2 (με το θ), καταλήγουμε και για τη μέθοδο ΠΚ στη συνθήκη

$$\sum_{i=0}^{m} \lambda_i \nabla f_i(\bar{x})(y - \bar{x}) \geq 0, \quad \text{για κάθε } y \in U.$$

Από τον ορισμό των λ_i^j, έχουμε $\lambda_i^j \geq 0$, $i = 0, \ldots, l$, άρα και στο όριο $\lambda_i \geq 0$, $i = 0, \ldots, l$, όπου τα λ_i δεν είναι όλα 0, αφού $\lambda_0 := 1$. Τέλος, αν $f_i(\bar{x}) < 0$ για κάποιο δείκτη $i \in \{1, \ldots, l\}$, τότε $f_i(x^j) < 0$ και $\lambda_i^j = 0$ για j αρκετά μεγάλο, από τον ορισμό του λ_i^j και αφού $f(x^j) \to f(\bar{x})$, και άρα στο όριο $\lambda_i = 0$, δηλ. ισχύουν οι συνθήκες εγκαρσιότητας (3).

Σχόλιο

Στην πράξη, για να εξασφαλίσουμε ότι οι ακολουθίες (λ_i^j) παραμένουν φραγμένες, καλό είναι να διαλέγουμε ακολουθίες (M_i^j) που συγκλίνουν σχετικά ***αργά*** στο $+\infty$ και μια ακολουθία (β^j) που συγκλίνει σχετικά ***γρήγορα*** στο 0 (αυτό δικαιολογείται και από την καθαρή μέθοδο ποινών).