## ΜΕΘΟΔΟΙ ΠΡΟΒΕΒΛΗΜΕΝΗΣ ΚΛΙΣΗΣ ΚΑΙ FRANK-WOLFE

---

Έστω $f \in C^1(\mathbb{R}^n)$ μια συνάρτηση και $U \subset \mathbb{R}^n$ ένα ***κυρτό κλειστό*** σύνολο. Θεωρούμε το ακόλουθο ***πρόβλημα ελαχιστοποίησης με περιορισμό (χωρίς δεσμεύσεις)***

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

Από το γνωστό θεώρημα, αν το $\bar{x}$ είναι λύση του προβλήματος (1), τότε ικανοποιεί την αναγκαία συνθήκη

(2) $\nabla f(\bar{x})(y - \bar{x}) \geq 0$, για κάθε $y \in U$.

Έστω $\gamma > 0$, $b, c \in (0,1)$, και $(s_k)$ μία ακολουθία στο $(0,1]$. Οι ***μέθοδοι προβεβλημένης κλίσης και Frank-Wolfe, με βέλτιστο βήμα, ή με βήμα Armijo***, είναι επαναληπτικές μέθοδοι καθόδου οι οποίες βρίσκουν στο όριο σημεία $\bar{x}$ που ικανοποιούν τη συνθήκη (2). Περιγράφονται από τον ακόλουθο αλγόριθμο.

**Αλγόριθμος 2**

*Βήμα 1.* Θέτουμε $k := 0$ 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 δύο παρακάτω τρόπους:

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

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

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

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

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

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

*Βήμα 3.* Αν $\delta_k = 0$, σταματάμε. Αλλιώς, πηγαίνουμε στο Βήμα 4.

*Βήμα 4.* Βρίσκουμε ένα ***βήμα*** $\alpha_k$ με έναν από τους δύο παρακάτω τρόπους:

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

$$f(x_k + \alpha_k(y_k - x_k)) - f(x_k) = \min_{\alpha \in [0,1]}[f(x_k + \alpha(y_k - x_k)) - f(x_k)].$$

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

$$f(x_k + \alpha(y_k - x_k)) - f(x_k) \leq \alpha b \delta_k,$$

θέτουμε επαναληπτικά $\alpha := c\alpha$ και επιλέγουμε για $\alpha_k$ το ***πρώτο*** $\alpha$ που την ικανοποιεί. Αν την ικανοποιεί, θέτουμε επαναληπτικά $\alpha := \alpha / c$ και επιλέγουμε για $\alpha_k$ το ***τελευταίο*** $\alpha \in (0,1]$ (ώστε να ισχύει $x_{k+1} \in U$ στο Βήμα 4) 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.

**Θεώρημα 2.** Στη μέθοδο προβεβλημένης κλίσης, υποθέτουμε ότι το σύνολο

$$S := \{x \in U \,|\, f(x) \leq f(x_0)\}$$

είναι συμπαγές, και στη μέθοδο Frank-Wolfe ότι το $U$ είναι συμπαγές. Αν ο αλγόριθμος σταματά για κάποιο $k$, τότε $\delta_k = 0$ και το $\bar{x} := x_k$ ικανοποιεί τη συνθήκη (2). Αλλιώς, η ακολουθία $(x_k)$ που κατασκευάζει ο Αλγόριθμος 2 είναι άπειρη, κάθε όριο $\bar{x}$ συγκλίνουσας υπακολουθίας της $(x_k)$ ικανοποιεί τη συνθήκη (2), και $\delta_k \to 0$ (και $\zeta_k \to 0$ στη μέθοδο ΠΚ), για ολόκληρες τις ακολουθίες.

**Απόδειξη.** Θα δείξουμε πρώτα επαγωγικά ότι, εφόσον $\delta_k \neq 0$, η ακολουθία $(x_k)$ είναι καλά ορισμένη, ότι $x_k \in S$ για κάθε $k$, και ότι η ακολουθία $(f(x_k))$ είναι φθίνουσα. Έχουμε $x_0 \in S$. Ας υποθέσουμε ότι $x_k \in S$ και $\delta_k \neq 0$.

(ΠΚ) Εύκολα βλέπουμε ("συμπληρώνοντας το τετράγωνο") ότι η ελαχιστοποίηση (3) ισοδυναμεί με την ελαχιστοποίηση ως προς $y$ του τετράγωνου νόρμας (άρα και της νόρμας)

$$\|y - [x_k - (1/\gamma)\nabla f(x_k)]\|_2^2$$

στο κυρτό κλειστό σύνολο $U$, δηλ. το $y_k$ είναι η μοναδική προβολή του διανύσματος

$$z_k := x_k - (1/\gamma)\nabla f(x_k)$$

πάνω στο $U$, και γράφουμε $y_k = P_U z_k$.

(FW) Η ελαχιστοποίηση (4) μιας γραμμικής, άρα συνεχούς, συνάρτησης πάνω στο συμπαγές (από την υπόθεση) σύνολο $U$ έχει τουλάχιστο μια λύση $y_k$.

Παρατηρώντας πρώτα ότι $\delta_k \leq 0$ και στις δύο περιπτώσεις (ΠΚ), (FW), δείχνουμε, παρόμοια με το Θεώρημα 1, ότι το βήμα $\alpha_k$ και το $x_{k+1}$ υπάρχουν και στις δύο περιπτώσεις 4α και 4β. Επιπλέον, τα βήματα 4α, 4β και 5 δείχνουν ότι σε όλες τις περιπτώσεις ιχύει $f(x_{k+1}) - f(x_k) \leq 0$, και άρα $x_{k+1} \in S$. Συνεπώς, $x_k \in S$ για κάθε $k$, και η ακολουθία $(f(x_k))$ είναι πάντα φθίνουσα.

Αν ο αλγόριθμος σταματά για κάποιο $k$, τότε $\delta_k = 0$ (και $\zeta_k = 0$ στη μέθοδο ΠΚ, αφού $\delta_k \leq \zeta_k$). Θέτουμε $\bar{x} := x_k$.

(ΠΚ) Έχουμε εδώ

$$\nabla f(\bar{x})(y - \bar{x}) + \frac{1}{\gamma}\|y - \bar{x}\|_2^2 \geq \zeta_k = 0, \quad \text{για κάθε } y \in U.$$

Αντικαθιστώντας το $y$ με $\bar{x} + \theta(y - \bar{x})$, διαιρώντας με $\theta \in (0,1]$, και παίρνοντας το όριο όταν $\theta \to 0$, βρίσκουμε

$$\nabla f(\bar{x})(y - \bar{x}) \geq 0, \quad \text{για κάθε } y \in U.$$

(FW) Έχουμε εδώ

$$\nabla f(\bar{x})(y - \bar{x}) \geq \delta_k = 0, \quad \text{για κάθε } y \in U.$$

Αν τώρα ότι η ακολουθία $(x_k)$ είναι άπειρη, θα δείξουμε ότι τότε $\delta_k \to \delta := 0$.

(ΠΚ) Επειδή το $S$ είναι συμπαγές και $x_k \in S$ για κάθε $k$, υπάρχει υπακολουθία $(x_k)_{k\in K}$ που συγκλίνει σε κάποιο $\bar{x} \in S$. Επειδή $f \in C^1(\mathbb{R}^n)$, ισχύει

$$z_k := x_k - (1/\gamma)\nabla f(x_k) \to \bar{z} := \bar{x} - (1/\gamma)\nabla f(\bar{x}), \quad \text{όταν } k \to \infty,\ k \in K.$$

Επειδή $y_k = P_U z_k$, από τη συνέχεια του τελεστή προβολής $P_U$ έχουμε $y_k \to \bar{y}$, $k \in K$, όπου $\bar{y} = P_U \bar{z}$. Έχουμε προφανώς $\delta_k \leq \zeta_k \leq 0$, άρα

$$\zeta_k := \nabla f(x_k)(y_k - x_k) + \frac{\gamma}{2}\|y_k - x_k\|_2^2 \to \zeta := \nabla f(\bar{x})(\bar{y} - \bar{x}) + \frac{\gamma}{2}\|\bar{y} - \bar{x}\|_2^2 \leq 0,$$

$$\delta_k := \nabla f(x_k)(y_k - x_k) \to \delta := \nabla f(\bar{x})(\bar{y} - \bar{x}) \leq \zeta \leq 0.$$

(FW) Επειδή εδώ το $U$ είναι συμπαγές και $x_k \in U$ για κάθε $k$, υπάρχουν υπακολουθίες $(x_k)_{k\in K}$, $(y_k)_{k\in K}$ που συγκλίνουν σε κάποια $\overline{x}, \overline{y} \in U$, αντίστοιχα. Έχουμε $\delta_k \le 0$, άρα

$$\delta_k := \nabla f(x_k)(y_k - x_k) \to \delta := \nabla f(\overline{x})(\overline{y} - \overline{x}) \le \zeta \le 0 .$$

Ας υποθέσουμε τώρα ότι $\delta < 0$. Από το Θεώρημα Μέσης Τιμής και αφού $b \in (0,1)$, έχουμε

$$\begin{aligned} f(x_k + \alpha(y_k - x_k)) - f(x_k) &= \alpha \nabla f(x_k + \mu_\alpha (y_k - x_k))^T (y_k - x_k) \quad (\mu_\alpha \in (0,\alpha)) \\ &= \alpha[\nabla f(\overline{x}) + \varepsilon_{k\alpha}]^T [(\overline{y} - \overline{x}) + \varepsilon'_k] = \alpha(\delta + \eta_{k\alpha}) \le \alpha b \delta \quad \textbf{και} \quad \le \alpha b \delta_k , \end{aligned}$$

για $\alpha \in [0, \alpha']$ και $k \ge k'$, για κάποια $\alpha', k'$, όπου $\varepsilon_{k\alpha} \to 0$, $\varepsilon'_k \to 0$, $\eta_{k\alpha} \to 0$, όταν $k \to \infty$, $k \in K$, και $\alpha \to 0^+$.

Και στις δύο περιπτώσεις (ΠΚ), (FW), καταλήγουμε έτσι σε άτοπο όπως στο Θεώρημα 1, και άρα $\delta_k \to \delta = 0$, για ολόκληρη η ακολουθία. Στην περίπτωση (ΠΚ), έχουμε επιπλέον $\zeta_k \to \zeta = 0$, αφού $\delta_k \le \zeta_k \le 0$.

Στην περίπτωση (ΠΚ), έστω τυχόν $y \in U$. Από το Βήμα 2(ΠΚ), έχουμε

$$\nabla f(x_k)(y - x_k) + \frac{\gamma}{2} \| y - x_k \|_2^2 \ge \zeta_k ,$$

και παίρνοντας το όριο όταν $k \to \infty$, $k \in K$, βρίσκουμε

$$\nabla f(\overline{x})(y - \overline{x}) + \frac{\gamma}{2} \| y - \overline{x} \|_2^2 \ge 0 , \quad \text{για κάθε } y \in U .$$

Συμπεραίνουμε όπως πιο πάνω, με το $\theta$-επιχείρημα, ότι

$$\nabla f(\overline{x})(y - \overline{x}) \ge 0 , \quad \text{για κάθε } y \in U .$$

Στην περίπτωση (FW), έστω τυχόν $y \in U$. Από το Βήμα 2(FW), έχουμε

$$\nabla f(x_k)(y - x_k) \ge \delta_k ,$$

και παίρνοντας το όριο όταν $k \to \infty$, $k \in K$, βρίσκουμε

$$\nabla f(\overline{x})(y - \overline{x}) \ge 0 , \quad \text{για κάθε } y \in U .$$

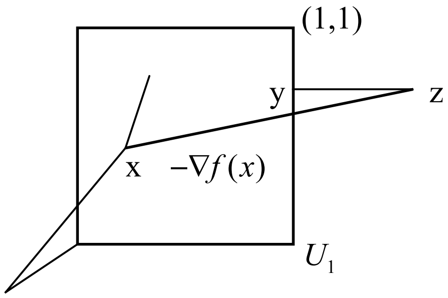
**Σχόλια**

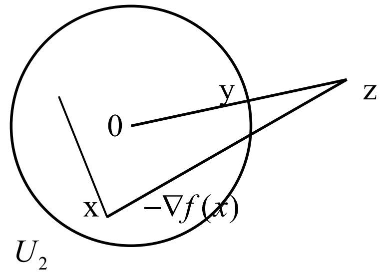
1. Αν η $f$ είναι επιπλέον κυρτή, τότε η παραπάνω μέθοδοι υπολογίζουν σημεία ολικού ελαχίστου, και αν είναι αυστηρά κυρτή, τότε υπολογίζουν το μοναδικό σημείο ελαχίστου, οπότε ολόκληρη η ακολουθία $(x_k)$ συγκλίνει σε αυτό το σημείο.
2. Η μέθοδος προβεβλημένης κλίσης είναι συνήθως πολύ ταχύτερη από τη μέθοδο Frank-Wolfe.
3. Οι παραπάνω μέθοδοι εφαρμόζονται όταν το σύνολο $U$ είναι απλής γεωμετρικής μορφής, π.χ. υπερκύβος, μπάλα κλπ. (βλ. Παραδείγματα).

**Παραδείγματα.**

Έστω τα σύνολα $U_1 := [-1,1]^2$ και $U_2 := \{x \in \mathbb{R}^2 \mid \|x\|_2 \le 1\}$. Εφαρμόζουμε το Βήμα 2 για τις δύο παραπάνω μεθόδους για την εύρεση του σημείου κατεύθυνσης $y := y_k$, και για τα δύο αυτά σύνολα περιορισμών.

<u>Μέθοδος Προβεβλημένης Κλίσης</u>: $y :=$ προβολή του $z := x - \nabla f(x)/\gamma$ στο $U$ :

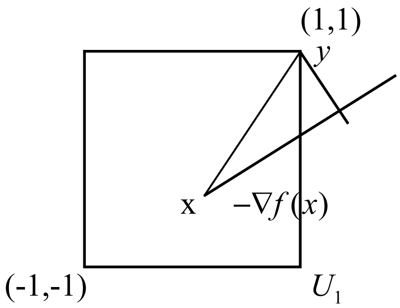


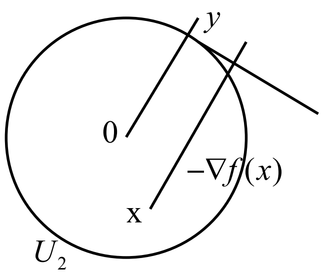




$$y_i := \begin{cases} z_i & \text{αν } |z_i| \le 1 \\ 1 & \text{αν } z_i > 1 \\ -1 & \text{αν } z_i < -1 \end{cases}$$

$$y := \begin{cases} z & \text{αν } \|z\|_2 \le 1 \\ \dfrac{z}{\|z\|_2} & \text{αν } \|z\|_2 > 1 \end{cases}$$

<u>Μέθοδος Frank-Wolfe</u>




$$y_i := \begin{cases} 1 & \text{αν } -\nabla f(x)_i \ge 0 \\ -1 & \text{αν } -\nabla f(x)_i < 0 \end{cases}$$

$$y := \frac{-\nabla f(x)}{\|\nabla f(x)\|_2}$$

---