ΜΕΘΟΔΟΙ ΤΗΣ ΚΛΙΣΗΣ

Έστω $f \in C^1(\mathbb{R}^n)$ μια συνάρτηση. Θεωρούμε το ακόλουθο ***πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς***

(1) Να βρεθεί $\bar{x} \in \mathbb{R}^n$ τέτοιο ώστε $f(\bar{x}) = \min_{x \in \mathbb{R}^n} f(x)$.

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

(2) $\nabla f(\bar{x}) = 0$.

Έστω $b, c \in (0,1)$ (συνήθως $b = c = 0.5$), $d > 0$, και (s_k) μία ακολουθία με $s_k \in (0, d]$. Η ***μέθοδος της κλίσης, με βέλτιστο βήμα, ή με βήμα Armijo***, είναι μια επαναληπτική μέθοδος καθόδου η οποία βρίσκει στο όριο σημεία $\bar{x}$ που ικανοποιούν τη συνθήκη (2). Εκμεταλλεύεται σε κάθε επανάληψη την κατεύθυνση καθόδου $-\nabla f$, που δίνει τοπικά τη μεγαλύτερη μείωση, και περιγράφεται από τον ακόλουθο αλγόριθμο.

Αλγόριθμος 1

Βήμα 1. Θέτουμε $k := 0$ και επιλέγουμε ένα αρχικό $x_0 \in \mathbb{R}^n$.

Βήμα 2. Υπολογίζουμε το $\nabla f(x_k)$ και θέτουμε $\delta_k := -\|\nabla f(x_k)\|_2^2$.

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

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

(α) ***Βέλτιστο Βήμα***: Βρίσκουμε ένα $\alpha_k \geq 0$ τέτοιο ώστε

(3) $f(x_k - \alpha_k \nabla f(x_k)) - f(x_k) = \min_{\alpha \geq 0} [f(x_k - \alpha \nabla f(x_k)) - f(x_k)]$.

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

(4) $f(x_k - \alpha \nabla f(x_k)) - f(x_k) \leq \alpha b \delta_k$,

θέτουμε επαναληπτικά $\alpha := c\alpha$ (π.χ. $c = 0.5$) 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*** α που την ικανοποιεί.

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

Θεώρημα 1. Υποθέτουμε ότι το κλειστό σύνολο

$$S := \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$$

είναι και φραγμένο, άρα 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 1 είναι άπειρη, κάθε όριο $\bar{x}$ συγκλίνουσας υπακολουθίας της (x_k) ικανοποιεί τη συνθήκη (2), και $\delta_k \to 0$ για ολόκληρη την ακολουθία.

Απόδειξη. Θα δείξουμε πρώτα επαγωγικά ότι, εφόσον $\delta_k \neq 0$, η ακολουθία (x_k) είναι καλά ορισμένη (δηλ. υπάρχει x_k για κάθε k), 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 $\delta_k \neq 0$, δηλ. $\nabla f(x_k) \neq 0$.

Εύκολα βλέπουμε ότι η ελαχιστοποίηση (3) ισοδυναμεί με την ελαχιστοποίηση της f στο συμπαγές σύνολο

$$\{x := x_k - \alpha\nabla f(x_k) \big| \alpha \geq 0\} \cap S ,$$

και άρα υπάρχει α_k για κάθε k.

Εξάλλου, από τη ιδιότητα της παραγώγου θετικά κατά κατεύθυνση, έχουμε

$$\lim_{\alpha\to 0^+} \frac{f(x_k - \alpha\nabla f(x_k)) - f(x_k)}{\alpha} = \nabla f(x_k)^T[-\nabla f(x_k)] = -\|\nabla f(x_k)\|_2^2 = \delta_k < 0 ,$$

άρα

$$\frac{f(x_k - \alpha\nabla f(x_k)) - f(x_k)}{\alpha} \leq b\delta_k , \quad (\text{αφού } b \in (0,1) , \text{ π.χ. } b = 0.5),$$

για α αρκετά μικρό. Αυτό δείχνει ότι, αν η ανισότητα (4) δεν ικανοποιείται αρχικά, τότε ικανοποιείται μετά από ένα πεπερασμένο αριθμό επαναλήψεων $\alpha := c\alpha$. Επίσης, αν η ανισότητα αυτή ικανοποιείται αρχικά, επειδή το σύνολο S είναι φραγμένο, εύκολα βλέπουμε ότι ικανοποιείται μόνο για ένα πεπερασμένο αριθμό επαναλήψεων $\alpha := \alpha / c$.

Επομένως, το βήμα α_k και το x_{k+1} υπάρχουν και στις δύο περιπτώσεις 4α και 4β. Επιπλέον, τα βήματα 4α, 4β και 5 δείχνουν ότι και στις δύο περιπτώσεις ισχύει $f(x_{k+1}) - f(x_k) \leq 0$, και άρα $x_{k+1} \in S$. Συνεπώς, $x_k \in S$ για κάθε k, και η ακολουθία $(f(x_k))$ είναι φθίνουσα.

Τώρα, αν ο αλγόριθμος σταματά για κάποιο k, τότε $\delta_k = 0$ και $\nabla f(x_k) = 0$. Αλλιώς, η ακολουθία (x_k) είναι άπειρη. Ας δείξουμε ότι τότε $\delta_k \to \delta := 0$. Επειδή το S είναι συμπαγές και $x_k \in S$ για κάθε k, υπάρχει υπακολουθία $(x_k)_{k\in K}$ που συγκλίνει σε κάποιο $\overline{x} \in S$. Επειδή $f \in C^1(\mathbb{R}^n)$, έχουμε

$$\delta_k := -\|\nabla f(x_k)\|_2^2 \to \delta := -\|\nabla f(\overline{x})\|_2^2 \leq 0 , \quad \text{όταν } k \to \infty ,\ k \in K .$$

Ας υποθέσουμε ότι $\delta < 0$. Εύκολα βλέπουμε ότι η παράγωγος της συνάρτησης $\phi(\alpha) := f(x_k - \alpha\nabla f(x_k))$ είναι $\phi'(\alpha) = -\nabla f(x_k - \alpha\nabla f(x_k))^T \nabla f(x_k)$. Από το Θεώρημα Μέσης Τιμής και αφού $b \in (0,1)$, έχουμε τότε

$$\begin{aligned}(5) \qquad & f(x_k - \alpha\nabla f(x_k)) - f(x_k) = \phi(\alpha) - \phi(0) = -\alpha\nabla f(x_k - \mu_\alpha\nabla f(x_k))^T \nabla f(x_k) \\ & = \alpha[-\nabla f(\overline{x})^T \nabla f(\overline{x}) + \varepsilon_{k\alpha}] = \alpha(\delta + \varepsilon_{k\alpha}) \leq \alpha b\delta \quad \textbf{\textit{και}} \quad \leq \alpha b\delta_k ,\end{aligned}$$

για $\alpha \in [0,\alpha']$ και $k \geq k'$, για κάποια α', k', όπου $\mu_\alpha \in (0,\alpha)$, και $\varepsilon_{k\alpha} \to 0$ όταν $k \to \infty$, $k \in K$, και $\alpha \to 0^+$.

(α) Περίπτωση Βέλτιστου Βήματος: Έχουμε εδώ

$$\begin{aligned} f(x_{k+1}) - f(x_k) &= \min_{\alpha\geq 0}[f(x_k - \alpha\nabla f(x_k)) - f(x_k)] \\ &\leq \min_{\alpha\in[0,\alpha']}[f(x_k - \alpha\nabla f(x_k)) - f(x_k)] \leq \min_{\alpha\in[0,\alpha']} \alpha b\delta = \alpha' b\delta , \quad \text{για } k \in K ,\ k \geq k' .\end{aligned}$$

Επειδή ολόκληρη η ακολουθία $(f(x_k))$ είναι φθίνουσα και $\delta < 0$, συμπεραίνουμε ότι $f(x_k) \to -\infty$, που είναι άτοπο αφού $f(x_k) \to f(\overline{x})$, $k \in K$. Συνεπώς, $\delta = 0$.

(β) Περίπτωση Βήματος Armijo: Από την κατασκευή του βήματος Armijo και την πιο πάνω $2^η$ ανισότητα (5), εύκολα βλέπουμε ότι έχουμε αναγκαστικά $\alpha_k \geq c\alpha'$, και άρα

$$f(x_{k+1}) - f(x_k) = f(x_k - \alpha_k\nabla f(x_k)) - f(x_k) \leq \alpha_k b\delta_k \leq c\alpha' b\delta_k \leq c\alpha' b\delta / 2,$$

για $k \in K$, $k \geq k'$. Καταλήγουμε έτσι σε άτοπο, όπως στο (α). Συνεπώς, $\delta = 0$.

Τέλος, και στις δύο περιπτώσεις (α) και (β), από τη μοναδικότητα του ορίου 0 προκύπτει ότι $\delta_k \to 0$ για ολόκληρη η ακολουθία.

Σχόλια

1. Η επιλογή βήματος Armijo είναι συνήθως ταχύτερη από την επιλογή βέλτιστου βήματος, η οποία προυποθέτει μια άπειρη διαδικασία, π.χ. τη μέθοδο της Χρυσής Τομής, για την εύρεσή του. Με την επιλογή του βέλτιστου βήματος, όμως, υπάρχει μεγαλύτερη πιθανότητα η μέθοδος να συγκλίνει σε ένα απόλυτο ελάχιστο.
2. Αν η f είναι επιπλέον κυρτή, τότε η μέθοδος της κλίσης υπολογίζει σημεία ολικού ελαχίστου, και αν είναι αυστηρά κυρτή, τότε υπολογίζει το μοναδικό σημείο ελαχίστου, οπότε ολόκληρη η ακολουθία (x_k) συγκλίνει σε αυτό το σημείο.
3. Μπορούμε να επιλέξουμε ένα σταθερό αρχικό βήμα Armijo $s_k := s$ για κάθε k. Μια πιο αποτελεσματική, προσαρμοστική, επιλογή (πολύ χρήσιμη κυρίως στις μικτές μεθόδους ποινών) είναι να θέσουμε $s_0 := 1$, και μετά $s_k := \alpha_{k-1}$, για $k \geq 1$.

Παράδειγμα. Έστω η τετραγωνική συνάρτηση

$$f(x) = \frac{1}{2} x^T Ax - b^T x,$$

όπου ο πίνακας A είναι συμμετρικός ($A^T = A$) και θετικά ορισμένος ($x^T Ax > 0$ για κάθε $x \neq 0$). Εδώ, μπορούμε να υπολογίσουμε το βέλτιστο βήμα α_k ακριβώς, αν $\delta_k \neq 0$. Πράγματι, έχουμε $\nabla f(x) = Ax - b$, και θέτοντας

$$g_k := \nabla f(x_k) = Ax_k - b,$$

υπολογίζουμε

$$P(\alpha) := f(x_k - \alpha g_k) - f(x_k) = \frac{1}{2}(x_k - \alpha g_k)^T A(x_k - \alpha g_k) - b^T(x_k - \alpha g_k)$$

$$-\frac{1}{2} x_k{}^T Ax_k + b^T x_k$$

$$= \frac{1}{2}\alpha^2 g_k{}^T Ag_k - \alpha x_k{}^T A^T g_k + \alpha b^T g_k$$

$$= \frac{1}{2}\alpha^2 g_k{}^T Ag_k - \alpha(Ax_k - b)^T g_k = \frac{1}{2}\alpha^2 g_k{}^T Ag_k - \alpha g_k{}^T g_k.$$

Αν $\delta_k \neq 0$, δηλ. $g_k \neq 0$, τότε $g_k{}^T Ag_k > 0$, αφού ο A είναι θετικά ορισμένος, και άρα το τετραγωνικό πολυώνυμο $P(\alpha)$ ελαχιστοποιείται, για $\alpha \in \mathbb{R}$, αν

$$P'(\alpha) = \alpha g_k{}^T Ag_k - g_k{}^T g_k = 0,$$

δηλ. για

$$\alpha = \alpha_k := \frac{g_k{}^T g_k}{g_k{}^T Ag_k} > 0,$$

άρα και για $\alpha \geq 0$, που δίνει και το ζητούμενο βέλτιστο βήμα.
