Αριθμητικές Μέθοδοι Βελτιστοποίησης

Μέθοδος αναζήτησης γραμμής. Κριτήριο Wolfe. Μέθοδος Απότομης Καθόδου

Αναζήτηση γραμμής (άλλη μέθοδος)

Σκοπός είναι να βρούμε a ώστε $\min_{a>0} F(a) = \min_{a>0} f(x_k + ap_k)$

Με το κριτήριο Armijo βρίσκουμε a_k ώστε $f(x_k + a_k p_k) < f(x_k)$

Αντί για αυτό θα μπορούσαμε να αναζητήσουμε a_k ώστε $F'(a_k) \approx 0$
ή διαφορετικά

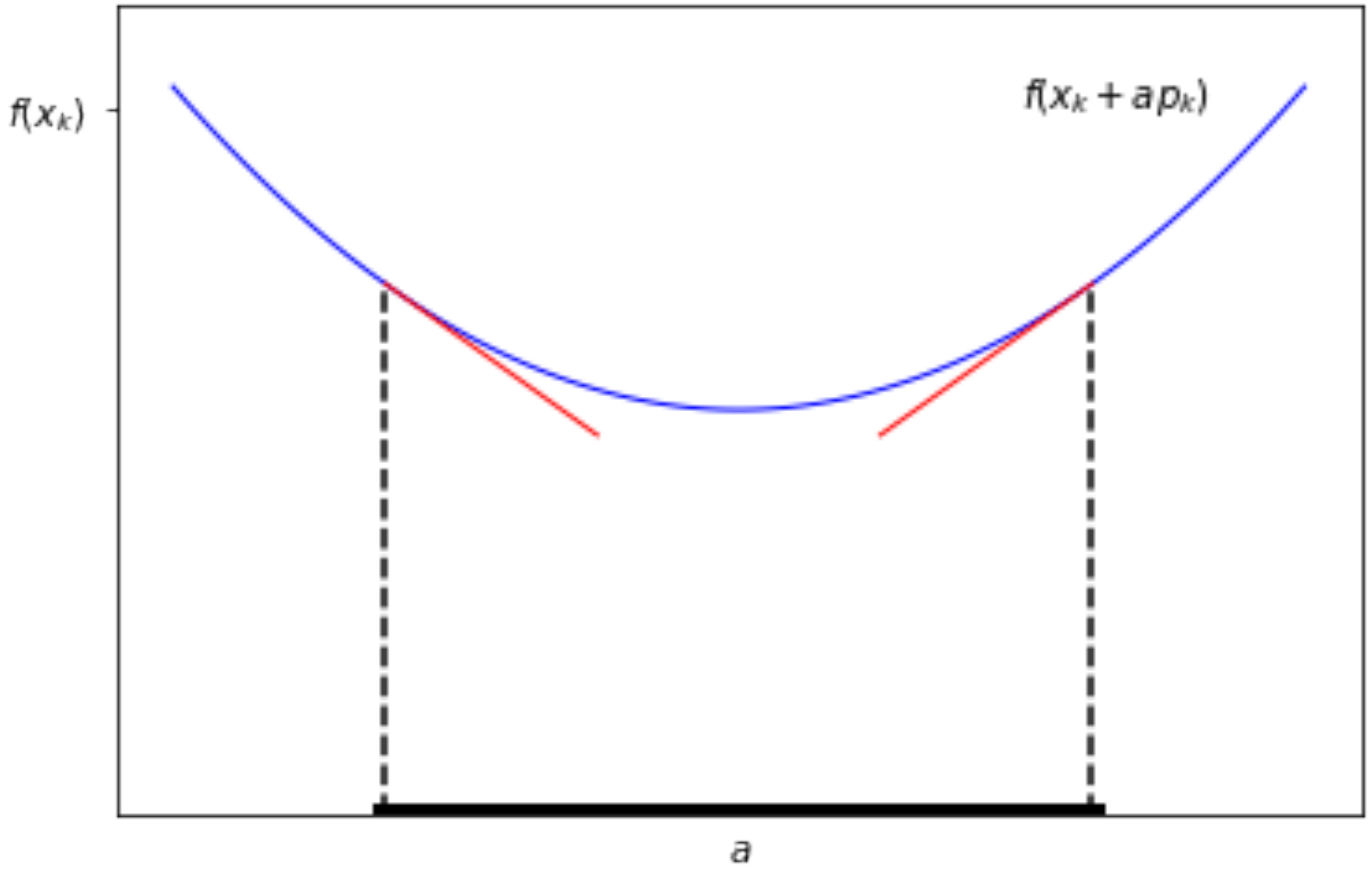
$$|F'(a_k)| \leq \eta |F'(0)|, 0 \leq \eta < 1.$$

Συνθήκες Wolfe

$$F(a) = f(x_k + ap_k)$$

$$F'(a) = p_k^T \nabla f(x_k + ap_k)$$

$$|F'(a_k)| \leq \eta |F'(0)|$$



$$|p^T \nabla f(x_k + ap)| \leq \eta |p^T \nabla f(x_k)|$$

Αν $\eta = 0$, τότε το a_k θα δίνει το σημείο που ελαχιστοποιείται η F.

Επιλέγοντας $\eta > 0$ θα βρούμε a_k που $|F'(a_k)| \approx 0$

Αυτή η επιλογή a_k μπορεί να οδηγήσει σε τοπικό μέγιστο και όχι ελάχιστο. Για αυτό η μέθοδος θα πρέπει να δίνει a_k που να ικανοποιεί και μια συνθήκη καθόδου Armijo,

$$f(x_k + a_k p_k) \leq f(x_k) + \mu a_k p_k^T \nabla f(x_k), \quad 0 < \mu < \eta$$

Κατασκευάζουμε a_k που ανήκει σε ένα διάστημα $[\underline{a}, \overline{a}]$, που να περιέχει κατά προσέγγιση το ελάχιστο της $F(a) = f(x_k + ap_k)$.

Αφού η F' είναι συνεχής θέλουμε να βρούμε ένα διάστημα με $F'(\underline{a}) < 0$ και $F'(\overline{a}) > 0$ οπότε θα υπάρχει σημείο $a \in [\underline{a}, \overline{a}]$ με $F'(a) = 0$.

Επειδή p_k είναι κατεύθυνση καθόδου ισχύει $F'(0) = p_k^T \nabla f(x_k) < 0$, οπότε $0 < \underline{a}$

Μπορούμε να δοκιμάσουμε τιμές a, π.χ. $a = 1,2,4,\ldots$ και επιλέγουμε $\underline{a}$ τη μεγαλύτερη τιμή a που $F'(a) < 0$

Αν κατα τη διάρκεια την διαδικασίας κιβωτισμού βρούμε a που ικανοποιεί τη συνθήκη Wolfe, τερματίζουμε τη διαδικασία και θέτουμε αυτό το a ως a_k

Αν έχουμε βρει ένα διάστημα $[\underline{a}, \overline{a}]$ προσεγγίζουμε το a το οποίο ελαχιστοποιεί την F με το σημείο a_k το οποίο ελαχιστοποιεί μια κατάλληλη πολυωνυμική παρεμβολή της F.

Θεωρούμε το πολυώνυμο παρεμβολής $p_3 \in \mathbb{P}_3$ τέτοιο ώστε

$$p_3(\underline{a}) = F(\underline{a}), \quad p_3(\overline{a}) = F(\overline{a})$$

$$p_3'(\underline{a}) = F'(\underline{a}), \quad p_3'(\overline{a}) = F'(\overline{a})$$

Υπολογίζουμε το μοναδικό ελάχιστο της p_3 στο $[\underline{a}, \overline{a}]$ και θέτουμε αυτό ως a_k

Αν το a_k ικανοποιεί τη συνθήκη του Wolfe, υπολογίζουμε την επόμενη προσέγγιση $x_{k+1} = x_k + a_k p_k$, διαφορετικά αντικαθιστούμε το $\underline{a}$ ή το $\overline{a}$ με το a_k, ανάλογα με την τιμή $F'(a_k)$ και επαναλαμβάνουμε τη διαδικασία

Παράδειγμα

$$F(a) = 5 - a - \log(4.5 - a), \quad \eta = 0.1$$

$a = 0, \quad F(0) = 3.4959, \quad F'(0) = -0.7778 < 0$

Συνθήκη Wolfe: $|F'(a_k)| < 0.07778$

$a = 1, \quad F(1) = 2.7472, \quad F'(1) = -0.7143$

$a = 2, \quad F(2) = 2.0837, \quad F'(2) = -0.6000$

$a = 4, \quad F(4) = 1.6931, \quad F'(4) = 1.$

Άρα $[\underline{a}, \overline{a}] = [2,4]$

Κανένα από τα a που βρήκαμε δεν ικανοποιεί τη συνθήκη του Wolfe, οπότε θεωρούμε το $p_3 \in \mathbb{P}_3$, ώστε να παρεμβάλει τις F, F' στα $2{,}4$.

Μπορούμε να βρούμε ότι το p_3 θα έχει ελάχιστο στο $a = 3.3826$ και

$$F(a) = 1.5064 \quad \text{και} \quad F'(a) = -0.1050$$

Αφού $F'(a) < 0$ το νέο διάστημα που θα βρίσκεται το ελάχιστο είναι $[3.3826{,}4]$. Το νέο $p_3 \in \mathbb{R}^3$ που παρεμβάλλει τις F, F' στο $[3.3826{,}4]$ έχει ελάχιστο στο $a = 3.5294$ και

$$F(a) = 1.5004 \quad \text{και} \quad F'(a) = 0.0303$$

Αυτό το σημείο ικανοποιεί τη συνθήκη Wolfe και θέτουμε $a_k = 3.5294$

Μέθοδος απότομης καθόδου

Μέθοδος απότομης καθόδου (steepest descent)

Στο γενικό αλγόριθμο βελτιστοποίησης, επιλέγοντας ως κατεύθυνση

$$p_k = -\nabla f(x_k)$$

η οποία είναι μια κατεύθυνση καθόδου γιατί

$$p_k^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0$$

Στη συνέχεια χρησιμοποιούμε μια μέθοδο αναζήτησης γραμμής για να προσδιορίσουμε το a_k ώστε

$$f(x_{k+1}) = f(x_k + a_k p_k) < f(x_k)$$

Μέθοδος απότομης καθόδου (steepest descent)

Σε προηγούμενο θεώρημα για τη σύγκλιση είχαμε υποθέσει ότι η κατεύθυνση καθόδου p_k θέλουμε να ικανοποιεί

$$-\frac{p_k^T \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} \geq \epsilon > 0, \quad \epsilon > 0 \text{ και } \|p_k\| \geq m\|\nabla f(x_k)\|, \quad m > 0$$

Αν $p_k = -\nabla f(x_k)$ τότε

$$-\frac{p_k^T \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} = 1 > 0 \text{ και } \|p_k\| = \|\nabla f(x_k)\|$$

Στη συνέχεια χρησιμοποιούμε μια μέθοδο αναζήτησης γραμμής για να προσδιορίσουμε το a_k ώστε

$$f(x_{k+1}) = f(x_k + a_k p_k) < f(x_k)$$

Μέθοδος απότομης καθόδου (steepest descent)

Στην περίπτωση που η $f(x) = \frac{1}{2}x^T A x - b^T x$, είναι δηλαδή έχουμε ένα τετραγωνικό πρόβλημα, μπορούμε να υπολογίσουμε ακριβώς το βήμα a_k που ελαχιστοποιεί την $f(x_k + a_k p_k)$, αν $p_k = -\nabla f(x_k)$.

Λήμμα: Έστω $A \in \mathbb{R}^{n\times n}$, συμμετρικός, $b \in \mathbb{R}^n$, και

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

Αν $x_k \in \mathbb{R}^n$, $p_k = -\nabla f(x_k)$ και $a > 0$, η $\phi(a) = f(x_k + ap_k)$,

λαμβάνει την ελάχιστη τιμή της για $a_k = \dfrac{\|p_k\|^2}{p_k^T A p_k}$

Απόδειξη

Απόδειξη: $\phi(a) = \frac{1}{2}(x_k + ap_k)A(x_k + ap_k) - b^T(x_k + ap_k)$

$$= \frac{1}{2}x_k^T A x_k - b^T x_k + \frac{1}{2}a^2 p_k^T A p_k - ab^T p_k - ap_k^T A x_k$$

$$= f(x_k) + \frac{1}{2}a^2 p_k^T A p_k + ap_k^T (Ax_k - b)$$

$$= f(x_k) + \frac{1}{2}a^2 p_k^T A p_k + ap_k^T p_k$$

Απόδειξη (συνέχεια)

Άρα το ϕ είναι ένα τριώνυμο 2ου βαθμού ως προς a. Εύκολα βλέπουμε ότι η ελάχιστη τιμή του ϕ λαμβάνεται για $a = \dfrac{p_k^T p_k}{p_k^T A p_k}$

Παράδειγμα

Θέλουμε να ελαχιστοποιήσουμε την $f(x) = \frac{1}{2}x^T A x - b^T x$, με

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 25 \end{pmatrix} \quad \text{και} \quad b = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

Το ελάχιστο είναι $x^\star = (-1, -\frac{1}{5}, \frac{1}{25})^T$. Αν εφαρμόσουμε τη μέθοδο απότομης καθόδου και θέσουμε ως κριτήριο τερματισμού όταν $||\nabla f(x_k)|| \leq 10^{-8}$, θα χρειαστούμε 216 επαναλήψεις.

Παράδειγμα

k	x_k	$f(x_k)$	$\|\nabla f(x_k)\|$
0	[0. 0. 0.]	0.0	1.7321
1	[-0.0968 -0.0968 -0.0968]	-0.1452	1.7598
2	[-0.1500 -0.1272 -0.0131]	-0.2365	1.1437
3	[-0.2375 -0.1647 -0.0823]	-0.3038	1.3163
….			
216	[-1 -0.2 -0.04]	-0.62	9.0092×10^{-09}

Ταχύτητα σύγκλισης

Λήμμα: Έστω $\{x_k\}$ μια ακολουθία προσεγγίσεων του ελάχιστου της $f(x) = \frac{1}{2}x^T Ax - b^T x, A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n$ με τη μέθοδο απότομης καθόδου, όπου έχουμε υπολογίσει ακριβώς το βέλτιστο βήμα a_k σε κάθε επαναληπτικό βήμα της μεθόδου. Τότε για κάθε αρχική προσέγγιση $x_0 \in \mathbb{R}^n$, η μέθοδος συγκλίνει γραμμικά με

$$\frac{f(x_{k+1}) - f(x^\star)}{f(x_k) - f(x^\star)} \leq \left(\frac{\kappa(A) - 1}{\kappa(A) + 1}\right)^2$$

όπου $\kappa(A)$ είναι ο δείκτης κατάστασης του A.

Παράδειγμα (προσέγγιση τάξεως σύγκλισης προηγούμενου παραδείγματος)

k	$f(x_k)$	$\dfrac{f(x_k) - f(x^\star)}{f(x_k) - f(x^\star)}$	p
0	0.0		
1	-0.1452	0.7659	
2	-0.2365	0.8077	0.8006
3	-0.3038	0.8246	0.9033
4	-0.3560	0.8348	0.9358
5	-0.3988	0.8379	0.9798