# Αριθμητικές Μέθοδοι Βελτιστοποίησης

**Μέθοδος αναζήτησης γραμμής. Κριτήριο Armijo. Σύγκλιση.**

## Παράδειγμα

Θεωρούμε τη συνάρτηση

$$f(x_1, x_2) = 5x_1^2 + 7x_2^2 - 3x_1x_2$$

Έστω $x_k = (2,3)^T, p_k = (5, -7)^T$. Τότε $f(x_k) = 65$ και επειδή

$$\nabla f(x_k) = (10x_1 - 3x_2, 14x_2 - 3x_1),$$

$\nabla f(x_k) = (11,36)$. Οπότε $p_k^T \nabla f(x_k) < 0$, δηλαδή $p_k$ κατεύθυνση καθόδου.

Αν θέσουμε $x_{k+1} = x_k + p_k$, τότε $f(x_{k+1}) = 121 > f(x_k) = 65$.

Όμως αν $x_{k+1} = x_k + \dfrac{1}{2}p_k$, τότε $f(x_{k+1}) = \dfrac{9}{4} < f(x_k) = 65$

# Γενικός Αλγόριθμος

1. Θεωρούμε μια αρχική προσέγγιση $x_0 \in \mathbb{R}^n$ και ένα $\epsilon > 0$
2. Για $k = 0,1,2,\ldots$
   (i) Αν το $\|\nabla f(x_k)\| < \epsilon$ έχουμε βρει "σχεδόν" το ελάχιστο και σταματάμε
   (ii) Βρείτε μια κατεύθυνση καθόδου $p_k$ (Μέθοδος Νεύτωνα ή κάποια άλλη)
   (iii) Βρείτε ένα βήμα $a_k$, ώστε $x_{k+1} = x_k + a_k p_k$ να είναι καλύτερη προσέγγιση από το $x_k$

# Μέθοδοι Αναζήτησης Γραμμής

## Κριτήρια καθόδου

Οι μέθοδοι αναζήτησης γραμμής χρησιμοποιούνται συχνά για την προσέγγιση τοπικού ελάχιστου της $f(x_k + ap_k)$ ως προς $a > 0$.

Αν υποθέσουμε ότι γνωρίζουμε μια κατεύθυνση καθόδου $p_k$ από το $x_k$,$(p^T \nabla f(x_k) < 0)$, η νέα προσέγγιση θα δίνεται ως

$$x_{k+1} = x_k + a_k p_k$$

Όπου $a_k > 0$, τέτοιο ώστε

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

Η μέθοδος υπολογισμού ενός $a_k > 0$ που να αποτελεί προσέγγιση του ελαχίστου $a$, καλείται ***μέθοδος αναζήτησης γραμμής*** ή ***γραμμική αναζήτηση***.

## Παράδειγμα καθόδου

Κατασκευάζοντας προσεγγίσεις $x_k$, χρησιμοποιώντας μια κατεύθυνση καθόδου $p_k$, δεν είναι σίγουρο ότι θα συγκλινούμε σε ένα τοπικό ελάχιστο της $f$.

Έστω $f(x) = x^2$ και $x_0 = -3$. Επίσης αν έχουμε την κατεύθυνση καθόδου $p_k = 1,\ k = 0,1,2,\ldots,$ και θέσουμε το βήμα $a_k = \frac{1}{2^k}$, τότε $x_{k+1} = x_k + \frac{1}{2^k}$

Παίρνουμε $x_0 = -3, x_1 = -2, x_2 = -3\frac{3}{2}, x_4 = -\frac{5}{4}, \ldots,\ x_k = -(1 + 2^{1-k})$

## Παράδειγμα καθόδου

Η κατεύθυνση $p_k$ είναι κατεύθυνση καθόδου γιατί

$$p_k^T \nabla f(x_k) = 1 \cdot 2x_k = -2(1+2^{1-k}) < 0 \text{ και } f(x_{k+1}) < f(x_k)$$

Πράγματι, έχουμε

$$x_{k+1}^2 = (x_k + 2^{-k})^2 < x_k^2$$

$$\text{ή } x_k^2 + 2\frac{1}{2^k}x_k + \frac{1}{2^{2k}} < x_k^2$$

$$\text{ή } -\frac{1}{2^{k-1}}(1+2^{1-k}) + \frac{1}{2^{2k}} < 0, \text{ το οποίο είναι αληθές}$$

**Παράδειγμα καθόδου**

Παρατηρούμε ότι $x_k \to -1$ και $f'(-1) = -2 \neq 0$. Το $x^\star = 0$ είναι το ελάχιστο.

Καταλαβαίνουμε ότι χρειάζονται κάποιες επιπλέον συνθήκες για να εξασφαλίσουν τη σύγκλιση σε ένα ελάχιστο

## “Ικανοποιητική” κάθοδος

Αν το $p_k$ είναι καθοδική κατεύθυνση $p_k^T \nabla f(x_k) < 0$, αλλά είναι “σχεδόν” κάθετο στο $\nabla f(x_k)$, θα μειώνει “ελάχιστα” την τιμή της $f$.

Για να το αποφύγουμε αυτό υποθέτουμε ότι

$$-\frac{p_k^T \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} \geq \epsilon > 0, \quad k > k_0$$

Συνθήκη γωνίας: $\cos(\theta_k) \geq \epsilon > 0$

Αν $p_k \perp \nabla f(x_k)$ τότε $\cos(\theta_k) = 0$

**Συνθήκη κλίσης**

Μια ακόμα συνθήκη είναι

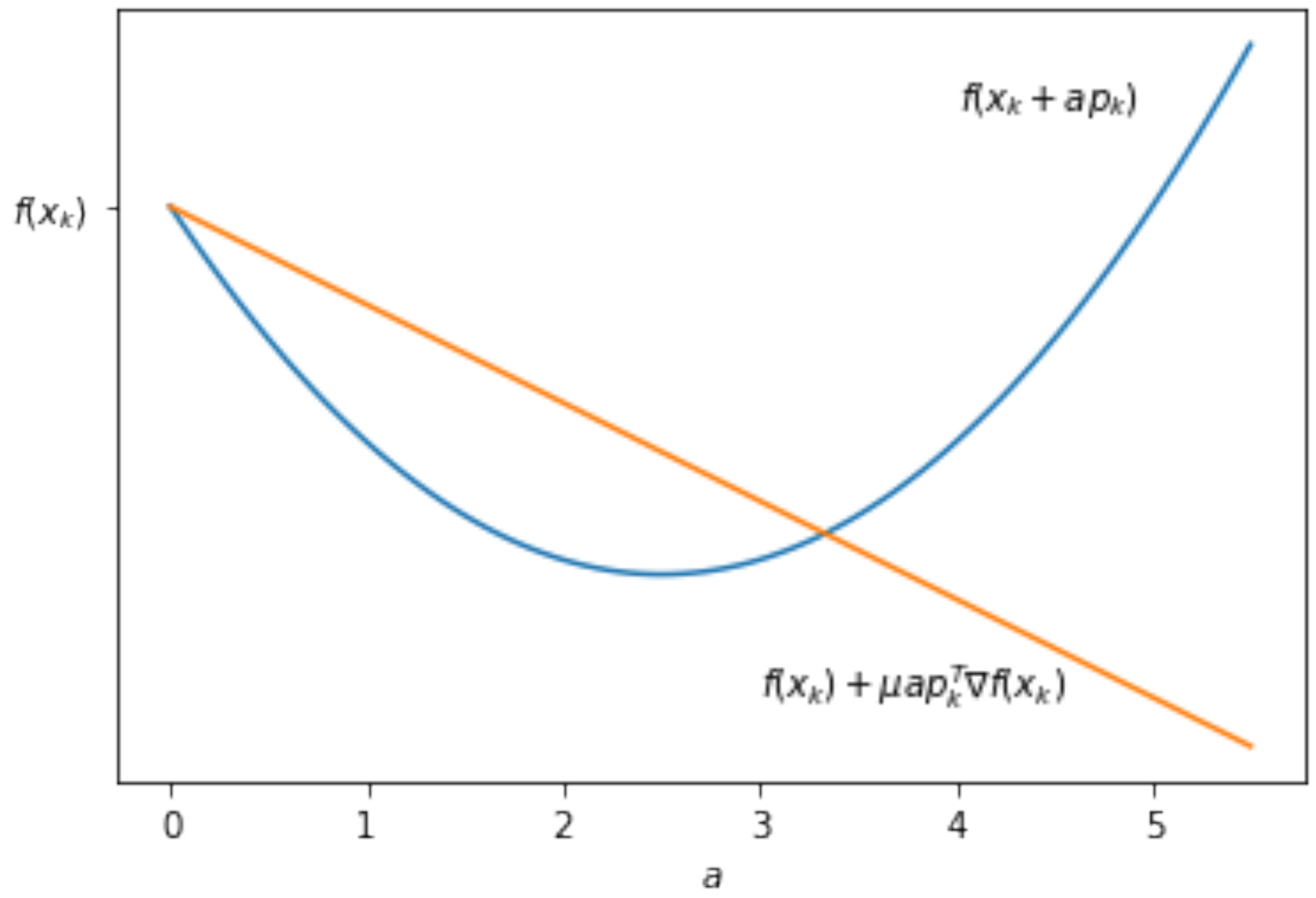
$$\|p_k\| \geq m\|\nabla f(x_k)\|, \quad m > 0$$

Δεν μπορεί η κατεύθυνση αναζήτησης $p_k$ να γίνει πολύ μικρή σε "μήκος" σε σχέση με τη $\nabla f(x_k)$

## Συνθήκη Armijo

$$f(x_k + a_k p_k) \leq f(x_k) + \mu a_k p_k^T \nabla f(x_k), \quad 0 < \mu < 1$$

Το $a_k$ πρέπει να τέτοιο ώστε να έχουμε μια ικανοποιητική μείωση της τιμής της $f$ στην κατεύθυνση $p_k$



# Σύγκλιση μεθόδου αναζήτησης γραμμής

***Θεώρημα:*** Έστω $f : \mathbb{R}^n \to \mathbb{R}$, $x_0 \in \mathbb{R}^n$ αρχική προσέγγιση και $\{x_k\}$ ακολουθία προσεγγίσεων που ορίζεται ως $x_{k+1} = x_k + a_k p_k$, όπου $p_k$ κατεύθυνση καθόδου και $a_k \geq 0$. Υποθέτουμε ότι

(i) Το $S = \{x : f(x) \leq f(x_0)\}$ είναι φραγμένο

(ii) $\nabla f$ είναι Lipschitz

(iii) $-\dfrac{p_K^T \nabla f(x)}{\|\nabla f(x)\| \|p_k\|} \geq \epsilon > 0$, για $\epsilon > 0$

(iv)$\|p_k\| \geq m\|\nabla f(x_k)\|$ και $\|p_k\| \leq M, \quad k > 0$

(v) $a_k \in \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots\}$, το μεγαλύτερο που ικανοποιεί την

$$f(x_k + a_k p_k) < f(x_k) + \mu a_k p_k^T \nabla f(x_k), \quad 0 < \mu < 1$$

Τότε $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$

***Απόδειξη:*** Τα βήματα στην απόδειξη είναι τα ακόλουθα: Δείχνουμε ότι η $f$ είναι κάτω φραγμένη στο $S$. Μετά θα δείξουμε ότι υπάρχει το $\lim_{k\to\infty} f(x_k)$ και ότι $\lim_{k\to\infty} a_k\|\nabla f(x_k)\|^2 = 0$. Αν $a_k < 1$ τότε υπάρχει σταθερά $\gamma > 0$ τέτοια ώστε $a_k \geq \gamma\|\nabla f(x_k)\|^2$. Τελικά δείχνουμε το ζητούμενο ότι $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$

**Απόδειξη (συνέχεια)**

Επειδή η $f$ είναι συνεχής το σύνολο $S$ είναι κλειστό και επειδή είναι φραγμένο η $f$ θα λαμβάνει ελάχιστη τιμή σε ένα σημείο $x^\star$. Άρα υπάρχει σταθερά $C$ τ.ω. $C \leq f(x), \quad \forall x \in S$.

Θα δούμε ότι το $\lim_{k\to\infty} f(x_k)$ υπάρχει. Η συνθήκη ικανοποιητικής καθόδου εξασφαλίζει ότι $f(x_{k+1}) < f(x_k) \leq f(x_0)$, διότι

$$f(x_{k+1}) = f(x_k + a_k p_k) \leq f(x_k) + \mu a_k p_k^T \nabla f(x_k)$$

**Απόδειξη (συνέχεια)**

Η $\{f(x_k)\}$ είναι μονότονα φθίνουσα και κάτω φραγμένη, άρα συγκλίνει σε κάποιο αριθμό $\rho$.

## Απόδειξη (συνέχεια)

$$f(x_0) - f(x_{j+1}) = (f(x_0) - f(x_1)) + (f(x_1) - f(x_2)) + \ldots + (f(x_j) - f(x_{j+1}))$$

$$= \sum_{j=0}^{j} f(x_k) - f(x_{k+1}) \geq -\mu \sum_{j=0}^{j} a_k p_k^T \nabla f(x_k)$$

Αν

$$\geq \mu\epsilon \sum_{j=0}^{j} a_k \|p_k\| \|\nabla f(x_k)\| \geq m\mu\epsilon \sum_{j=0}^{j} a_k \|\nabla f(x_k)\|^2$$

τώρα $j \to \infty$ παίρνουμε $f(x_0) - \rho \geq m\mu\epsilon \sum_{j=0}^{\infty} a_k \|\nabla f(x_k)\|^2$

**Απόδειξη (συνέχεια)**

Συνεπώς $\lim_{k\to\infty} a_k \|\nabla f(x_k)\|^2 = 0$.

Δείχνουμε τώρα ότι αν $a_k < 1$ τότε υπάρχει σταθερά $\gamma > 0$ τέτοια ώστε $a_k \geq \gamma \|\nabla f(x_k)\|^2$.

Αν $a_k < 1$, τότε η $f(x_{k+1}) = f(x_k + ap_k) \leq f(x_k) + \mu a p_k^T \nabla f(x_k)$ δεν ισχύει για $a = 2a_k$, οπότε

$$f(x_{k+1}) = f(x_k + 2a_k p_k) \geq f(x_k) + \mu 2 a_k p_k^T \nabla f(x_k)$$

## Απόδειξη (συνέχεια)

Επειδή η $\nabla f$ είναι Lipschitz συνεχής θα έχουμε ότι

$$|f(x + 2a_k pk) - f(x_k) - 2a_k p_k^T \nabla f(x_k)| \leq \frac{L}{2} \|2a_k p_k\|^2$$

Οπότε

$$f(x_k) - f(x_k + 2a_k p_k) \geq -\frac{L}{2} \|2a_k p_k\|^2 - 2a_k p_k^T \nabla f(x_k)$$

**Απόδειξη (συνέχεια)**

Προσθέτωντας τις δύο προηγούμενες ανισότητες παίρνουμε

$$\begin{aligned} a_k L\|p_k\|^2 &\geq -(1-\mu)p_k^T \nabla f(x_k) \geq (1-\mu)\epsilon\|p_k\|\|\nabla f(x_k)\| \\ &\geq (1-\mu)\epsilon m\|\nabla f(x_k)\|^2 \end{aligned}$$

Και επειδή $\|p_k\| \leq M$ έχουμε

$$a_k \geq \gamma\|\nabla f(x_k)\|^2$$

με $\gamma = \dfrac{(1-\mu)\epsilon m}{M^2 L} > 0$

## Απόδειξη (συνέχεια)

Συνεπώς είτε $a_k = 1$ είτε $a_k \geq \gamma \|\nabla f(x_k)\|^2$, άρα

$$a_k \geq \min\{1, \gamma \|\nabla f(x_k)\|^2\}$$

και

$$a_k \|\nabla f(x_k)\|^2 \geq \min\{1, \gamma \|\nabla f(x_k)\|^2\} \|\nabla f(x_k)\|^2 \geq 0$$

Οπότε αφού $\lim_{k \to \infty} a_k \|\nabla f(x_k)\|^2 = 0$, θα έχουμε

$$\lim_{k \to \infty} \|\nabla f(x_k)\|^2 = 0$$