Διάλεξη 8

Βελτιστοποίηση χωρίς περιορισμούς (πολλές μεταβλητές), Τρίτη 20/3/18

Περιεχόμενα

1. Μια ικανή συνθήκη: κυρτότητα

> ***ΘΕΩΡΗΜΑ 8.1 [Ικανή συνθήκη για βέλτιστη λύση]: Εάν $\vec{x}^* \in A$ όπου $\nabla f(\vec{x}^*) = 0$ και η f είναι κυρτή συνάρτηση τότε η $\vec{x}^*$ είναι βέλτιστη λύση.***

Απόδειξη: Θα δείξουμε ότι για κάθε $\vec{x} \in A$ ισχύει $f(\vec{x}^*) \leq f(\vec{x})$.
Θεωρήστε οποιοδήποτε $\vec{x} \in A$ και έστω $\vec{\delta} = \vec{x} - \vec{x}^*$. Η συνάρτηση $g(t) = f(\vec{x}^* + t\vec{\delta})$ είναι κυρτή (λόγω κυρτότητας της f) και $g'(0) = \nabla f(\vec{x}^*) \cdot \vec{\delta} = 0$, άρα από την ικανή συνθήκη για βέλτιστη λύση του προβλήματος $\min_t g(t)$ (Θεώρημα 6.3) η $t^* = 0$ είναι βέλτιστη λύση. Συνεπώς, $g(0) \leq g(1)$. Όμως, $g(0) = f(\vec{x}^*), g(1) = f(\vec{x})$ ο.ε.δ. ■

Παράδειγμα

Λύστε το πρόβλημα $\min_{x_1, x_2 \in \mathbb{R}} x_1^2 - 2x_1 + x_1x_2 + x_2^2$.

Πρώτα θα βρούμε το σημείο $\vec{x}^* = (x_1^*, x_2^*)$ μηδενισμού του διανύσματος κλίσης της αντικειμενικής συνάρτησης:

$$\nabla f(\vec{x}^*) = 0 \Leftrightarrow \begin{cases} 2x_1^* - 2 + x_2^* = 0 \\ x_1^* + 2x_2^* = 0 \end{cases} \Leftrightarrow \begin{cases} x_1^* = -\frac{4}{3} \\ x_2^* = \frac{2}{3} \end{cases}$$

Τώρα για κάθε $\vec{\delta} = (\delta_1, \delta_2)$ έχουμε

$$\vec{\delta}^T H(f, \vec{x}^*)\vec{\delta} = \begin{bmatrix} \delta_1 & \delta_2 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \delta_1^2 + \delta_2^2 + (\delta_1 + \delta_2)^2 \geq 0$$

άρα η f είναι κυρτή (από Θεώρημα 7.3)

Συνεπώς από την ικανή συνθήκη, η βέλτιστη λύση είναι $x_1^* = -\frac{4}{3}, x_2^* = \frac{2}{3}$. ■

2. Λογισμός κυρτών συναρτήσεων

Για να δείξουμε την κυρτότητα μιας συνάρτησης, βολεύει να αποδείξουμε ότι η συνάρτηση αυτή παράγεται από απλούστερες κυρτές συναρτήσεις σύμφωνα με τους κανόνες του θεωρήματος που ακολουθεί:

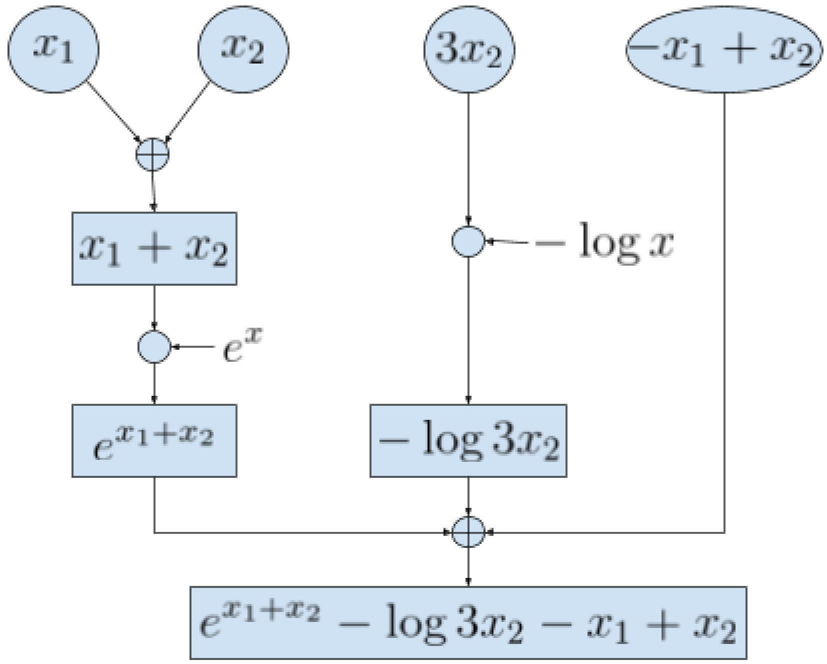
ΘΕΩΡΗΜΑ 8.2 [Κανόνες παραγωγής κυρτών συναρτήσεων]:

1. ***Οι γραμμικές συνάρτησεις*** $\sum_i a_i x_i + b$ ***είναι κυρτές.***
2. ***Εάν*** f_1, f_2 ***κυρτές τότε το άθροισμα*** $f_1 + f_2$ ***είναι κυρτή συνάρτηση.***
3. ***Εάν*** f ***κυρτή και*** $c > 0$ ***τότε η*** cf ***είναι κυρτή.***
4. ***Εάν*** $f : \mathbb{R}^m \to \mathbb{R}$ ***κυρτή και*** A ***πίνακας*** $m \times n$ ***τότε η συνάρτηση*** $h : \mathbb{R}^n \to \mathbb{R}$ ***με*** $h(\vec{x}) = f(A\vec{x})$ ***είναι κυρτή.***
5. ***Εάν*** $f : \mathbb{R}^n \to \mathbb{R}$ ***κυρτή και η συνάρτηση*** $h : \mathbb{R} \to \mathbb{R}$ ***είναι αύξουσα και κυρτή τότε η σύνθεση τους*** $h(f(\vec{x}))$ ***είναι κυρτή.***

Παράδειγμα

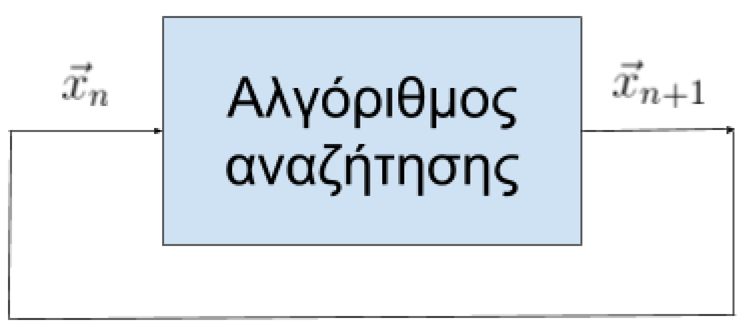
Δείξτε ότι η συνάρτηση $f(x_1, x_2) = e^{x_1 + x_2} - \log 3x_2 - x_1 + x_2$ *είναι κυρτή.*

Η f είναι κυρτή συνάρτηση εφόσον παράγεται από τους κανόνες παραγωγής του παραπάνω θεωρήματος, ξεκινώντας από γραμμικές συναρτήσεις από πάνω προς τα κάτω στο σχήμα:



3. Αλγόριθμοι αναζήτησης σημείου μηδενισμού του διανύσματος κλίσης

Όπως και στην περίπτωση της μιας μεταβλητής, οι αλγόριθμοι αναζήτησης είναι επαναληπτικοί, δηλαδή δέχονται ως είσοδο μια εκτίμηση $\vec{x}_n$ του σημείου μηδενισμού του διανύσματος κλίσης και σε κάθε επανάληψη υπολογίζουν μια νέα (συνήθως καλύτερη) εκτίμηση $\vec{x}_{n+1}$. Στην επόμενη επανάληψη χρησιμοποιείται ως είσοδος η προηγούμενη έξοδος $\vec{x}_{n+1}$ κ.ο.κ.



Ο υπολογισμός της νέας εκτίμησης σε κάθε επανάληψη γίνεται σε δύο διακριτά στάδια:

A. Επιλογή της διεύθυνσης $\vec{\delta}$ κατά την οποία θα αναζητηθεί το $\vec{x}_{n+1}$.
B. Με δεδομένη τη διεύθυνση $\vec{\delta}$ επιλέγεται t και $\vec{x}_{n+1} = \vec{x}_n + t\vec{\delta}$.

Για κάθε στάδιο υπάρχουν αλγόριθμοι οι οποίοι συνδυάζονται με διάφορους τρόπους για να δώσουν ένα ολοκληρωμένο αλγόριθμο αναζήτησης. Αλγόριθμοι του σταδίου Α ονομάζονται **αλγόριθμοι αναζήτησης διεύθυνσης**, ενώ του Β, **αλγόριθμοι αναζήτησης σε γραμμή (line search)**.

Κριτήρια τερματισμού

Τερματίζουμε τη λήψη εκτιμήσεων όταν θεωρούμε ότι η τρέχουσα προσέγγιση $\vec{x}_n$ είναι αρκετά καλή. Βέβαια δεν είναι ποτέ γνωστό πόσο μακρυά βρίσκεται η προσέγγιση από τη βέλτιστη λύση $\vec{x}^*$. Για αυτό τον λόγο επιλέγεται ένα από τα ακόλουθα έμμεσα κριτήρια τερματισμού:

ΚΡΙΤΗΡΙΟ ΙΚΑΝΟΠΟΙΗΤΙΚΗΣ ΠΡΟΟΔΟΥ:
Ο αλγόριθμος τερματίζει όταν η νέα εκτιμήση $\vec{x}_{n+1}$ βρίσκεται αρκετά κοντά στην προηγούμενη εκτίμηση $\vec{x}_n$, πχ, όταν το διανυσματικό μήκος της διαφοράς τους, $|\vec{x}_{n+1} - \vec{x}_n|$, είναι μικρότερο μιας προκαθορισμένης μικρής τιμής, πχ, όταν $|\vec{x}_{n+1} - \vec{x}_n| \leq 10^{-5}$.

ΚΡΙΤΗΡΙΟ ΙΚΑΝΟΠΟΙΗΤΙΚΟΥ ΜΗΔΕΝΙΣΜΟΥ ΤΟΥ ΔΙΑΝΥΣΜΑΤΟΣ ΚΛΙΣΗΣ:
Ο αλγόριθμος τερματίζει όταν το διάνυσμα κλίσης είναι αρκετά κοντά στο να είναι μηδενικό, πχ, το διανυσματικό του μήκος είναι μιας προκαθορισμένης μικρής τιμής. Για παράδειγμα όταν $|\nabla f(\vec{x}_n)| \leq 10^{-5}$.

3.1 Επιλογή διεύθυνσης: η μέθοδος της πιο απότομης κατάβασης

Όπως μαρτυρά και το όνομα, η μέθοδος αυτή επιλέγει τη διεύθυνση προς την οποία η τιμή της αντικειμενικής συνάρτησης ελαττώνεται περισσότερο (κοιτώντας κοντινά σημεία μόνο).

Συγκεκριμένα αν η τρέχουσα εκτίμηση είναι το σημείο $\vec{x}_n$, μια μικρή μεταβολή προς τη διεύθυνση $\vec{\delta}$ μεταβάλλει την τιμή της αντικειμενικής συνάρτησης κατά $f(\vec{x}_n + \epsilon\vec{\delta}) - f(\vec{x}_n)$, όπου $\epsilon > 0$ σταθερά που ελέγχει το μέγεθος της μεταβολής. Για μικρές τιμές του ϵ, η μεταβολή είναι προσεγγιστικά $\epsilon f'(\vec{x}_n + t\vec{\delta})\Big|_{t=0} = \epsilon \nabla f(\vec{x}_n) \cdot \vec{\delta} = \epsilon |\nabla f(\vec{x}_n)| \cdot |\vec{\delta}| \cos\phi$, όπου ϕ η γωνία μεταξύ των διανυσμάτων $\nabla f(\vec{x}_n), \vec{\delta}$. Μεταξύ των διευθύνσεων $\vec{\delta}$ με $|\vec{\delta}| = 1$, η μεταβολή ελαχιστοποιείται για $\cos\phi = -1$, δηλαδή όταν το $\vec{\delta}$ δείχνει στην αντίθετη κατεύθυνση από το διάνυσμα κλίσης. (Παρόμοια, μπορούμε να συμπεράνουμε ότι η κατεύθυνση της μεγαλύτερης αύξησης είναι προς το $\nabla f(\vec{x}_n)$.)

Για αυτό τον λόγο η διεύθυνση $\vec{\delta} = -\nabla f(\vec{x}_n)$ καλείται **η διεύθυνση της πιο απότομης κατάβασης** (steepest descent). Ο *αλγόριθμος* της πιο απότομης κατάβασης σε κάθε επανάληψη επιλέγει να κινηθεί προς την κατεύθυνση αυτή:

ΜΕΘΟΔΟΣ ΤΗΣ ΠΙΟ ΑΠΟΤΟΜΗΣ ΚΑΤΑΒΑΣΗΣ:

Είσοδος: αρχικό σημείο $\vec{x}_0$

1. **Θέσε** $n = 0$
2. **Υπολογισμός διεύθυνσης** $\vec{\delta} = -\nabla f(\vec{x}_n)$
3. **Επιλογή επόμενου σημείου** $\vec{x}_{n+1} = \vec{x}_n + t\vec{\delta}$ **με ακριβή αναζήτηση ή οπισθοδρόμηση**
4. **Επανάληψη από το βήμα 2**