Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

# ΗΥ-111 Απειροστικός Λογισμός ΙΙ
*Εαρινό εξάμηνο 2020*

## Άσκηση Bonus
## Παράδοση: 24/7/2020

**Γενικές οδηγίες**

Το όνομα του παραδοτέου πρέπει να είναι της μορφής **askBonus_AM.pdf** (όπου ΑΜ ο αριθμός μητρώου). Η βαθμολογία της άσκησης είναι προσθετική (bonus) στο τελικό βαθμό του μαθήματος και ίση με το 10% (μέγιστος τελικός βαθμός 110/100). Η αξιολόγηση θα γίνει κυρίως με βάση την αναφορά, γραμμένη σε Word ή latex, η οποία θα περιλαμβάνει την ανάλυση της μεθόδου όπως περιγράφεται από τα επιμέρους μέρη και θα συμπεριλαμβάνει και τον αντίστοιχο κώδικα σε παράθεμα (appendix) στο τέλος. Η παράδοση της άσκησης θα γίνει ηλεκτρονικά μέχρι και τις 24/7/2020 και ώρα 23:59 στην ιστοσελίδα του μαθήματος στο eLearn.

---

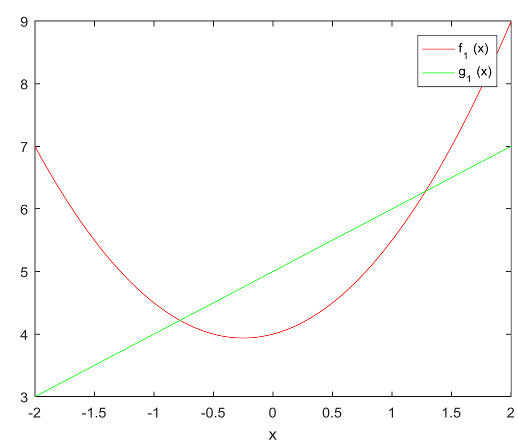
Σκοπός αυτής της άσκησης είναι να εξοικειωθούμε με την μέθοδο gradient descent για την ελαχιστοποίηση συναρτήσεων υπό περιορισμούς (constraint optimization problems). Συγκεκριμένα, υποθέτουμε ότι στόχος είναι να βρούμε τα $x$ τα οποία ελαχιστοποιούν τις παρακάτω συναρτήσεις, με δεδομένους περιορισμούς:

i) $f_1(x) = x^2 + 0.5x + 4$ υπό τον περιορισμό ότι $g_1(x) = x + 5$

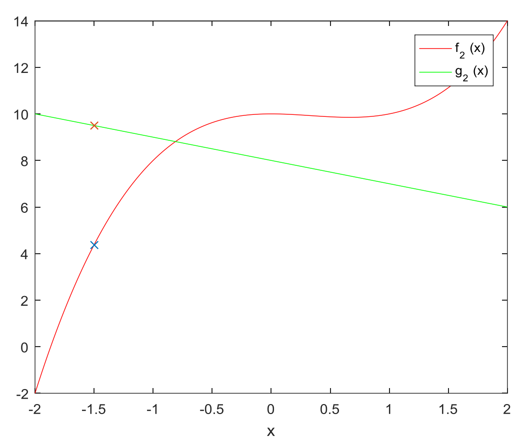
ii) $f_2(x) = x^3 - x^2 + 10$ υπό τον περιορισμό ότι $g_2(x) = -x + 8$

iii) $f_3(x) = e^x - 1$ υπό τον περιορισμό ότι $g_3(x) = -10e^x + 2$

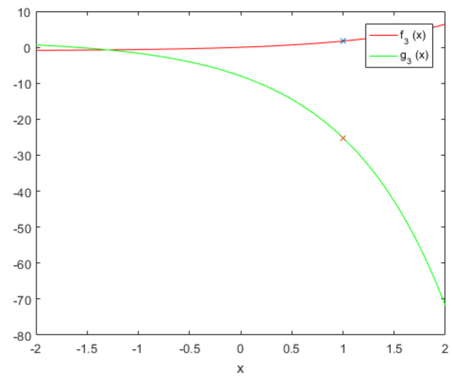
Οι ακόλουθες γραφικές παραστάσεις αντιστοιχούν στις συναρτήσεις




$f_1(x)$ και $g_1(x)$




$f_2(x)$ και $g_2(x)$




$f_3(x)$ $και$ $g_3(x)$

Γενικά, για να βρεθούν τα ελάχιστα μιας συνάρτησης, η μέθοδος gradient descent, ξεκινάει από μια αρχική εκτίμηση για το που ελαχιστοποιείται η συνάρτηση, το $x^0$, και υπολογίζει επαναληπτικά νέες εκτιμήσεις που δίνονται από την εξίσωση

$$x^t = x^{t-1} - a\nabla f(x) = x^{t-1} - a\frac{df(x)}{dx}$$

Η παράμετρος $a$ καθορίζει το μέγεθος του κάθε βήματος, ενώ το $\nabla f(x)$ είναι το διάνυσμα κλίσης της συνάρτησης και $\frac{df(x)}{dx}$ η μερική παράγωγος ως προς $x$. Μεγαλύτερα βήματα (π.χ. $a = 0.1$) κάνουν τη μέθοδο να συγκλίνει γρηγορότερα στο ελάχιστο, αλλά έχουν λιγότερη ακρίβεια στην εξεύρεση λύσης, ενώ μικρότερα βήματα (π.χ. $a = 0.001$) έχουν την ανάποδη συμπεριφορά.

Στην περίπτωση ελαχιστοποίησης με περιορισμούς, εκτός από την συνάρτηση που θέλουμε να ελαχιστοποιήσουμε πχ. $f(x)$, πρέπει να εξασφαλίσουμε ότι ικανοποιούνται και περιορισμοί $g(x) = 0$. Η προσέγγιση που θα δούμε, προτείνει την δημιουργία μιας νέα συνάρτηση, της $L'(x) = f(x) - g(x)$ και την ελαχιστοποίηση αυτής. Μάλιστα, για να μην μας επηρεάζουμε τα πρόσημα της διαφοράς, μια ακόμα καλύτερη λύση είναι να ελαχιστοποιήσουμε την συνάρτηση $L(x) = (f(x) - g(x))^2$. Σε αυτή την περίπτωση, οι επαναληπτικές εκτιμήσεις δίνονται από την εξίσωση

$$x^t = x^{t-1} - a\nabla L(x) = x^{t-1} - a\frac{d(f(x) - g(x))}{dx}$$

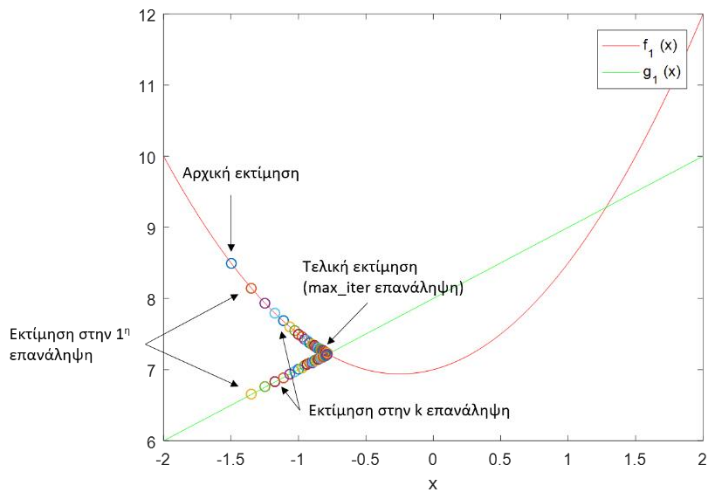
Η επαναληπτική διαδικασία σταματάει όταν ικανοποιηθούν κάποια κριτήρια ελέγχου, για παράδειγμα, όταν ο αριθμός των επαναλήψεων φτάσει την τιμή max_iteration. Η ποιότητα της εκτίμησης δίνεται από το λάθος $L(x^t) - L(x^*)$, όπου $x^*$ είναι η θεωρητική τιμή που ελαχιστοποιεί τη συνάρτηση, ή όταν δεν υπάρχει αυτή, από την διαφορά των δύο συναρτήσεων $|f(x^t) - g(x^t)|$.

Ο ψευδο-κώδικας της μεθόδου gradient descent για προβλήματα με περιορισμούς είναι

- Είσοδος: Αρχικοποίηση της $\boldsymbol{x}^0$ με μία τιμή από το διάστημα ορισμού της $f$ και της $g$. Δημιουργία έκφρασης για την παράγωγο της $L$ (**προσοχή** η $L$ είναι σύνθετη συνάρτηση των $x$ και άρα η παραγώγιση θα πρέπει να είναι η αντίστοιχη).

- Για επανάληψη από 1 ενώ max_iteration
  - $x^t = x^{t-1} - a\nabla L(x)$
- Έξοδος: η τιμή $\boldsymbol{x}$ που ελαχιστοποιεί την $L$

Στην παρακάτω γραφική παράσταση φαίνονται οι διαφορετικές εκτιμήσεις της μεθόδου gradient descent στις διαφορετικές επαναλήψεις, όπως και η τελική εκτίμηση.




## ΕΡΩΤΗΜΑΤΑ

**1ο Μέρος**

Να βρεθούν τα ελάχιστα των συναρτήσεων $[f_1 - g_1]^2, [f_2 - g_2]^2, [f_3 - g_3]^2$.

**2ο Μέρος**

Να υλοποιηθεί η μέθοδος του gradient descent (δείτε τον αντίστοιχο ψευδο-κώδικα) σε κατάλληλη γλώσσα προγραμματισμού (Matlab, Octave, C, Python).

**3ο Μέρος**

Για τις συναρτήσεις $f_1, f_2$ και $f_3$ , υποθέστε ένα συγκεκριμένο αριθμό επαναλήψεων (max_iteration=20) και αρχική τιμή το $x^0 = -0.5$ για την $f_1$, $x^0 = -1.5$ για την $f_2$ και $x^0 = 0$ για την $f_3$ και αξιολογήστε την ποιότητα εκτίμησης ως προς τις διάφορες τιμές της παραμέτρου $a$ (π.χ. 0.001, 0.01 , 0.1).

**4ο Μέρος**

Να εξετάσετε την ποιότητα της προσέγγισης, σαν συνάρτηση του αριθμού επαναλήψεων (max_iteration =5, 10, και 20) από αρχικό σημείο $x^0$ και για συγκεκριμενη τιμή παραμέτρου $a$ (π.χ. 0.01) .

**5ο Μέρος**

Αν οι δύο συναρτήσεις δεν έχουν κοινό σημείο, πχ αν $f_1(x) = x^2 + 0.5x + 4$ και $\hat{g}_1(x) = x + 3 = 0$, η μέθοδος δουλεύει, και αν ναι, τι χαρακτηριστικά έχει το σημείο που βρίσκει?