ΗΥ-111 Απειροστικός Λογισμός ΙΙ
Εαρινό εξάμηνο 2019

4η σειρά ασκήσεων – Bonus
Παράδοση: 19/6/2019

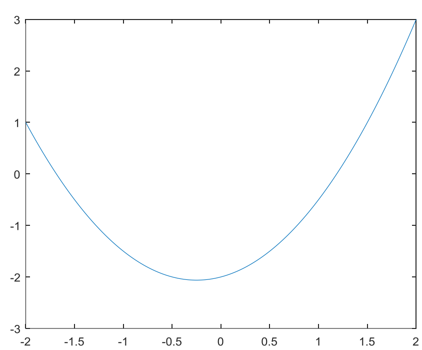
Γενικές οδηγίες

Το όνομα του παραδοτέου πρέπει να είναι της μορφής **ask4_AM** (όπου AM ο αριθμός μητρώου). Η βαθμολογία της άσκησης είναι προσθετική (bonus) στο τελικό βαθμό του μαθήματος και ίση με το 10% (μέγιστος βαθμός 110/100). Η αξιολόγηση θα γίνει με βάση μια 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). Η παράδοση των ασκήσεων θα γίνει ηλεκτρονικά μέχρι και τις 19/6/2019 και ώρα 23:59 από την ιστοσελίδα του μαθήματος στο eLearn.

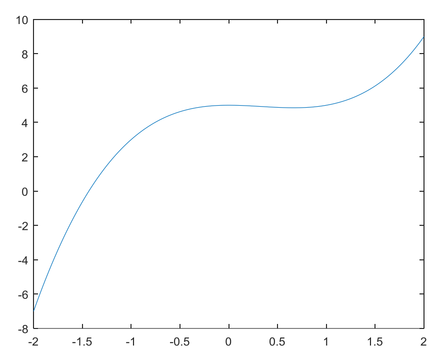
Σκοπός αυτής της άσκησης είναι να εξοικειωθούμε με την μέθοδο gradient descent για την ελαχιστοποίηση συναρτήσεων. Συγκεκριμένα, υποθέτουμε ότι στόχος είναι να βρούμε τα x και y τα οποία ελαχιστοποιούν τις (<u>προσοχή</u> οι i, ii, και iii είναι συναρτήσεις μιας μεταβλητής ενώ η iv είναι δύο μεταβλητών):

i) $f_1(x) = x^2 + 0.5x - 2$
ii) $f_2(x) = x^3 - x^2 + 5$
iii) $f_3(x) = e^x + x^2 + 1$
iv) $f_4(x, y) = x^2 + y^2 + e^x$

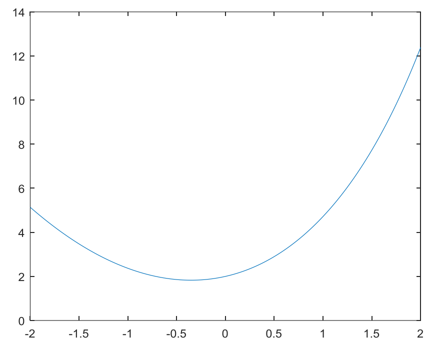
Οι ακόλουθες γραφικές παραστάσεις αντιστοιχούν στις συναρτήσεις



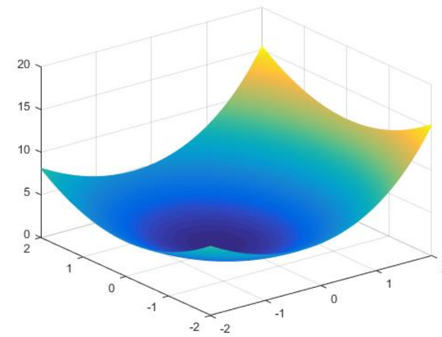
$f_1(x)$



$f_2(x)$



$f_3(x)$



$f_4(x,y)$

Για να βρεθούν τα ελάχιστα, η μέθοδος gradient descent, ξεκινάει από μια αρχική εκτίμηση για το που ελαχιστοποιείται η συνάρτηση x^0 (ή $[x^0, y^0]$ για 2 μεταβλητές) και υπολογίζει επαναληπτικά νέες εκτιμήσεις που δίνονται από την εξίσωση

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

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

Αν η συνάρτηση είναι δύο μεταβλητών x και y, η εξίσωση γράφεται σε διανυσματική της μορφή

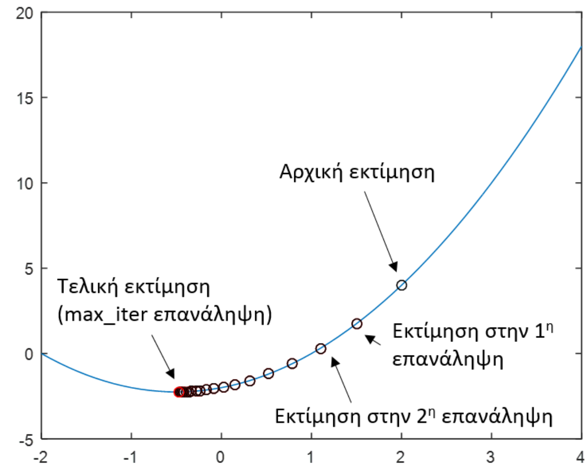
$$[x^t, y^t] = [x^{t-1}, y^{t-1}] - a[\frac{\partial f(x,y)}{dx}, \frac{\partial f(x,y)}{dy}]$$

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 $f(x^t) - f(x^*)$, όπου x^* είναι η θεωρητική τιμή που ελαχιστοποιεί τη συνάρτηση (αντίστοιχα $f(x^t, y^t) - f(x^*, y^*)$ για 2 μεταβλητές).

Ο ψευδο-κώδικας της μεθόδου gradient descent είναι

- Αρχικοποίηση της $\boldsymbol{x}^0$ με μία τιμή από το διάστημα ορισμού της f
- Για επανάληψη από 1 ενώ max_iteration
    - $\boldsymbol{x} = \boldsymbol{x} - a\nabla f(\boldsymbol{x})$
- Η συνάρτηση f ελαχιστοποιείται για $\boldsymbol{x}$ και η τιμή της είναι η $f(\boldsymbol{x})$

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



1ο Μέρος

Να βρεθούν τα ελάχιστα των συναρτήσεων f_1, f_2, f_3, f_4 με την κλασσική μέθοδο ελαχιστοποίησης (θέτοντας το $\nabla f = 0$).

2ο Μέρος

Να υλοποιηθεί η μέθοδος του gradient descent (δείτε τον αντίστοιχο ψευδο-κώδικα) σε κατάλληλη γλώσσα προγραμματισμού (Matlab, Octave, C, Python) για **1** μεταβλητή.

3ο Μέρος

Για τις συναρτήσεις f_1, f_2 και f_3 , υποθέστε ένα συγκεκριμένο αριθμό επαναλήψεων (max_iteration=20) και αρχική τιμή το $x^0 = 2$, 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 $x^0 = 2$ και για συγκεκριμενη τιμή παραμέτρου a (π.χ. 0.01) .

5ο Μέρος

Επεκτείνεται των κώδικα του 2ου Μέρους ασκήσεων από μια διάσταση σε **2** με χρήση της διανυσματικής έκδοσης της μεθόδου gradient descent.

6ο Μέρος

Εξετάστε τη συμπεριφορά της μεθόδου για την συνάρτηση f_4 όσον αφορά την τιμή της παραμέτρου a, και τον αριθμό των επαναλήψεων max_iteration.