Αριθμητικές Μέθοδοι Βελτιστοποίησης

Βασικές έννοιες (Παράγωγοι - Γραμμική Άλγεβρα)

Γραμμική Άλγεβρα
Νόρμες διανυσμάτων - εσωτερικό γινόμενο

Ορισμός: Μια νόρμα $\|\cdot\|$ στον $\mathbb{R}^n$ είναι μια πραγματική συνάρτηση τέτοια ώστε

(i) $\|v\| \geq 0, \quad v \in \mathbb{R}^n$ και $\|v\| = 0$ αν και μόνο αν $v = 0 \in \mathbb{R}^n$

(ii) $\|\lambda v\| = |\lambda|\|v\|, \quad v \in \mathbb{R}^n, \lambda \in \mathbb{R}$

(iii) $\|v + w\| \leq \|v\| + \|w\|, \quad v, w \in \mathbb{R}^n$

Θα χρησιμοποιήσουμε την Ευκλείδεια νόρμα, $\|v\|^2 = \sum_{i=1}^{n} v_i^2$ καθώς και το συνηθισμένο εσωτερικό γινόμενο

$$(v, w) = \sum_{i=1}^{n} v_i w_i = v^T w$$

Γραμμική Άλγεβρα
Νόρμες πινάκων

Ορισμός: Μια νόρμα πίνακα $\| \cdot \|$ στον $\mathbb{R}^{n\times n}$ είναι μια πραγματική συνάρτηση τέτοια ώστε

(i) $\|A\| \geq 0, \quad A \in \mathbb{R}^{n\times n}$ και $\|A\| = 0$ αν και μόνο αν $A = 0 \in \mathbb{R}^{n\times n}$

(ii) $\|\lambda A\| = |\lambda| \|A\|, \quad v \in \mathbb{R}^{n\times n}, \lambda \in \mathbb{R}$

(iii) $\|A + B\| \leq \|A\| + \|B\|, \quad A, B \in \mathbb{R}^{n\times n}$

(iv) $\|AB\| \leq \|A\| \|B\|, \quad A, B \in \mathbb{R}^{n\times n}$

Γραμμική Άλγεβρα
Νόρμες πινάκων - Φυσική νόρμα

Ορισμός: Μια νόρμα πίνακα $\|\cdot\|$ στον $\mathbb{R}^{n\times n}$ καλείται φυσική νόρμα αν παράγεται από μια νόρμα διανυσμάτων ως

$$\|A\| = \sup_{v\in\mathbb{R}^n,\ v\neq 0} \frac{\|Av\|}{\|v\|}$$

Θα χρησιμοποιήσουμε τη φυσική νόρμα πινάκων που παράγεται από την Ευκλείδεια νόρμα διανυσμάτων, και μπορεί να αποδειχθεί ότι

$$\|A\| = \|A\|_2 = \sqrt{\rho(A^T A)}$$

όπου $\rho(A)$ είναι η μέγιστη κατ'απόλυτο τιμή ιδιοτιμή του A.

Γραμμική Άλγεβρα
Θετικά ορισμένοι πίνακες

Ορισμός: Ένας πίνακας $A \in \mathbb{R}^{n \times n}$ καλείται

(i) Θετικά ορισμένος αν $x^T A x > 0, \quad x \in \mathbb{R}^n, x \neq 0$

(ii) Θετικά ημιορισμένος αν $x^T A x \geq 0, \quad x \in \mathbb{R}^n$

(iii) Αρνητικά ημιορισμένος αν $x^T A x \leq 0, \quad x \in \mathbb{R}^n$

(iv) Αρνητικά ορισμένος αν $x^T A x < 0, \quad x \in \mathbb{R}^n, x \neq 0$

Γραμμική Άλγεβρα
Θετικά ορισμένοι πίνακες

Θεώρημα: Αν $A \in \mathbb{R}^{n \times n}$ συμμετρικός, τότε έχει n πραγματικές ιδιοτιμές $\lambda_1, \ldots, \lambda_n$, όχι αναγκαστικά διαφορετικές ανά δύο, και αντίστοιχα ιδιοδιανύσματα $v_1, \ldots, v_n$, τα οποία δημιουργούν μια ορθοκανονική βάση του $\mathbb{R}^n$.

Θεώρημα: Έστω $A \in \mathbb{R}^{n \times n}$ συμμετρικός. Ο A είναι θετικά ορισμένος αν και μόνο αν όλες οι ιδιοτιμές του είναι γνήσια θετικές.

Γραμμική Άλγεβρα
Επίλυση γραμμικών συστημάτων - Απαλοιφή Gauss

Για την επίλυση ενός γραμμικού συστήματος $Ax = b$, με $A \in \mathbb{R}^{n \times n}$, μπορούμε να χρησιμοποιήσουμε τη μέθοδο της απαλοιφής Gauss η οποία ισοδύναμα μπορεί να γραφεί ως εξής: Υπάρχει κάτω τριγωνικός πίνακας L με μονάδες στη διαγώνιο, άνω τριγωνικός πίνακας U και ένας πίνακας μετάθεσης P ,$(P^{-1} = P^T)$ τέτοιοι ώστε

$$A = PLU$$

Το κόστος υπολογισμού της παραπάνω ανάλυσης είναι $\dfrac{1}{3}n^3 + O(n^2)$ πράξεις

Η ανάλυση αυτή καλείται ανάλυση LU του πίνακα A

Γραμμική Άλγεβρα

Επίλυση γραμμικών συστημάτων - Απαλοιφή Gauss

Αλγόριθμος επίλυσης

Αν γνωρίζουμε ότι $A = PLU$, τότε για να λύσουμε το $Ax = b$, πραγματοποιούμε τα εξής βήματα

(i) Λύνουμε $Ly = P^T b$, (αλγόριθμος ευθείας αντικατάστασης)

(ii) Λύνουμε $Ux = y$, (αλγόριθμος οπισθοδρομικής αντικατάστασης)

Γραμμική Άλγεβρα
Ανάλυση Cholesky

- Αν ο πίνακας έχει μια συγκεκριμένη δομή μπορούμε να την εκμεταλευτουμε για να δημιουργήσουμε δυο πίνακες, έναν άνω τριγωνικό και έναν κάτω τριγωνικό, οι οποίοι να δημιουργούνται με λιγότερες πράξεις από ότι οι πίνακες στην ανάλυση LU

- Αν $A \in \mathbb{R}^{n \times n}$ συμμετρικός και θετικά ορισμένος η ανάλυση Cholesky δίνει $A = LL^T$ όπου L κάτω τριγωνικός με θετικά διαγώνια στοιχεία.

Γραμμική Άλγεβρα
Ανάλυση QR

- Αν $A \in \mathbb{R}^{n \times n}$ αντιστρέψιμος η ανάλυση QR δίνει $A = QR$ όπου Q ορθομοναδιαίος πίνακας $(QQ^T = I)$ και R άνω τριγωνικός.

Απειροστικός λογισμός πολλών μεταβλητών
Κλίση

Ορισμός: Μια συνεχή συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$ είναι συνεχώς παραγωγίσιμη στο $x \in \mathbb{R}^n$ αν $\dfrac{\partial f}{\partial x_i}(x)$ υπάρχει και είναι συνεχής για $i = 1, \ldots, n$. Η κλίση της f στο x ορίζεται ως

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right)^T$$

Απειροστικός λογισμός πολλών μεταβλητών

Κατά κατεύθυνση παράγωγος

Λήμμα: Έστω $f : \mathbb{R}^n \to \mathbb{R}$ μια συνεχώς παραγωγίσιμη συνάρτηση σε ένα ανοικτό κυρτό σύνολο $D \subset \mathbb{R}^n$. Για $x \in D$ και $p \in \mathbb{R}^n$, μια μη-μηδενική κατεύθυνση. Η κατά κατεύθυνση παράγωγος της f στο x στη κατεύθυνση p ορίζεται ως

$$\frac{\partial f}{\partial p}(x) = \lim_{\epsilon \to 0} \frac{f(x + \epsilon p) - f(x)}{\epsilon}$$

και είναι ίσο με $p^T \nabla f(x)$. Επίσης υπάρχει $z = x + \alpha p$, $\alpha \in [0,1]$, τέτοιο ώστε

$$f(x + z) = f(x) + p^T \nabla f(z)$$

Απειροστικός λογισμός πολλών μεταβλητών

Εσσιανή

Ορισμός: Μια συνεχή συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$ είναι δύο φορές συνεχώς παραγωγίσιμη στο $x \in \mathbb{R}^n$ αν $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ υπάρχει και είναι συνεχής για $i, j = 1, \ldots, n$. Η Εσσιανή της f στο x ορίζεται ως $n \times n$ πίνακας με στοιχεία

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \quad 1 \leq i, j \leq n$$

Απειροστικός λογισμός πολλών μεταβλητών
Κατά κατεύθυνση παράγωγος

Λήμμα: Έστω $f : \mathbb{R}^n \to \mathbb{R}$ μια συνεχώς παραγωγίσιμη συνάρτηση σε ένα ανοικτό κυρτό σύνολο $D \subset \mathbb{R}^n$. Για $x \in D$ και $p \in \mathbb{R}^n$, μια μη-μηδενική κατεύθυνση. Η δεύτερη τάξεως κατά κατεύθυνση παράγωγος της f στο x στη κατεύθυνση p ορίζεται ως

$$\frac{\partial^2 f}{\partial p^2}(x) = \lim_{\epsilon \to 0} \frac{\frac{\partial f}{\partial p}(x + \epsilon p) - \frac{\partial f}{\partial p}(x)}{\epsilon}$$

και είναι ίσο με $p^T \nabla^2 f(x) p$. Επίσης υπάρχει $z = x + \alpha p$, $\alpha \in [0,1]$, τέτοιο ώστε

$$f(x + z) = f(x) + p^T \nabla f(x) + p^T \nabla^2 f(z) p$$

Απειροστικός λογισμός πολλών μεταβλητών

Ιακωβιανή

Ορισμός: Μια συνεχή συνάρτηση $F : \mathbb{R}^n \to \mathbb{R}^m$ είναι συνεχώς παραγωγίσιμη στο $x \in \mathbb{R}^n$ αν κάθε συνιστώσα συνάρτηση $f_i,\ i = 1,\ldots,m$ είναι συνεχώς παραγωγίσιμη στο x. Η παράγωγος της F στο x είναι ένας πίνακας και καλείται ***Ιακωβιανή της*** F ***στο*** x ορίζεται ως $m \times n$ πίνακας με στοιχεία

$$[F'(x)]_{ij} = \frac{\partial f_i}{\partial x_j}(x), \quad 1 \le i,j \le n$$

και συμβολίζεται ακόμα ως $F'(x) = J(x) = \nabla F(x)^T$

Απειροστικός λογισμός πολλών μεταβλητών

Ενδιάμεση τιμή - Taylor

Λήμμα: Έστω $F : \mathbb{R}^n \to \mathbb{R}^m$ μια συνεχώς παραγωγίσιμη συνάρτηση σε ένα ανοικτό κυρτό σύνολο $D \subset \mathbb{R}^n$ και έστω F' Lipschitz συνεχής στο $x \in D$ με σταθερά Lipschitz L, και $p \in \mathbb{R}^n$, μια μη-μηδενική κατεύθυνση. Τότε

$$\|F(x+p) - F(x) - F'(x)p\| \leq \frac{L}{2}\|p\|^2$$

Απειροστικός λογισμός πολλών μεταβλητών

Λήμμα: Έστω $f : \mathbb{R}^n \to \mathbb{R}$ μια δύο φορές συνεχώς παραγωγίσιμη συνάρτηση σε ένα ανοικτό κυρτό σύνολο $D \subset \mathbb{R}^n$ και έστω $\nabla^2 f(x)$ αντιστρέψιμη, στο $x \in D$. Τότε υπάρχουν σταθερές $\epsilon > 0$, και $\alpha, \beta > 0$, τέτοιες ώστε

$$\alpha \|x - y\| \leq \|\nabla f(y) - \nabla f(x)\| \leq \beta \|x - y\|, \text{ για } y \text{ τέτοια ώστε } \|x - y\| \leq \epsilon$$

Απειροστικός λογισμός πολλών μεταβλητών

Τοπικό ελάχιστο

Λήμμα: Αν μια πραγματική συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$ είναι συνεχώς παραγωγίσιμη σε ένα ανοικτό και κυρτό $D \subset \mathbb{R}^n$. Το σημείο $x^\star \in D$ είναι τοπικό ελάχιστο της f αν $\nabla f(x^\star) = 0$.

Απειροστικός λογισμός πολλών μεταβλητών
Τοπικό ελάχιστο

Λήμμα: Αν μια πραγματική συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$ είναι δύο φορές συνεχώς παραγωγίσιμη σε ένα ανοικτό και κυρτό $D \subset \mathbb{R}^n$ και έστω ότι υπάρχει σημείο $x^\star \in D$ τέτοιο ώστε $\nabla f(x^\star) = 0$. Αν $\nabla^2 f(x^\star)$ θετικά ορισμένος τότε το $x^\star$ είναι τοπικό ελάχιστο της f. Αν $\nabla^2 f$ είναι Lipschitz συνεχής στο $x^\star$, τότε αν $\nabla^2 f(x^\star)$ είναι θετικά ημιορισμένος τότε το $x^\star$ είναι τοπικό ελάχιστο της f

Απειροστικός λογισμός πολλών μεταβλητών
Τοπικό ελάχιστο

Πόρισμα: Αν μια πραγματική συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$ είναι δύο φορές συνεχώς παραγωγίσιμη σε ένα ανοικτό και κυρτό $D \subset \mathbb{R}^n$. Αν $x^\star \in D$, τέτοιο ώστε $\nabla f(x^\star) = 0$, $\nabla^2 f$ είναι Lipschitz συνεχής στο $x^\star$ και $\nabla^2 f(x^\star)$ αντιστρέφεται, τότε το $x^\star$ είναι τοπικό ελάχιστο της f αν και μόνο αν $\nabla^2 f(x^\star)$ είναι θετικά ορισμένος.

Κυρτότητα

Θεώρημα: Έστω $f : \mathbb{R}^n \to \mathbb{R}$ συνεχώς παραγωγίσιμη σε ένα κυρτό σύνολο $D \subset \mathbb{R}^n$. Τότε

(i) Η f είναι κυρτή αν και μόνο αν

$$f(x) + \nabla f(x)(y - x) \leq f(x), \quad \forall x, y \in D$$

(ii) Η f είναι αυστηρά κυρτή αν και μόνο αν

$$f(x) + \nabla f(x)(y - x) < f(x), \quad \forall x, y \in D$$

Κυρτότητα

Πόρισμα: Έστω $f : \mathbb{R}^n \rightarrow \mathbb{R}$, κυρτή και συνεχώς παραγωγίσιμη σε ένα κυρτό σύνολο $D \subset \mathbb{R}^n$. Τότε κάθε κρίσιμο σημείο της f είναι ολικό ελάχιστο.

Θεώρημα: Έστω $f : \mathbb{R}^n \rightarrow \mathbb{R}$ δύο φορές συνεχώς παραγωγίσιμη σε ένα κυρτό σύνολο $D \subset \mathbb{R}^n$. Αν η $\nabla^2 f(x)$ είναι θετικά ημιορισμένη στο D, τότε η f είναι κυρτή στο D.