Αριθμητικές Μέθοδοι Βελτιστοποίησης

Αλγόριθμος ελαχιστοποίησης - Μέθοδος Νεύτωνα

Γενικός Αλγόριθμος

1. Θεωρούμε μια αρχική προσέγγιση $x_0 \in \mathbb{R}^n$
2. Για $k = 0,1,2,\ldots$

    (i) Αν το x_k είναι βέλτιστο σταματάμε

    (ii) Βρείτε μια κατεύθυνση p_k

    (iii) Βρείτε ένα βήμα a_k, ώστε $x_{k+1} = x_k + a_k p_k$ να είναι καλύτερη προσέγγιση από το x_k

Παράδειγμα

Για $f(x) = c^T x$ η p_k είναι κατεύθυνση καθόδου αν

$$c^T(x_k + ap_k) = c^T x_k + ac^T p_k < c^T x_k, \text{ με } a > 0,$$

δηλαδή αν $c^T p_k < 0$.

Αν γνωρίζουμε το p_k, τότε θέλουμε να βρούμε το a_k ώστε να ελαχιστοποιήσουμε

$$\min_{a>0} f(x_k + ap_k)$$

Παρατήρηση

Μπορεί να μην μπορούμε να βρούμε ακριβώς το

$$\min_{a>0} f(x_k + ap_k)$$

Προσπαθούμε να βρούμε ένα a_k που να ελαττώνει αρκετά την f ή να δίνει ένα κατά προσέγγιση ελάχιστο στην κατεύθυνση p_k

Ταχύτητα Σύγκλισης

Θεωρούμε την ακολουθία σφαλμάτων $e_k = x_k - x^\star$, $\lim_{k\to\infty} x_k = x^\star$

Θα λέμε ότι $\{x_k\}$ συγκλίνει στο $x^\star$ με τάξη r αν υπάρχει σταθερά $0 \leq C < \infty$

$$\lim \frac{\|e_{k+1}\|}{\|e_k\|^r} = C$$

Αν $r = 1$ θα λέμε ότι έχουμε γραμμική σύγκλιση

Αν $r = 1$ και $C = 0$, θα λέμε ότι έχουμε υπεργραμμική σύγκλιση

Αν $r = 2$ θα λέμε ότι έχουμε τετραγωνική σύγκλιση

Ταχύτητα Σύγκλισης

Η υπεργραμμική σύγκλιση περιλαμβάνει όλες τις περιπτώσεις που ταχύτητα σύγκλισης είναι $r > 1$, γιατί

$$\lim_{k\to\infty} \frac{\|e_{k+1}\|}{\|e_k\|} = \lim_{k\to\infty} \frac{\|e_{k+1}\|^r}{\|e_k\|} \|e_k\|^{r-1} = C \lim_{k\to\infty} \|e_k\|^{r-1} = 0$$

Ταχύτητα Σύγκλισης

Παράδειγμα:

$$x_{k+1} = \frac{1}{2}(x_k + \frac{4}{x_k}) = \frac{x_k}{2} + \frac{2}{x_k}, \quad x_0 = 4$$

Μπορούμε να δούμε ότι $x_k \to 2$ και ότι $e_{k+1} \equiv x_{k+1} - 2 = \ldots = \frac{1}{2x_k}e_k^2$

Οπότε $\lim \frac{\|e_{k+1}\|}{\|e_k\|^2} = \frac{1}{2 \cdot 2} = \frac{1}{4}$ Δηλαδή έχουμε τετραγωνική σύγκλιση.

Μέθοδος Νεύτωνα

Για την εύρεση της ρίζας μιας μη γραμμικής συνάντησης f μπορούμε να χρησιμοποιήσουμε τη μέθοδο του Νεύτωνα.

Δημιουργούμε μια αναδρομική ακολουθία $\{x_k\}$ σύμφωνα με τον αλγόριθμο

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)},\ k = 0,1,\ldots,$$

Με x_0 μια αρχική προσέγγιση της ρίζας $x^\star$ της f.

Μέθοδος Νεύτωνα

Ένας τρόπος για να προκύψει η μέθοδος είναι να χρησιμοποιήσουμε το ανάπτυγμα Taylor με κέντρο το σημείο x_k

$$f(x_k + p) \approx f(x) + pf'(x_k)$$

για p “μικρό”. Αν είμαστε “κοντά” στη ρίζα $x^\star$μπορούμε να υποθέσουμε ότι $f(x_k + p) \approx 0$

Θέτουμε λοιπόν $f(x) + pf'(x_k) = 0$, οπότε

$$p = -f(x)/f'(x_k)$$

Μέθοδος Νεύτωνα
Ταχύτητα σύγκλισης

Αν θέσουμε $e_k = x_k - x^\star$, παίρνουμε

$$0 = f(x^\star) = f(x_k - e_k) = f(x_k) - e_k f'(x_k) + \frac{1}{2} e_k^2 f''(\xi_k), \quad \xi_k \text{ ανάμεσα σε } x_k \text{ και } x^\star$$

Τότε έχουμε

$$e_k - \frac{f(x_k)}{f'(x_k)} = \frac{1}{2} e_k^2 \frac{f''(\xi_k)}{f'(x_k)}$$

ή διαφορετικά,

$$x_{k+1} - x^\star = x_k - \frac{f(x_k)}{f'(x_k)} - x^\star = \frac{1}{2}(x_k - x^\star)^2 \frac{f''(\xi_k)}{f'(x_k)}$$

Μέθοδος Νεύτωνα

Ταχύτητα σύγκλισης

Αν $\{x_k\}$ συγκλίνει, τότε $\xi_k \to x^\star$, τότε

$$\frac{|x_{k+1} - x^\star|}{|x_k - x^\star|^2} = \frac{1}{2}\frac{|f''(\xi_k)|}{|f'(x_k)|} \to \frac{1}{2}\frac{|f''(x^\star)|}{|f'(x^\star)|}$$

- Αν $f'(x^\star) \neq 0, f''(x^\star) \neq 0$, η σύγκλιση είναι τετραγωνική.
- Αν $f'(x^\star) = 0$, τότε δεν ορίζεται το δεξιό μέλος και η ταχύτητα δεν μπορεί να είναι τετραγωνική
- Αν τώρα $f'(x^\star) \neq 0$ και $f''(x^\star) = 0$, η ταχύτητα θα μπορούσε να είναι μεγαλύτερη από τετραγωνική.

Μέθοδος Νεύτωνα

Ταχύτητα σύγκλισης - Προσέγγιση

Μπορούμε να υπολογίσουμε προσεγγιστικά την ταχύτητα σύγκλισης μιας ακολουθίας ως εξής.

Αν ισχύει ότι

$$\frac{|e_{k+1}|}{|e_k|^p} \to C$$

με $e_k = x_k - x^\star$, τότε για k μεγάλο θα ισχύει ότι $|e_{k+1}| \approx C|e_k|^p$ και $|e_k| \approx C|e_{k-1}|^p$

οπότε $p \approx \dfrac{\log(\frac{e_{k+1}}{e_k})}{\log(\frac{e_k}{e_{k-1}})}$

Μέθοδος Νεύτωνα
Παραδείγματα

Έστω

$$f(x) = 7x^4 + 3x^3 + 2x^2 + 9x + 4,$$

θεωρούμε τη μέθοδο του Νεύτωνα για να βρουμε τη ρίζα $x^\star$, αρχίζοντας με $x_0 = 0$.

Μέθοδος Νεύτωνα
Παραδείγματα

k	x_k	$f(x_k)$	$\lvert x_k - x^\star \rvert$	p
0	0	4	0.511041	
1	-0.444444444444444	0.4048163	0.066597	0.205303
2	-0.506325574893409	0.0264508	0.004716	1.299281
3	-0.511009242860438	0.0001812	3.2545×10^{-05}	1.879446
4	-0.511041786445413	8.8632×10^{-09}	1.5914×10^{-09}	1.994669
5	-0.511041788036866	0.0	0.0	

Μέθοδος Νεύτωνα

Παραδείγματα

Έστω

$$f(x) = x^4 - 7x^3 + 17x^2 - 17x + 6 = (x-1)^2(x-2)(x-3),$$

Υπάρχουν τρεις ρίζες $x^\star = 1,2,3$.

Θεωρούμε τη μέθοδο του Νεύτωνα για να βρουμε τη ρίζα $x^\star$.

Αρχίζοντας με $x_0 = 1.1$ θα προσεγγίσουμε τη ρίζα $x^\star = 1$, $f'(x^\star) = 0$.

Μέθοδος Νεύτωνα

Παραδείγματα

k	x_k	$f(x_k)$	$\lvert x_k - x^\star \rvert$	p
0	1.1	0.017099	0.1	
1	1.04554140127389	0.003868	0.04554140	1.261282
2	1.02193239599271	0.000930	0.02193239	0.928941
…	…	…	…	…
19	1.00000016298492	5.5067×10^{-14}	1.6298×10^{-07}	0.973256
20	1.00000007851839	1.4210×10^{-14}	7.8518×10^{-08}	1.073948

Μέθοδος Νεύτωνα
Παραδείγματα

Έστω

$$f(x) = x^4 - 7x^3 + 17x^2 - 17x + 6 = (x-1)^2(x-2)(x-3),$$

Υπάρχουν τρεις ρίζες $x^\star = 1,2,3$.

Θεωρούμε τη μέθοδο του Νεύτωνα για να βρουμε τη ρίζα $x^\star$.

Αρχίζοντας με $x_0 = 1.9$ θα προσεγγίσουμε τη ρίζα $x^\star = 2$, $f'(x^\star) \neq 0$.

Μέθοδος Νεύτωνα
Παραδείγματα

k	x_k	$f(x_k)$	$\lvert x_k - x^\star \rvert$	p
0	1.9	0.089100	0.1	
1	2.01511627906978	-0.015341	0.015116	0.13008
2	2.00021508490357	-0.000215	0.000215	2.25071
3	2.00000004622174	-4.6221×10^{-08}	4.6221×10^{-08}	1.98597
4	2	7.1054×10^{-15}	2.2204×10^{-16}	2.26797

Μέθοδος Νεύτωνα

Παραδείγματα

Έστω

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}},$$

Υπάρχει μια ρίζα $x^\star = 0$.

Θεωρούμε τη μέθοδο του Νεύτωνα για να βρουμε τη ρίζα $x^\star$.

Αρχίζοντας με $x_0 = 1$ θα προσεγγίσουμε τη ρίζα $x^\star = 0$.

Μέθοδος Νεύτωνα
Παραδείγματα

k	x_k	$f(x_k)$	$\lvert x_k - x^\star \rvert$	p
0	1	0.761594	1	
1	-0.813430203923509	-0.67147	0.813430	-
2	0.409402316583386	0.38796	0.409402	3.3248
3	-0.0473049164556157	-0.04726	0.047304	3.1433
4	$7.0602803644 \times 10^{-05}$	7.0602×10^{-05}	7.0602×10^{-05}	3.0153
5	$-2.345539398 \times 10^{-13}$	-2.345×10^{-13}	2.3455×10^{-13}	3.0001

Μέθοδος Νεύτωνα
Παραδείγματα

Έστω

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}},$$

Υπάρχει μια ρίζα $x^\star = 0$.

Θεωρούμε τη μέθοδο του Νεύτωνα για να βρουμε τη ρίζα $x^\star$.

Αρχίζοντας με $x_0 = 1.1$ θέλουμε να προσεγγίσουμε τη ρίζα $x^\star = 0$, όμως αποτυγχάνουμε.

Μέθοδος Νεύτωνα
Παραδείγματα

k	x_k	$f(x_k)$	$\lvert x_k - x^\star \rvert$	p
0	1.1	0.800499	1.1	
1	-1.12855258526795	-0.810523	1.128552	-
2	1.2341311330391	0.8437734	1.234131	-
3	-1.6951659799228	-0.934802	1.695165	-
4	5.71536010037965	0.999978	5.71536	-
5	-23021.3564857236	-	23021.35	-

Μέθοδος Νεύτωνα

Σε συστήματα μη γραμμικών εξισώσεων

Χρησιμοποιώντας το ανάπτυγμα Taylor για μια $f : \mathbb{R}^n \to \mathbb{R}^n$ θα έχουμε

$$f(x_k + p) \approx f(x_k) + \nabla f(x_k)^T p,$$

Όπου $\nabla f(x)$ είναι η ιακωβιανή (πίνακας $\mathbb{R}^{n\times n}$)

Οπότε με ανάλογο τρόπο προκύπτει η μέθοδος Νεύτωνα για συστήματα εξισώσεων:

$$x_{k+1} = x_k + p$$

$$p = -\nabla f(x_k)^{-T} f(x_k)$$

Βελτιστοποίηση χωρίς περιορισμούς

Έστω $f : \mathbb{R}^n \to \mathbb{R}$, θα αναζητούμε τοπικά ελάχιστα της f γιατί είναι πολύ ευκολότερο να το εξακριβώσουμε από ότι για ένα ολικό ελάχιστο

- $$f(x^\star) \le f(x), \quad \|x - x^\star\| < \epsilon$$ (τοπικό ελάχιστο)
- $$f(x^\star) < f(x), \quad \|x - x^\star\| < \epsilon$$ (αυστηρά τοπικό ελάχιστο)

Βελτιστοποίηση χωρίς περιορισμούς

Θα θεωρήσουμε ότι η $f : \mathbb{R}^n \to \mathbb{R}$ είναι 2 φορές συνεχώς παραγωγίσιμη.

Αν $x^\star$ τοπικό ελάχιστο τότε θα έχουμε $\nabla f(x^\star) = 0$ (κρίσιμο σημείο), και επιπλέον ο πίνακας $\nabla^2 f(x^\star)$ (εσσιανή) θα πρέπει να είναι θετικά ημιορισμένος

Μέθοδος Νεύτωνα

Αναζητούμε τις ρίζες $x^\star$της $\nabla f(x^\star) = 0$, με τη μέθοδο του Νεύτωνα.

Έστω x_0 αρχική προσέγγιση δημιουργούμε την ακολουθία $\{x_k\}$ σύμφωνα με τον αλγόριθμο:

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

$$\nabla^2 f(x_k) p_k = -\nabla f(x_k)$$

Μέθοδος Νεύτωνα

Τετραγωνική σύγκλιση

Θεώρημα: Έστω $f : S \to \mathbb{R}$, $S \subset \mathbb{R}^n$, κυρτό. Υποθέτουμε ότι η $\nabla^2 f(x)$ είναι Lipschitz συνεχής στο S,

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|, \quad \forall x, y \in S,$$

Θεωρούμε την ακολουθία που παράγεται από

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

Αν $x^\star$ είναι τοπικό ελάχιστο της f στο S, $\nabla^2 f(x^\star)$ είναι θετικά ορισμένος και $\|x_0 - x^\star\|$είναι αρκετά μικρό, τότε η $\{x_k\}$ συγκλίνει τετραγωνικά στο $x^\star$