Optimizavimo metodai

Optimizavimas be apribojimų

Optimizavimo uždavinys be apribojimų

- Sprendžiamas optimizavimo uždavinys, kurio kintamiesiems neformuluojami jokie apribojimai, t.y. leistinoji sritis sutampa su visa n-mate Euklido erdve $\mathbb{R}^n$

$$\min_{X \in \mathbb{R}^n} f(X),\; X = (x_1, \ldots, x_n).$$

- Pagal minimumo apibrėžimą patikrinti, ar taškas X yra minimumo tašku, įmanoma tik tada, kai leistinoji sritis A baigtinė, nes tik tada galima palyginti $f(X)$ su visomis kitomis funkcijos reikšmėmis $f(Y),\ Y \in A$.
- Nagrinėjamuoju atveju to padaryti negalima iš principo, nes leistinoji sritis sutampa su $\mathbb{R}^n$.
- Todėl reikia suformuluoti konstruktyvius (patikrinamus) požymius minimumo taškui atpažinti.

Minimumo taškas

- Taškas X^* vadinamas funkcijos $f(X)$ lokaliojo minimumo tašku, jei egzistuoja tokia X^* aplinka $S(X^*)$, kad $f(X^*) \leq f(X)$, $X \in S(X^*)$.
- Taškas X^* vadinamas funkcijos $f(X)$ globaliojo minimumo tašku, jei $f(X^*) \leq f(X)$, $X \in \mathbb{R}^n$.

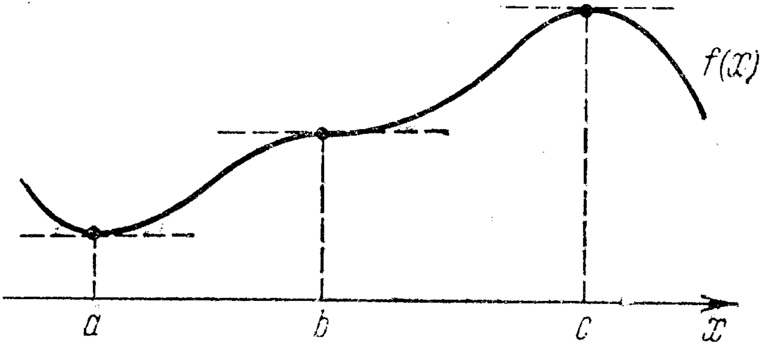
Funkcijos gradientas

- Sakysim, kad funkcija $f(X)$ diferencijuojama.
- Vektorius, sudarytas iš funkcijos dalinių išvestinių, apskaičiuotų taške X, vadinamas funkcijos gradientu ir žymimas
$$\nabla f(X) = \left(\frac{\partial f(X)}{\partial x_1}, \ldots, \frac{\partial f(X)}{\partial x_n} \right).$$

Būtina minimumo sąlyga

- Jei X^* yra $f(X)$ lokaliojo minimumo taškas, tai
$$\nabla f(X^*) = 0.$$
- Įrodoma matematinės analizės kurse remiantis funkcijos Teiloro eilutės pirmosios eilės nariais.
- Trumpai šis įrodymas toks: sakykime taške X^* gradientas nelygus nuliui. Tada, žengus pakankamai trumpą žingsnį antigradiento kryptimi $X = X^* - \gamma \cdot \nabla f(X^*)$, $\gamma > 0$, gaunama funkcijos $f(X)$ reikšmė mažesnė už $f(X^*)$. Reiškia, taškas, kuriame gradientas nelygus nuliui negali būti lokaliojo minimumo tašku.

Būtina minimumo sąlyga nepakankama



- Geometriškai gradiento lygybės nuliui sąlyga reiškia, kad lokaliojo minimumo taške liečiamoji hiperplokštuma yra ortogonali funkcijos reikšmių ašiai.

Hesse matrica

- Sakykime, kad $f(X)$ yra du kartus diferencijuojama.
- Funkcijos antrųjų išvestinių taške X matricą vadinsime Hesse matrica ir žymėsime $H(X)$

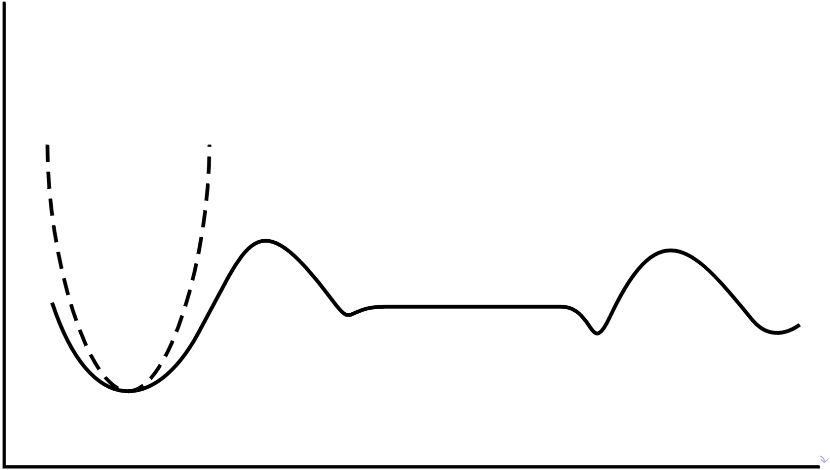
$$H(X) = \begin{pmatrix} \frac{\partial}{\partial x_1^2} f(X) & \frac{\partial}{\partial x_1 \partial x_2} f(X) & \cdots & \frac{\partial}{\partial x_1 \partial x_n} f(X) \\ \frac{\partial}{\partial x_2 \partial x_1} f(X) & \frac{\partial}{\partial x_2^2} f(X) & \cdots & \frac{\partial}{\partial x_2 \partial x_n} f(X) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n \partial x_1} f(X) & \frac{\partial}{\partial x_n \partial x_2} f(X) & \cdots & \frac{\partial}{\partial x_n^2} f(X) \end{pmatrix}.$$

Pakankamos minimumo sąlygos

- X^* yra funkcijos $f(X)$ minimumo taškas, jei jame
  1. $\nabla f(X^*) = 0$,
  2. $H(X^*)$ teigiamai apibrėžta.
- Įrodomos panašiai kaip ir būtinos. Remiamasi tikslo funkcijos skleidiniu Teiloro eilute iki antrosios eilės narių. Sakykime, taške X^* patenkintos pakankamos minimumo sąlygos ir iš jo žengiamas pakankamai trumpas žingsnis bet kuria kryptimi $X = X^* - \gamma \cdot S$. Tada patenkinama nelygybė $f(X) > f(X^*)$ reiškianti, kad X_* yra lokalaus minimumo taškas.

Pakankamų minimumo sąlygų patenkinimo iliustracija

- Hesse funkcijos teigiamas apibrėžtumas reiškia, kad kvadratinė funkcijos aproksimacija lokaliojo minimumo taške yra visomis kryptimis augantis hiperparaboloidas.
- Situacija, kai funkcijos reikšmės kokiame nors intervale pastovios, vadinama “plato”.



Nusileidimo metodų apibūdinimas

- Nusileidimo metodai pagrįsti informacija apie pirmąsias ir antrąsias dalines tikslo funkcijos $f(X)$ išvestines.
- Gradientas nukreiptas funkcijos greičiausio augimo kryptimi. Todėl priešinga – antigradiento kryptis yra reikšminga ieškant minimumo.
- Naudojantis antrosiomis išvestinėmis dažnai galima nustatyti kryptį, kuria galima žengti gan ilgą žingsnį link minimumo, t.y. informacija apie antrąsias išvestines leidžia greičiau artėti prie ieškomo minimumo taško.
- Nusileidimo metodai yra iteraciniai. Pradedama nuo vartotojo parinkto taško X_0 ir generuojama taškų seka $X_{i+1} = X_i + \gamma_i S_i$, čia S_i yra krypties vektorius (dažniausiai vienetinio ilgio), o γ_i žingsnio daugiklis (reiškiantis žingsnio ilgį, jei $\|S_i\| = 1$).
- Norint užtikrinti metodo konvergavimą, reikia, kad žingsnio ilgis tam tikru greičiu trumpėtų.

Gradientinis nusileidimas

- Idėja: žengti žingsnį antigradiento kryptimi.
- Nusileidimo krypties vektorius sutampa su antigradientu $S_i = -\nabla f(X_i)$, o žingsnio daugiklis γ pastovus:

$$X_{i+1} = X_i - \gamma \nabla f(X_i).$$

- Artėjant prie minimumo taško, žingsnio ilgis trumpėja, nes mažėja gradiento norma.
- Žingsnio daugiklis negali būti pernelyg didelis, nes vietoj norimo nuoseklaus artėjimo link minimumo taško nuo pat pradžių gali prasidėti nereguliarus šokinėjimas ilgėjančiais šuoliais.

Teorema

- Sakykime, $f(X), X \in \mathbb{R}^n$, diferencijuojama ir aprėžta iš apačios
$$f(X) \geq f_- > -\infty,$$
o jos gradientas patenkina Lipshitz'o sąlygą
$$\|\nabla f(X) - \nabla f(Z)\| \leq L \cdot \|X - Z\|.$$
Jei žingsnio daugiklis patenkina sąlygą $0 < \gamma < 2/L$, tai gradientinio nusileidimo metodas generuoja taškų seką X_i, kuri monotoniškai mažėja tikslo funkcijos reikšmių prasme, t.y. $f(X_{i+1}) < f(X_i)$, ir konverguoja gradiento prasme, t.y.
$$\lim_{i \to \infty} \nabla f(X_i) = 0.$$

Teorema

- Sakykime, $f(X)$ du kartus diferencijuojama, ir jos antrųjų išvestinių matricos $H(X)$ tikrinės reikšmės λ_i, $i = 1, \dots, n$, patenkina nelygybę $0 < \lambda_0 \leq \lambda_i \leq \lambda_*$ visiems X. Tada gradientinio nusileidimo metodas su žingsnio daugikliu γ, patenkinančiu nelygybę $0 < \gamma < 2/\lambda_*$, konverguoja geometrinės progresijos greičiu į minimumo tašką X^*

$$\|X_i - X^*\| \leq q^i \|X_0 - X^*\|,$$

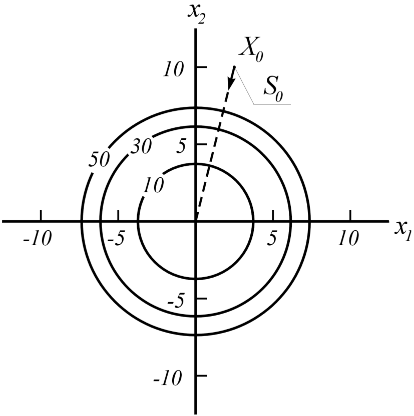
$$q = \max(|1 - \gamma\lambda_0|, |1 - \gamma\lambda_*|) < 1.$$

Metodas konverguoja greičiausiai, kai $\gamma = \gamma^* = 2/(\lambda_0 + \lambda_*)$, nes tada $q = q^* = \frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0}$, yra minimalus.

Palankus atvejis gradientinio nusileidimo algoritmui

- Gautasis konvergavimo greičio įvertis rodo, kad lygybės $\lambda_* = \lambda_0$ atveju tikslus sprendinys gaunamas jau pirmojoje iteracijoje.

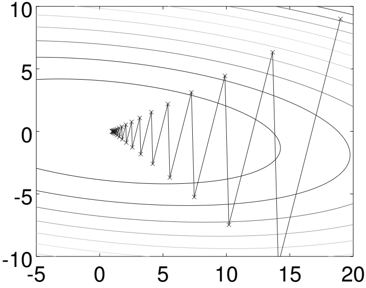
- Bet kuriame taške tokios funkcijos antigradiento kryptis rodo į minimumo tašką. Parinkę žingsnio daugiklį γ^* taip, kaip rekomenduojama teoremos formulavime, jau pirmuoju žingsniu pateksime tiesiai į minimumo tašką.



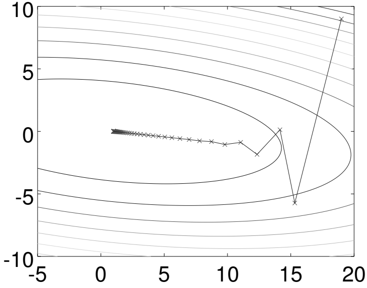
Iliustracija

- Kvadratinė iškila funkcija
- Optimizuojama, kol gradiento norma didesnė už 10^{-4}

80 žingsnių



118 žingsnių



Gradientinio nusileidimo trūkumai

- Teoremos rezultatas rodo, kad esant stipriai besiskiriančioms λ_* ir λ_0 reikšmėms, q yra artimas vienetui. Kadangi santykis tarp maksimalios ir minimalios tikrinių reikšmių praktiniuose uždaviniuose būna 10^6 ir daugiau, tai tokiais atvejais gradientinio metodo konvergavimas būtų akivaizdžiai per lėtas.
- Praktiniu požiūriu gradientinio nusileidimo metodo taikymas susijęs su žingsnio daugiklio parinkimo sunkumais, nes įvertinti Lipshitz'o konstantą gradientui, o juo labiau rėžius tikrinėms matricos $H(X)$ reikšmėms įmanoma tik labai retais atvejais.

Greičiausias nusileidimas

- Idėja: vienmatis minimizavimas antigradiento kryptimi.
- Nusileidimo krypties vektorius sutampa su antigradientu $S_i = -\nabla f(X_i)$ arba $S_i = \frac{-\nabla f(X_i)}{\|\nabla f(X_i)\|}$, o žingsnis parenkamas minimizuojant: $\gamma_i = \arg\min_{\gamma \geq 0} f(X_i + \gamma \cdot S_i)$:

$$X_{i+1} = X_i - \arg\min_{\gamma \geq 0} f(X_i - \gamma \cdot \nabla f(X_i)) \nabla f(X_i),$$

t.y. žingsnis žengiamas iki funkcijos minimumo antigradiento kryptimi.

Teorema

- Sakykime, $f(X), X \in \mathbb{R}^n$, tolygiai diferencijuojama ir aprėžta iš apačios
$$f(X) \geq f_{-} > -\infty.$$
Tada greičiausio nusileidimo metodas konverguoja gradiento prasme, t.y.
$$\lim_{i \to \infty} \nabla f(X_i) = 0.$$
- Šios teoremos sąlygos bendresnės negu anksčiau nagrinėtu (gradientinio nusileidimo metodo su pastoviu žingsnio daugikliu) atveju. Tačiau vykdant šį metodą, atliekamas minimizavimas antigradiento kryptimi, todėl konvergavimas įrodomas netgi paprasčiau.
- Ribinis metodo sugeneruotos sekos taškas yra stacionarus, tačiau su tam tikromis išlygomis galima teigti, kad sugeneruotos sekos ribinis taškas yra lokaliojo minimumo taškas.

Teigiamai apibrėžta kvadratinė funkcija

- Sakykime, kad

$$f(X) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j - \sum_{i=1}^{n} b_ix_i = \frac{1}{2}X \cdot A \cdot X - B \cdot X,$$

  čia matrica A yra teigiamai apibrėžta, t.y. visos jos tikrinės reikšmės yra teigiamos.
- Šios funkcijos gradientas

$$\nabla f(X) = A \cdot X - B.$$

- Funkcijos minimumo taškas gali būti užrašytas analitine formule $X^* = A^{-1}B$. Apskaičiavimams pagal šią formulę užtenka baigtinio operacijų skaičiaus, nors apversti matricą ne visada lengva.

Teorema

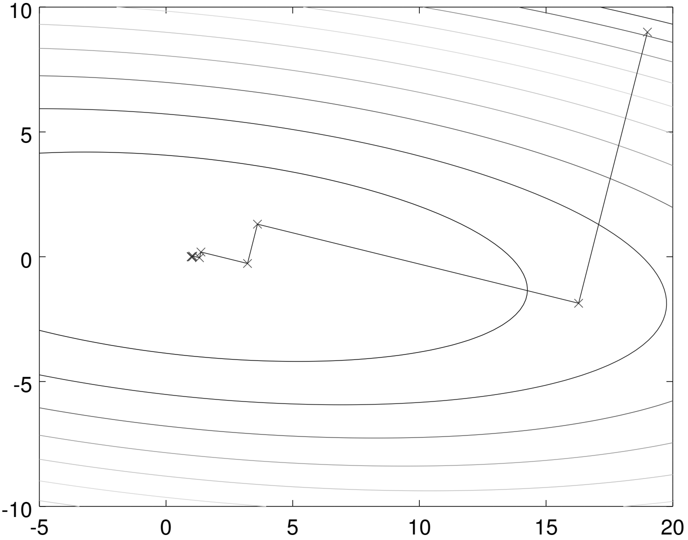
- Greičiausio nusileidimo metodas teigiamai apibrėžtos kvadratinės tikslo funkcijos atveju konverguoja geometrinės progresijos greičiu

$$\|X_i - X^*\| \leq \sqrt{\frac{2}{\lambda_0}\left(f(X_0) - f(X^*)\right)} \left(\frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0}\right)^i,$$

  čia λ_* yra didžiausia, o λ_0 – mažiausia matricos A tikrinės reikšmės.
- Gautas rezultatas nenuteikia optimistiškai: labai palankiu greičiausio nusileidimo metodui atveju, kai minimumo taškas antigradiento kryptimi apskaičiuojamas analitiškai, šio metodo konvergavimo greitis yra toks pat, kaip ir paprastesnio gradientinio nusileidimo metodo konvergavimo greitis platesnei funkcijų klasei.
- Iš gautų rezultatų darytina išvada, kad metodų, kurių nusileidimo kryptis sutampa su antigradiento kryptimi, konvergavimas ne greitesnis už tiesinį.

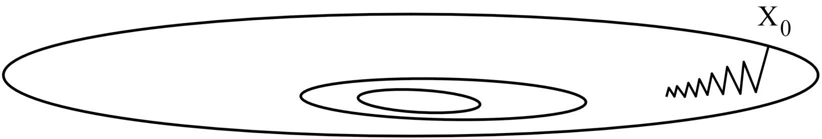
Iliustracija

- Po 16 žingsnių gradiento norma mažesnė už 10^{-4}



Lėto greičiausio nusileidimo metodo konvergavimo iliustracija

- Palyginę abiejų metodų konvergavimo greičio įverčius matome, kad jiems palankus simetriškos funkcijos atvejis. Bet tokia funkcija iš esmės tepriklauso nuo vieno kintamojo, ir šis atvejis praktiniu požiūriu nelabai įdomus.
- Esant stipriai nesimetrinėms funkcijoms, vaizdžiai apibūdinamoms “gilaus griovio” įvaizdžiu, abu metodai konverguoja lėtai.



Nusileidimo metodų apibūdinimas

- Nusileidimo metodai pagrįsti informacija apie pirmąsias ir antrąsias dalines tikslo funkcijos $f(X)$ išvestines.
- Gradientas nukreiptas funkcijos greičiausio augimo kryptimi. Todėl priešinga – antigradiento kryptis yra reikšminga ieškant minimumo.
- Pirmosios funkcijos išvestinės (gradientas) apibrėžia tiesinę funkcijos aproksimaciją. Pastaroji gali būti pakankamai tiksli tik tam tikroje, dažnai nedidelėje, taško, kuriame apskaičiuotos išvestinės, aplinkoje.
- Teoriškai ir antrųjų išvestinių informacija apie funkcijos kitimą patikima tik taško, kuriame jos apskaičiuotos, aplinkoje. Tačiau praktiškai tokia informacija dažnai leidžia prognozuoti funkcijos kitimo pobūdį didesnėje srityje.
- Naudojantis antrosiomis išvestinėmis dažnai galima nustatyti kryptį, kuria galima žengti gan ilgą žingsnį link minimumo, t.y. informacija apie antrąsias išvestines leidžia greičiau artėti prie ieškomo minimumo taško.

Niutono metodas

- Remdamiesi būtinosiomis minimumo sąlygomis, galime uždavinį spręsti netiesiogiai, pakeisdami jį lygčių sistemos $\nabla f(X) = 0$ sprendimu, ir pasinaudodami šių uždavinių sprendimo teorija bei sukaupta patirtimi.
- Niutono metodas yra vienas iš klasikinių metodų lygčių sistemoms spręsti. Kairiosiose lygčių sistemos pusėse esančios funkcijos ištiesinamos taške X_i ir, išsprendus tiesinę lygčių sistemą, turėtasis taškas pakeičiamas tiesinės lygčių sistemos sprendiniu X_{i+1}.
- Norėdami realizuoti šią idėją ir ištiesinti gradiento komponentes sudarančias funkcijas, pasinaudosime tikslo funkcijos $f(X)$ Teiloro eilutės nariais iki antrosios eilės:

$$\nabla f(X) \approx \nabla f(X_i) + H(X_i) \cdot (X - X_i).$$

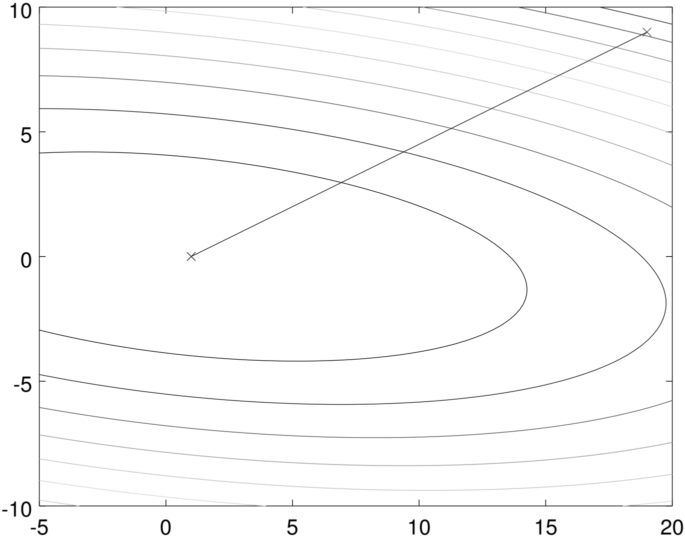
  Prilyginę nuliui, gausime iteracinę metodo formulę

$$X_{i+1} = X_i - H^{-1}(X_i) \cdot \nabla f(X_i).$$

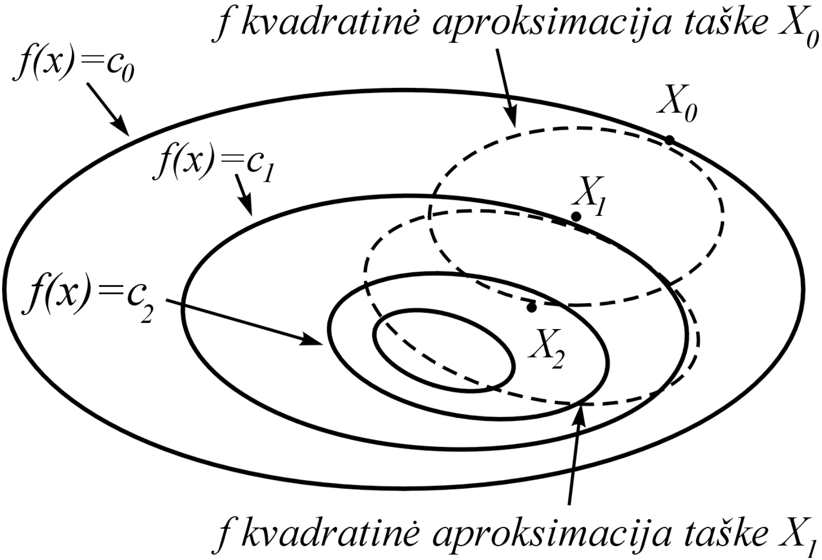
Niutono metodo interpretacija

- Iteracine formule išreiškiamas tikslo funkcijos $f(X)$ kvadratinės aproksimacijos minimumo taškas.
- Taške X_i tikslo funkcija aproksimuojama Teiloro eilute, įvertinant ne aukštesnės negu antros eilės narius.
- Apskaičiuojamas aproksimuojančiosios kvadratinės funkcijos minimumo taškas X_{i+1} ir juo pakeičiamas X_i.
- Metodas apibrėžtas naudojant pirmąsias ir antrąsias išvestines, bet nenaudojant tikslo funkcijos reikšmių.
- Jei tikslo funkcija būtų kvadratinė, jos minimumas būtų gautas vienu žingsniu.
- Bendru atveju formulė taikoma iteratyviai.

Iliustracija kvadratinei funkcijai



Niutono metodo iliustracija



Kita Niutono metodo interpretacija

- Vektoriaus daugyba iš matricos geometriškai reiškia tiesinę erdvės transformaciją, t.y. vektoriaus pasukimo ir ištempimo (suspaudimo) operacijas.
- Todėl vektorius $H^{-1}(X_i) \cdot \nabla f(X_i)$ gali būti interpretuojamas kaip gradiento vektorius tiesiškai transformuotoje erdvėje, o Niutono metodo žingsnio kryptis, kaip antigradiento kryptis matrica $H(X_i)$ apibrėžtoje metrikoje.

Teorema

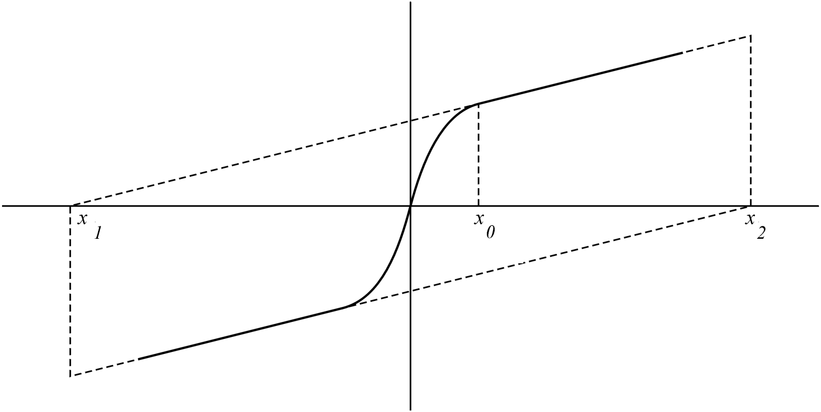
- Sakysime, kad egzistuoja taškas X^*, kuriame patenkintos pakankamos minimumo sąlygos ir taško X^* $\overline{\delta}$-aplinkoje $H(X)$ tolydi.
- Jei pradinis taškas X_0 parenkamas pakankamai artimoje minimumo taško aplinkoje, tai Niutono metodo generuojama seka konverguoja į minimumo tašką greičiau negu tiesiškai.
- Jei be to $H(X)$ patenkina Lipschitz'o sąlygą

  $$\|H(X) - H(Z)\| \leq L \cdot \|X - Z\|,$$

  tai Niutono metodas konverguoja kvadratiniu greičiu.
- Jei patenkintos teoremos sąlygos, tai Niutono metodas konverguoja labai greitai. Iš pirmo žvilgsnio tas tikrai patrauklu. Tačiau teoremos sąlygas galima užtikrinti tik parinkus tokį pradinį tašką, kuris būtų pakankamai artimas sprendiniui.

Niutono metodo užsiciklinimas

- Iš blogo pradinio taško Niutono metodas gali apskritai nekonverguoti.
- Tokią situaciją iliustruoja paveikslas, kuriame pavaizduotos funkcijos $f'(x)$ atveju šaknies $f'(x) = 0$ paieška Niutono metodu užsiciklina.



Niutono metodo trūkumai

- Pavyzdys rodo, kad Niutono metodas gali nekonverguoti net iškilosios tikslo funkcijos atveju.
- Ypač neaiškus metodo konvergavimo klausimas, jei metodo vykdymo metu gaunamuose taškuose antrųjų išvestinių matrica nėra teigiamai apibrėžta.
- Jei kokiame nors taške X_j matrica $H(X_j)$ būtų neigiamai apibrėžta, tai iš to taško būtų žengiamas žingsnis į kvadratinės aproksimacijos maksimumą.
- Vertinant praktinio panaudojimo požiūriu, Niutono metodo privalumas (didelis konvergavimo greitis) dažnai nustelbiamas jo trūkumų: siauros konvergavimo srities ir realizavimo sunkumų, sąlygojamų antrųjų išvestinių skaičiavimu bei jų matricos apvertimu.

Niutono metodo privalumai, modifikacijos

- Niutono metodas ypač vertingas tokiais atvejais, kai būtinas labai didelis sprendinio tikslumas ir turimas geras pradinis taškas, pavyzdžiui, surastas kitu metodu, turinčiu platesnę konvergavimo sritį.
- Siekiant susilpninti griežtas konvergavimo sąlygas, buvo pasiūlytos įvairios Niutono metodo modifikacijos.
- Viena iš jų tokia – modifikuojama kvadratinė tikslo funkcijos aproksimacija taip, kad antrosios eilės išvestinių matrica būtų visada teigiamai apibrėžta.
- Kitoje modifikacijoje kvadratinė aproksimacija vartojama tik mažėjimo krypčiai einamojo taško aplinkoje nustatyti; po to vienmatis minimizavimas atliekamas nustatytąja kryptimi panašiai, kaip greičiausio nusileidimo metode. Tokiu būdu praplečiama konvergavimo sritis, tačiau nebeužtikrinamas kvadratinis konvergavimo greitis.

Patikimos srities metodai

- Anksčiau nagrinėti metodai pagrįsti tiesine (gradientinis ir greičiausio nusileidimo metodai) arba kvadratine (Niutono metodas) aproksimacija.
- Tokios aproksimacijos gerai aprašo tikslo funkciją tam tikroje srityje, kuri paprastai sutampa su einamojo taško aplinka. Bet už tos aplinkos ribų aproksimuojančioji ir aproksimuojamoji funkcijos gali labai skirtis.
- Patikimos srities (trust region) metoduose aproksimacija naudojama tik tokioje srityje, kurioje aproksimacija yra tikrai priimtina.
- Šioje paskaitoje paaiškinsime tik pačią patikimos srities metodų idėją, ją panaudodami Niutono metodo modifikavimui.

Patikimos srities Niutono metodas

- Niutono metodo $i+1$-masis žingsnis žengiamas į $f(X)$ kvadratinės aproksimacijos $f^i(X)$ minimumo tašką, čia $f^i(X)$ yra $f(X)$ Teiloro eilutės taško X_i aplinkoje du pirmieji nariai

$$f^i(X) = f(X_i) + \nabla f(X_i) \cdot (X - X_i) + \frac{1}{2}(X - X_i) \cdot H(X_i) \cdot (X - X_i).$$

- Teoriškai $f^i(X)$ yra gera $f(X)$ aproksimacija taško X_i aplinkoje.
- Jei, savo ruožtu, X_i yra minimumo taško X^* aplinkoje, tai tikėtina, kad $f^i(X)$ minimumo taškas yra arti X^*. Tuo pagrįstas Niutono metodo greitas konvergavimas iš gero pradinio taško.
- Jei X_i yra toli nuo $f(X)$ minimumo taško X^*, tai $f^i(X)$ gali ne tik kad labai skirtis nuo $f(X)$ taško X^* aplinkoje, bet ir $f^i(X)$ minimumo taškas gali būti labai nutolęs nuo X^*. Kaip jau anksčiau minėjome, iš blogo pradinio taško Niutono metodas gali apskritai nekonverguoti.

Žingsnis patikimoje srityje

- Atsižvelgdami į tai, kad $f^i(X)$ yra gera $f(X)$ aproksimacija taško X_i aplinkoje, galime modifikuoti Niutono metodą šitaip
$$X_{i+1} = \arg \min_{\|X - X_i\| \leq \delta} f^i(X).$$
- Svarbu pažymėti, kad net tuo atveju, kai $H(X_i)$ nėra teigiamai apibrėžta, tokia iteracinė formulė sumažina tikslo funkcijos reikšmę.
- Naujasis taškas apskaičiuojamas sprendžiant kvadratinės funkcijos minimizavimo uždavinį esant apribojimui $\|X - X_i\| \leq \delta$.
- Kadangi uždavinyje tėra vienintelis apribojimas, tai jį galima palyginti lengvai įvertinti.

Patikimos srities metodo privalumas

- Žengiant žingsnį patikimoje srityje, galima užtikrinti monotonišką tikslo funkcijos mažėjimą ir metodo konvergavimą iš bet kokio pradinio taško.
- Priartėjus prie minimumo taško tiek, kad apribojimas $\|X - X_i\| \leq \delta$ nebeturi reikšmės, metodas nebesiskiria nuo Niutono metodo.
- Tuo užtikrinama ir plati metodo konvergavimo sritis, ir didelis konvergavimo greitis.

Kintamos metrikos metodai

- Niutono metodas buvo modifikuojamas įvairiai.
- Viena iš idėjų pasirodė ypač sėkminga ir ja remiantis buvo sukurti metodai, vadinami kintamos metrikos arba kvaziniutono metodais.
- Juose antrųjų išvestinių matrica nėra naudojama, o atvirkštinė antrųjų išvestinių matrica aproksimuojama remiantis tik informacija apie pirmąsias funkcijos išvestines.
- Vienas iš žymiausių šios klasės metodų – DFP (vadinamas jo autorių Davidon-Fletcher-Powell vardu).

DFP (Davidon-Fletcher-Powell) metodas

1. Inicializacija

- Parinkti pradinį tašką X_0.
- Suteikti reikšmę sustojimo parametrui ε.
- Suteikti reikšmes $n \times n$ matricos D_1 elementams taip, kad matrica būtų simetrinė ir teigiamai apibrėžta, pvz. vienetinė.
- $k = 0$.
- $j = 1$.
- $Y_1 = X_0$.

DFP (Davidon-Fletcher-Powell) metodas

2. Minimizavimas nusileidimo kryptimi

- Jei
$$\|\nabla f(Y_j)\| < \varepsilon$$
sustoti.
- Priešingu atveju tęsti: apskaičiuoti nusileidimo krypties vektorių $S_j = -\nabla f(Y_j) \cdot D_j$; rasti
$$\gamma_j = \arg\min_{\gamma > 0} f(Y_j + \gamma S_j),$$
ir pereiti į naują tašką
$$Y_{j+1} = Y_j + \gamma_j \cdot S_j,$$
$$j = j + 1.$$
- Jei $j \leq n + 1$, vykdyti 3 žingsnį.
- Priešingu atveju $k = k + 1$, $j = 1$, $X_k = Y_1 = Y_{n+1}$, kartoti 2 žingsnį.

DFP (Davidon-Fletcher-Powell) metodas

3. Matricos perskaičiavimas

- Perskaičiuojama D matrica:

$$D_{j+1} = D_j + \frac{P_j^t \cdot P_j}{P_j \cdot Q_j^t} - \frac{D_j \cdot Q_j^t \cdot Q_j \cdot D_j}{Q_j \cdot D_j \cdot Q_j^t},$$

  čia P_j ir Q_j yra vektoriai-eilutės

$$P_j = \gamma_j S_j = Y_{j+1} - Y_j,$$

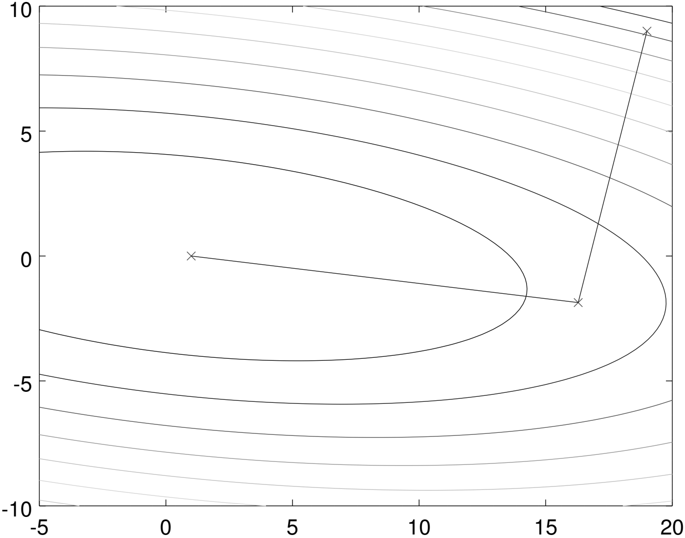
$$Q_j = \nabla f(Y_{j+1}) - \nabla f(Y_j).$$

- $j = j + 1$, vykdyti 2 žingsnį.
- Atkreipsime dėmesį į tai, kad kas n vienmačių minimizavimų matricai suteikiama pradinė reikšmė. Tas daroma siekiant išvengti paklaidų susikaupimo.

Kintamos metrikos metodų privalumai

- DFP yra vienas iš daugelio šios klasės metodų.
- Vienos iteracijos atlikimo laikas proporcingas n^2.
- Kvadratinės funkcijos atveju šie metodai išsprendžia uždavinį per n žingsnių.
- Šiuo požiūriu jų efektyvumas toks pat kaip ir Niutono metodo, kadangi Niutono metodo iteracijos atlikimo laikas proporcingas n^3.
- Kintamos metrikos metodai konverguoja iš bet kokio pradinio taško.
- Konvergavimo greitis yra virštiesinis n-žingsnių iteracijų skaičiaus atžvilgiu.
- Viena n-žingsnė iteracija apima n vienmačių minimizavimų.
- Kintamos metrikos metodai yra efektyvūs ir plačiai naudojami.

Iliustracija kvadratinei funkcijai



Jungtinių krypčių metodai

- Jungtinių krypčių (conjugate directions) metodai remiasi kvadratine tikslo funkcijos aproksimacija.
- Atskiros jų versijos buvo pasiūlytos apie 1950 metus.
- Metodų pavadinimas kilo iš to, kad kvadratinės funkcijos atveju šių metodų generuojamos nusileidimo kryptys yra jungtinės kvadratinę funkciją aprašančios matricos atžvilgiu.
- Kintamos metrikos metodai yra atskiras šios klasės metodų atvejis, kadangi kvadratinės tikslo funkcijos atveju vienmačiai minimizavimai vykdomi jungtinėmis kryptimis.
- Paprastai, kai kalbama apie jungtinių krypčių metodus, turimi galvoje metodai, kurių vienai iteracijai atlikti reikia laiko proporcingo n.
- Kitais atvejais jie vadinami siauresnės klasės vardu, pvz. kintamos metrikos metodais.

Jungtinių krypčių metodai

- Sakykime, tikslo funkcija yra kvadratinė, apibrėžiama formule
$$f(X) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j - \sum_{i=1}^{n} b_i x_i.$$
- Ortogonalia Euklido erdvės transformacija galima pakeisti kintamuosius, kad funkcija jais būtų užrašoma
$$f(V) = \sum_{j=1}^{n} a_i v_i^2.$$
- Naujojoje koordinačių sistemoje kvadratinės funkcijos minimumas iš bet kokio pradinio taško randamas atlikus n vienačių minimizavimų išilgai naujų koordinačių ašių.

Koordinačių transformavimas

- Minimas koordinačių transformavimo uždavinys ekvivalentus matricos A tikrinių vektorių apskaičiavimo uždaviniui, kuris skaičiavimų sudėtingumo požiūriu nėra lengvas.
- Bet nebūtina atlikti koordinačių transformavimo išreikšta forma. Tą patį efektą gautume, jei pradinėje koordinačių sistemoje atliktume minimizavimą naujosios sistemos koordinačių vektorių kryptimis.
- Teisingas dar bendresnis teiginys: minimumo tašką galima pasiekti atlikus n vienmačių minimizavimų kryptimis, atitinkančiomis matricos A jungtines kryptis. Pastarąsias galima apskaičiuoti truputį modifikavus greičiausio nusileidimo metodą.

Flecher-Reeves jungtinių gradientų metodas

- Viena iš jungtinių krypčių metodo versijų yra Flecher-Reeves jungtinių gradientų metodas, kuris aprašomas šitaip:

$$S_0 = -\nabla f(X_0),\ i = 0,$$

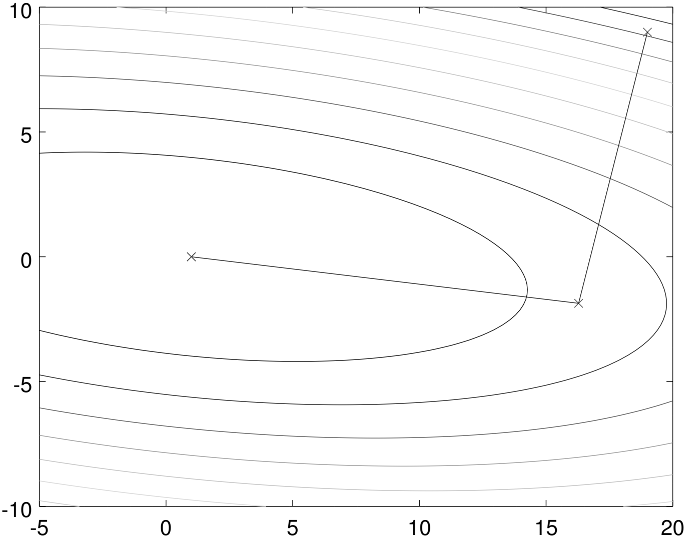
$$X_{i+1} = X_i + \gamma_i \cdot S_i,\ \gamma_i = \arg\min_{\gamma > 0} f(Y_i + \gamma S_i),$$

$$i = i + 1,$$

$$\beta_i = \frac{\nabla f(X_i) \cdot \nabla f(X_i)}{\nabla f(X_{i-1}) \cdot \nabla f(X_{i-1})},$$

$$S_i = -\nabla f(X_i) + \beta_i \cdot S_{i-1}.$$

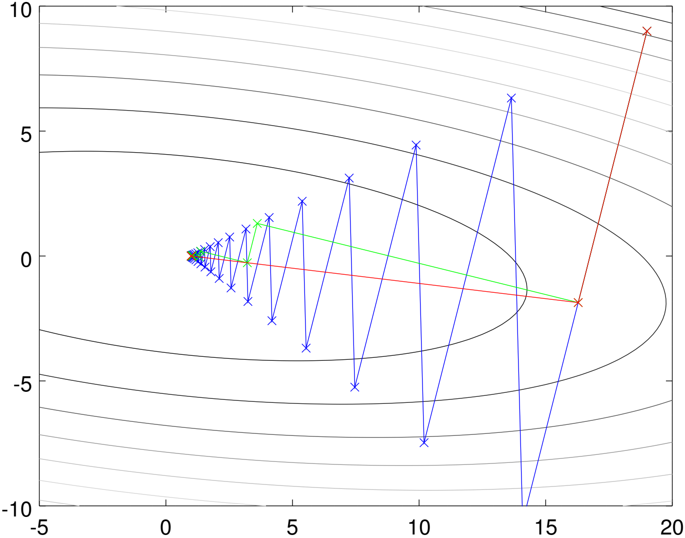
Iliustracija



Jungtinių krypčių metodų savybės

- Paeiliui einančias n vienmačių minimizavimų pavadinkime n-žingsne iteracija.
- Galima įrodyti, kad šiuo metodu kvadratinės funkcijos minimumas randamas per vieną n-žingsnę iteraciją.
- Bendruoju (nekvadratinės tikslo funkcijos) atveju konvergavimo greitis yra virštiesinis n-žingsnių iteracijų skaičiaus atžvilgiu.
- Praktinio efektyvumo požiūriu jungtinių krypčių metodai nusileidžia kintamos metrikos metodams ir yra jautresni išvestinių įvertinimo bei vienmačio minimizavimo paklaidoms.
- Tačiau jie naudotini tada, kai kintamųjų yra tiek daug, kad matricinėms operacijoms reikalingas n^2 proporcingas laikas yra per ilgas, o minėtos paklaidos yra paneigtinos.
- Programinė metodų realizacija ne ką sudėtingesnė už greičiausio nusileidimo metodo realizaciją.

Apibendrinimas



Tiesioginės paieškos metodai

- Nusileidimo metodai pagrįsti tam tikromis prielaidomis apie tikslo funkciją ir remiasi informacija apie tikslo funkcijos išvestines. Tačiau tokios prielaidos ne visada adekvačios sprendžiamam uždaviniui. Jei tikslo funkcija nėra glotni, tai jos hiperpaviršiaus nereguliarumai gali užgožti esmines funkcijos reikšmių mažėjimo kryptis: vaizdžiai kalbant, pernelyg pasikliaudami lokalia informacija (išvestinėmis), galime įstrigti artimiausioje baloje vietoj judėję kalno šlaitu žemyn.
- Paieškos metodai realizuoja įvairias labai skirtingas idėjas, todėl juos nelengva apibrėžti kaip metodų klasę. Kai kurie metodai modeliuoja gamtos reiškinius, gyvūnų elgseną ar net žmonių euristines sprendinio paieškos strategijas. Labiausiai charakteringa šių metodų savybė ta, kad jiems pagrįsti nėra sudaromi kiekvieno žingsnio lokalieji tikslo funkcijos modeliai.

Gauss-Seidel metodas (koordinatinė paieška)

- Kiekvienoje iteracijoje ieškoma tikslo funkcijos reikšmės vieno iš kintamųjų atžvilgiu kitiems esant fiksuotiems.
- Iš pradinio taško judama pagal pirmąją koordinatę x_1 kol pasiekiamas vienmatis minimumas. Iš gautojo taško judama pagal antrąją koordinatę kol pasiekiamas minimumas ir t.t. Formaliai metodą galima aprašyti šitaip

$$X_{k+1} = (x_{k1}, x_{k2}, \ldots, x_{ki}, x_{k+1\, i+1}, x_{k\, i+2}, \ldots, x_{kn}),$$

$$x_{k+1\, i+1} = \arg\min_t f(x_{k1}, x_{k2}, \ldots, x_{ki}, t, x_{k\, i+2}, \ldots, x_{kn}).$$

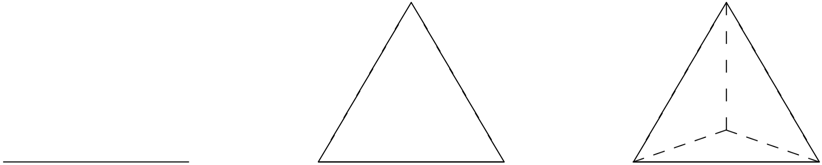
- Metodas konverguoja į globalųjį minimumą, jei tikslo funkcija iškiloji, o vieno kintamojo funkcija, gaunama iš $f(X)$ fiksavus $n-1$ kintamojo reikšmes, griežtai iškiloji.
- Metodas efektyvus daugiaekstremėms funkcijoms

$$f(X) = \sum_{i=1}^{n} f_i(x_i),$$

jei turimas efektyvus metodas funkcijoms $f_i(\cdot)$.

Simplekso metodas optimizavimui be apribojimų

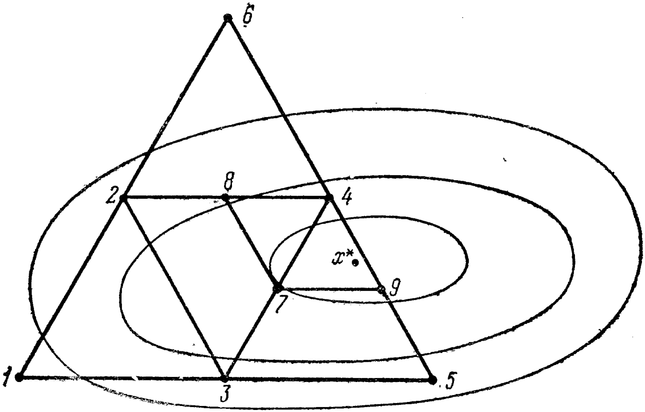
- Literatūroje aprašyta nemažai atvejų, kai deformuojamo simplekso metodu buvo sėkmingai išspręsti uždaviniai, kuriems dėl vienokių ar kitokių priežasčių netiko nusileidimo metodai.
- Simpleksas – trikampio apibendrinimas – n-matis iškilasis daugiakampis su $n+1$ tiesiškai nepriklausoma viršūne.



Taisyklingo simplekso atvejas

- Metodo idėją paaiškinsime taisyklingo simplekso atveju.
- Simpleksas vadinamas taisyklingu, jei jo viršūnės vienodai nutolusios viena nuo kitos.
- Sudaromas taisyklingas $n+1$ viršūnės simpleksas, kurio centras yra pasirinktasis pradinis taškas.
- Kraštinės ilgį rekomenduotina parinkti taip, kad jis būtų artimas nuspėjamam pradinio taško atstumui iki minimumo taško.
- Simplekso viršūnėse apskaičiuojamos tikslo funkcijos reikšmės.
- Naujasis taškas yra viršūnės, kurioje tikslo funkcijos reikšmė didžiausia, atspindys prieš ją esančios sienos atžvilgiu.

Simplekso algoritmo iliustracija



- Pradinis simpleksas yra trikampis 1,2,3.
- "Blogiausia" 1, jos atspindys 4 "geresnė", trikampis 2,3,4.
- "Blogiausia" 2, jos atspindys 5 yra "blogiausia" 3,4,5.
- Galima atspindėti 3, bet ir šis žingsnis nesėkmingas.
- Redukcija dalinant kraštines pusiau: 4,7,8.

Simplekso algoritmas

1. Sudaromas pradinis simpleksas ir visose jo viršūnėse apskaičiuojamos tikslo funkcijos reikšmės. Simplekso dydį reguliuoja parametras α.
2. Randama didžiausia $f(X)$ reikšmė ir ją atitinkanti viršūnė.
3. Ši viršūnė perstumiama tiese, nubrėžta per likusių viršūnių svorio centrą, ir tampa nauja simplekso viršūne.
4. 2-3 žingsniai kartojami tol, kol simpleksas neuždengs minimumo taško arba neprasidės ciklinis šokinėjimas tarp dviejų arba daugiau simpleksų.

Žingsnių kartojimo ciklo nutraukimas

1. Naujojoje viršūnėje funkcijos reikšmė didesnė negu kitose viršūnėse. Taip gali atsitikti, kai minimumo taškas uždengtas simpleksu arba kai simpleksas yra siauro griovio dugne. Tada toliau žengiama iš viršūnės, kuri pagal funkcijos reikšmės dydį yra antra.
2. Ciklinis judėjimas. Jei kuri nors viršūnė neišmetama per daugiau negu M iteracijų, būtina sumažinti simplekso dydį. Naujo simplekso bazinis taškas parenkamas ten, kur senajame simplekse buvo mažiausia funkcijos reikšmė. Kiekviena naujojo simplekso kraštinė sutrumpinama iki K-tosios buvusios kraštinės dalies, K yra simplekso mastelio koeficientas: $\alpha = K\alpha_{senas}$.
3. Patenkinamos sustojimo sąlygos:
    - simpleksas tampa mažas;
    - funkcijos reikšmės viršūnėse tampa panašios;
    - viršijamas leistinas funkcijos reikšmių skaičiavimų kiekis.

Skaičiavimų tipai

- Pagal šį algoritmą atliekami dviejų tipų skaičiavimai.
- Taisyklingo simplekso sudarymas pagal duotą bazinį tašką ir mastelio koeficientą α.
- Naujos simplekso viršūnės skaičiavimas.

Simplekso sudarymas

- Duotas bazinis taškas $X_0 = (x_{01}, \dots, x_{0n})$ ir koeficientas α.
- Taisyklingo simplekso viršūnės skaičiuojamos taip:

$$X_i = \begin{cases} x_{0j} + \delta_1, & \text{jei } j \neq i, \\ x_{0j} + \delta_2, & \text{jei } j = i, \end{cases} \quad i, j = 1, 2, \dots, n;$$

$$\delta_1 = \frac{\sqrt{n+1} + n - 1}{n\sqrt{2}}\alpha, \; \delta_2 = \frac{\sqrt{n+1} - 1}{n\sqrt{2}}\alpha.$$

- Parinkus $\alpha = 1$, simplekso kraštinės ilgis būtų lygus 1.

Naujos viršūnės skaičiavimas

- Jei atmetame X_j, kitų n viršūnių svorio centras lygus

$$X_c = \frac{1}{n} \sum_{i=0, i \neq j}^{n} X_i.$$

- Per taškus X_j ir X_c nubrėžtos tiesės lygtis:

$$X = X_j + \lambda (X_c - X_j).$$

- Jei $\lambda = 0$, tai $X = X_j$.
- Jei $\lambda = 1$, tai $X = X_c$.
- Jei $\lambda = 2$, tai X yra nauja simpleso viršūnė:

$$X_{naujas} = -X_j + 2X_c.$$

Algoritmo savybės

1. Paprastas.
2. Nedaug valdymo parametrų:
   - mastelio koeficientas α;
   - simplekso mažinimo daugiklis K;
   - skaičiavimo pabaigos parametrai.
3. Algoritmas efektyvus net kai tikslo funkcijos reikšmės skaičiuojamos su ganėtinai didelėmis paklaidomis.
4. Reikia nedaug kompiuterio atminties – $n(n+2)$ dydžio masyvo.

Trūkumai

1. Taisyklingas simpleksas nėra gerai tinkantis nesimetriškam tikslo funkcijos hiperpaviršiui.
2. Koeficientas α bendras visoms simplekso kraštinėms. Trūkumas šalinamas taip normuojant kintamuosius, kad suvienodėtų funkcijos kitimas visomis kryptimis. Pvz., jei x_1 – tonos, o x_2 – centimetrai, normuojama taip, kad žengus x_1 ar x_2 kryptimi to paties ilgio žingsniu, funkcija pakistų panašiai.
3. Lėtas, nes neįvertinami ankstesniais žingsniais gauti rezultatai.
4. Gali labai sumažėti α. Pvz., kai simpleksas apeina “kalną” arba praeina siaurą griovį.

Nelder-Mead deformuojamo simplekso metodas

- Siekiant išvengti dalies problemų Nelder ir Mead 1965 m. pasiūlė strategiją su netaisyklingu deformuojamu simpleksu – jis gali ir plėstis, ir trauktis bet kuria kryptimi.
- Tempiama ta kryptimi, kuria vyksta gerėjimas.
- Jei atspindys toks sėkmingas, kad naujoji funkcijos reikšmė geresnė net už geriausią iki tol žinotą, galima daryti išvadą, kad užčiuopta labai gera judėjimo kryptis ir ja žengiama toliau, ištempiant ta kryptimi gautąjį simpleksą.
- Jei atspindys nesėkmingas, tas nebūtinai reiškia, kad kryptis bloga. Gali būti, kad žingsnis per ilgas. Todėl nebūtina atsisakyti šios tikėtinai geros krypties, o pabandyti suspausti simpleksą atlikto žingsnio kryptimi.
- Tiek tempimo, tiek suspaudimo operacijos priklauso nuo empiriškai parinktų parametrų. Patyręs vartotojas gali atsisakyti siūlomų parametrų reikšmių ir jas parinkti pagal savo uždavinio ypatybes.

Simplekso deformavimo koeficientai

- Simpleksui išplėsti – γ ($\gamma > 1$).
- Simpleksui suspausti – β ir ν ($0 < \beta < 1$; $-1 < \nu < 0$).
- Simplekso deformavimo koeficientų reikšmės priklauso nuo sprendžiamų uždavinių. Eksperimentai parodė, kad geri rezultatai gaunami prie $\gamma = 2$, $\beta = 0{,}5$ ir $\nu = -0{,}5$.

Skaičiavimuose naudojamos viršūnės

- X_h, kurioje tikslo funkcijos reikšmė didžiausia (ji keičiama nauja);
- X_g, kurioje tikslo funkcijos reikšmė antra pagal dydį;
- X_l, kurioje tikslo funkcijos reikšmė mažiausia.
- Taisyklingo simplekso algoritme nauja viršūnė skaičiuojama pagal formulę

  $$X_{naujas} = X_h + \lambda(X_c - X_h)$$

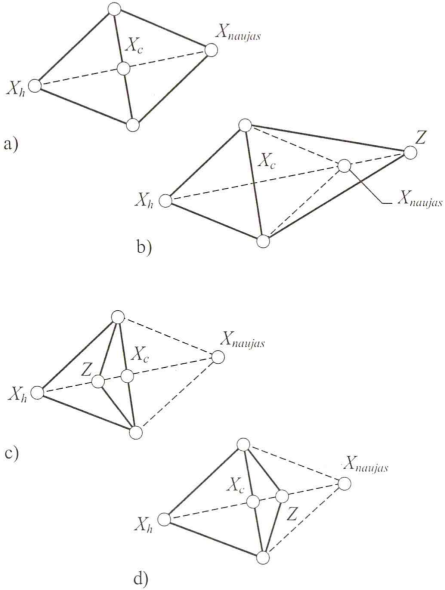
  arba, pažymėjus $\lambda = 1 + \Theta$,

  $$X_{naujas} = X_h + (1 + \Theta)(X_c - X_h).$$

- Kai $\Theta = 1$, gauname ankstesnį atvejį ($\lambda = 2$). Jei $-1 \leq \Theta < 1$, simpleksas spaudžiamas. Jei $\Theta > 1$ – ištempiamas.

Simplekso suspaudimas ir išplėtimas

- Apskaičiuojamas X_c ir X_{naujas}. Pagal $f(X_{naujas})$ nustatomas Θ ir skaičiuojamas naujas taškas Z:
  a) jei $f(X_l) < f(X_{naujas}) < f(X_g)$, tai $\Theta = 1$: dydis nekinta;
  b) jei $f(X_{naujas}) < f(X_l)$, tai $\Theta = \gamma > 1$: išplečiamas;
  c) jei $f(X_{naujas}) > f(X_h)$, tai $\Theta = \nu$: suspaudžiamas;
  d) jei $f(X_g) < f(X_{naujas}) < f(X_h)$, tai $\Theta = \beta$: suspaudžiamas.
- Toliau nustatomi nauji X_h, X_g, X_l ir vykdoma kita iteracija.



Algoritmo variantai

- Aptartas būdas paprastas, bet simpleksas apie minimumo tašką susispaudžia gana lėtai, nes spaudžiama tik $X_c - X_h$ kryptimi.
- Kiti metodo variantai dažniausiai skiriasi suspaudimo būdu, kai $f(X_{naujas}) > f(X_h)$.
- Dažnai kai $f(X_{naujas}) > f(X_h)$, iš geriausios reikšmės taško X_l formuojamas simpleksas su dvigubai trumpesnėmis nei pradinio iteracijos simplekso kraštinėmis.

Pavyzdys

- Rasti funkcijos $f(X) = (1 - x_1)^2 + (2 - x_2)^2$ minimumą.
- Tarkime, $X_0 = (0; 0)$ yra bazinis taškas, o koeficientas $\alpha = 2$, $n = 2$.
- Skaičiuojame kitas dvi pradinio simplekso viršūnes:

$$\delta_1 = \frac{\sqrt{3}+1}{2\sqrt{2}}\alpha = 1{,}9318; \ \delta_2 = \frac{\sqrt{3}-1}{2\sqrt{2}}\alpha = 0{,}5176;$$

  $X_1 = (0 + 0{,}5176; 0 + 1{,}9318) = (0{,}5176; 1{,}9318)$;
  $X_2 = (0 + 1{,}9318; 0 + 0{,}5176) = (1{,}9318; 0{,}5176)$.
- Skaičiuojamos funkcijos reikšmės $f(X_1) = 0{,}2374$; $f(X_2) = 3{,}0658$; $f(X_0) = 5$.

$$X_c = \frac{1}{2}\sum_{i=1}^{2} X_i = \frac{1}{2}(X_1 + X_2).$$

- Skaičiuojame naują tašką X_3 ir funkcijos reikšmę:

$$X_3 = X_1 + X_2 - X_0 = (2{,}4494; 2{,}4494),\ f(X_3) = 2{,}3027.$$

Hooke-Jeeves metodas

- Hooke-Jeeves metodu vykdoma paieška ortogonaliomis kryptimis.
- Metodas turi pranašumų tiek prieš simplekso metodą, tiek prieš koordinatinę paiešką.
- Simplekso metode kiekviename žingsnyje turime $n - 1$ simplekso viršūnę ir n galimų naujų judėjimo krypčių, bet tos kryptys neortogonalios.
- Kitas paieškos krypčių sistemos sudarymo būdas – sukurti ortogonalių krypčių sistemą kaip koordinatinės paieškos metodu.
- Hooke-Jeeves metode naudojamos papildomos kryptys, todėl jis efektyvesnis negu koordinatinė paieška.

Metodo idėja

- Paieška vykdoma kryptimi $d_i = X_i - X_{i-1}$.
- Tai greičiau lyginant su koordinatine paieška, nes taip sudarytos kryptys turi tendenciją išsidėstyti išilgai tikslo funkcijos paviršiaus griovio, jei toks yra.
- Svarbiausia – įvertinami ankstesnėse iteracijos gauti rezultatai.
- Metodas susideda iš dviejų dalių:
  - tiriamosios paieškos su cikliniu kintamųjų keitimu ir
  - greitėjančios paieškos pagal pavyzdį.

Tiriamoji paieška

- Tiriamoji paieška turi nustatyti lokalias tikslo funkcijos savybes ir rasti kryptį išilgai griovio.
- Tiriamosios paieškos taško $X_0 = (x_{01}, x_{02}, \ldots, x_{0n})$ aplinkoje algoritmas su prieaugiais Δ_i:

1. Startuojame iš bazinio taško X_0 ir formuojame naują tašką X_1, pradžioje priskiriame $X_1 = X_0$.
2. Pagalbinis taškas $\tilde{X} = X_1$.
3. Kiekvienai koordinatei i:
   - Keičiame taško $\tilde{X}$ koordinatę $\tilde{x}_i = x_{1i} + \Delta_i$ ir skaičiuojame $f(\tilde{X})$.
   - Jei $f(\tilde{X}) < f(X_1)$, tai $X_1 = \tilde{X}$.
   - Priešingu atveju $\tilde{x}_i = x_{1i} - \Delta_i$ ir skaičiuojame $f(\tilde{X})$; jei $f(\tilde{X}) < f(X_1)$, tai $X_1 = \tilde{X}$.

- Jei tiriamosios paieškos algoritmu gautas taškas X_1 nesutampa su tašku X_0, tai X_1 tampa nauju baziniu tašku.

Paieška pagal pavyzdį

- Einamasis bazinis taškas X_k,
- Ankstesnis bazinis taškas X_{k-1},
- Pagal pavyzdį judant gautas taškas $\bar{X}_{k+1}$,
- Naujas bazinis taškas X_{k+1}.
- Žengiama iš bazinio taško X_k išilgai tiesės, jungiančios šį tašką su ankstesniu baziniu tašku X_{k-1}:

$$\bar{X}_{k+1} = X_k + (X_k - X_{k-1}).$$

Hooke-Jeeves algoritmas

1. Nustatome algoritmo parametrus: pradinį bazinį tašką X_0, prieaugius Δ_i, žingsnio mažinimo koeficientą $\alpha > 0$, paieškos pabaigos parametrą $\varepsilon > 0$.
2. Atliekame tiriamąją paiešką iš bazinio taško.
3. Jei tiriamoji paieška sėkminga – radome tašką su mažesne tikslo funkcijos reikšme – eiti į 5 žingsnį, jei ne – į 4 žingsnį.
4. Tikriname pabaigos sąlygą – ar $||\Delta|| < \varepsilon$. Jei taip – baigti, jei ne – sumažinti prieaugius $\Delta_i = \Delta_i/\alpha$ ir eiti į 2 žingsnį.
5. Atliekame paiešką pagal pavyzdį: $\bar{X}_{k+1} = X_k + (X_k - X_{k-1})$.
6. Atliekame tiriamąją paiešką, kaip bazinį tašką naudodami $\bar{X}_{k+1}$, rezultatą pažymėkime X_{k+1}.
7. Ar $f(X_{k+1}) < f(X_k)$? Jei taip, $X_{k-1} = X_k$, $X_k = X_{k+1}$ ir eiti į 5 žingsnį. Ne – eiti į 4 žingsnį su baziniu tašku X_k.

Pavyzdys

- Rasti funkcijos $f(X) = 8x_1^2 + 4x_1x_2 + 5x_2^2$ minimumą.
- Pradinis bazinis taškas $X_0 = (-4; -4)$. $\Delta = (1; 1)$; $\alpha = 2$; $\varepsilon = 10^{-4}$; $f(X_0) = 272$.
- Tiriamoji paieška X_0 aplinkoje:
  - $X_1 = X_0 = (-4; -4)$.
  - $\tilde{X} = (-4 + 1; -4) = (-3; -4)$, $f(\tilde{X}) = 200 < 272$: $X_1 = \tilde{X} = (-3; -4)$.
  - $\tilde{X} = (-3; -4 + 1) = (-3; -3)$, $f(\tilde{X}) = 153 < 200$: $X_1 = \tilde{X} = (-3; -3)$.
- Kadangi tiriamoji paieška sėkminga, toliau vykdome paiešką pagal pavyzdį:

$$\bar{X}_2 = X_1 + (X_1 - X_0) = (-2; -2); \;\; f(\bar{X}_2) = 68.$$

- Tiriamoji paieška $\bar{X}_2$ aplinkoje. Tiriamosios paieškos žingsniai sėkmingi. Atlikus skaičiavimus gauname

$$X_2 = (-1; -1); \;\; f(X_2) = 17.$$

- Kadangi $f(X_2) < f(X_1)$, tai X_2 tampa nauju baziniu tašku.

Atsitiktinės paieškos metodai

- Vienas iš paprasčiausių metodų yra atsitiktinis taškų generavimas su tolygiu skirstiniu leistinojoje aibėje.
- Sugeneruotuose taškuose apskaičiuojamos tikslo funkcijos reikšmės ir išrenkamas taškas su mažiausia tikslo funkcijos reikšme.
- Ypač lengvai metodas realizuojamas, kai leistinoji sritis yra hiperstačiakampis.
- Metodo konvergavimas (tikimybine prasme) įrodomas esant minimalioms prielaidoms apie tikslo funkciją.
- Tačiau metodo teorinis konvergavimo greitis mažas, o praktiškai metodo efektyvumas dažniausiai nepakankamas.
- Tas ir natūralu, kadangi metodas pasyvus, t.y. einamasis taškas parenkamas nepriklausomai nuo anksčiau apskaičiuotų tikslo funkcijos reikšmių.
- Tolygi atsitiktinė paieška dažniausiai naudojama kombinuojant ją su kitais metodais.

Paieška atsitiktine kryptimi

- Tam kad nustatytume greičiausio funkcijos kitimo kryptį, turime apskaičiuoti funkcijos gradientą. Tas reikalauja tam tikrų skaičiavimo laiko sąnaudų. Kartais tokios sąnaudos gali neatsipirkti ta prasme, kad žingsnis "pigiau" nustatyta, nors ir "prastesne", kryptimi gali būti "naudingesnis".
- Pasirinkę tam tikrą žingsnio ilgį ρ, atsitiktinai su tolygiu skirstiniu ant sferos $S(X_i,\rho)$ paviršiaus generuojame tašką, ir jame apskaičiuojame tikslo funkcijos reikšmę.
- Jei gautoji funkcijos reikšmė mažesnė, naujas taškas X_{i+1}.
- Jei pagerėjimo nėra, $X_{i+1} = X_i$.
- Tikimybė gauti geresnę funkcijos reikšmę yra apie 0,5, kai funkcija nagrinėjamojoje srityje $S(X_i,\rho)$ artima tiesinei. Tokia prielaida priimtina, kai ρ pakankamai mažas ir X_i pakankami toli nuo minimumo taško.
- Buvo pasiūlytos įvairios metodo modifikacijos, kuriomis buvo siekiama padidinti sėkmingo žingsnio tikimybę esant kitokioms prielaidoms apie funkciją, pvz., kai einamasis taškas yra "griovio" dugne.

Stochastinio gradiento metodas

- Paminėsime vieną iš pastarojo metodo modifikacijų, kuri panaši į gradientinį nusileidimą tiek, kad net vadinama stochastinio gradiento metodu.
- Generuokime ant sferos ne vieną, o N taškų.
- Visuose sugeneruotuose taškuose apskaičiuokime tikslo funkcijos reikšmes ir išrinkime geriausią.
- Geriausias taškas tampa X_{i+1}.
- Kai sferos radiusas pakankamai mažas, o N pakankamai didelis, $X_{i+1} - X_i$ kryptis yra artima tikslo funkcijos antigradiento taške X_i krypčiai.

Paieškos metodų savybės

- Paieškos metodai pasižymi nelokaliu ieškojimo būdu, t.y. jų randamas sprendinys yra nebūtinai artimiausias pradiniam taškui lokalusis minimumas.
- Šiuo požiūriu paieškos metodai skiriasi nuo nusileidimo metodų.
- Tačiau ne visi paieškos metodai orientuoti globalaus minimumo paieškai. Tam skirta ištisa globaliojo optimizavimo metodų klasė.