

4. ТУЫНДЫ ҚҰРАЛДАРДЫ ҚОЛДАНУ ӘДІСТЕРІ

Түсіру әдістері және олардың әртүрлі модификациялары, функцияның мәндерін ғана қолданатын кездейсоқ іздеу әдістері (5-тармақ) 0-ші ретті әдістер деп аталады. Олар әдетте өте төмен конвергенция жылдамдығына ие. Сондықтан мақсатты функцияның бірінші және екінші туындыларын қолданатын бірқатар оңтайландыру әдістері жасалды (1-ші және 2-ші ретті әдістер).

Мұндай әдістерді қарастырмас бұрын бірқатар белгілерді енгізіп, кейбір анықтамаларды еске түсіріңіз.

n өлшемді кеңістіктің векторын бағанмен белгілейміз (R^n):

$$u = \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix}; \text{ тоғда } u^T = [x_1, \dots, x_n].$$

Егер $\frac{df(x)}{dx_1}$ туындылары болса, $f: R^n \rightarrow R$ функциясы $x \in R^n$ нүктесінде үздіксіз сараланады деп айтамыз $i = 1, \dots, n$ бар және үздіксіз. Содан кейін x нүктесіндегі f функциясының градиенті келесідей анықталады:

$$\nabla f(x) = \left[\frac{df(x)}{dx_1}, \dots, \frac{df(x)}{dx_n} \right]$$

$f(x)$ функциясы $D \subset R^n$ ашық аймағында үздіксіз сараланады, егер ол D -ден кез-келген нүктеде үздіксіз сараланса.

$f: R^n \rightarrow R$ қатардың кейбір ашық дөңес $D \subset R^n$ аймағында үздіксіз $x \in D$ дифференциалдансын. Содан кейін $x \in D$ жәнер $p \in R^n$ ерікті нөлдік емес түзету үшін туынды сызық 1-ші бағытта болады.... $p = [p_1, \dots, p_n]$ функциясы x нүктесінде $f(x)$, ретінде айқындалатын:

$$\frac{df(x)}{dp} \equiv \lim_{\varepsilon \rightarrow 0} \left(\frac{f(x+\varepsilon p) - f(x)}{\varepsilon} \right)$$

$\nabla f(x)^T * p$ қосылыстарының жиынтығы бар және оған тең, мұнда скалярлық өндіріс $\bullet$ белгісімен белгіленеді.

$$\frac{df(x)}{dp} = \nabla f(x)^T * p = \sum_{i=1}^n \frac{df(x)}{dx_i} p_i$$
$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j}, 1 \leq i, j \leq n$$

Егер туындылары бар және үздіксіз болса, $f: R^n \rightarrow R$ функциясы $x \in R^n$ -ге екі рет үздіксіз бөлінеді деп айтамыз.

x нүктесіндегі f функциясының гессиан (Гесс матрицасы) $n \times n$ өлшемінің матрицасы деп аталады және оның (i,j) -ші элементі

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

$f: R^n \rightarrow R$ $D \subset R^n$ ашық аймағында екі рет үздіксіз дифференциалдансын.

Содан кейін кез-келген $x \in D$ және еркін нөлдік емес өсу үшін $p \in R^n$ F функциясының x нүктесіндегі p бағыты бойынша екінші туынды ретінде анықталады

$$\frac{\partial^2 f(x)}{\partial p^2} \equiv \lim_{\varepsilon \rightarrow 0} \left(\frac{\frac{\partial f}{\partial p}(x + \varepsilon p) - \frac{\partial f}{\partial p}(x)}{\varepsilon} \right)^T$$

ол үшін теңдік бар және орындалады:

$$\frac{\partial^2 f(x)}{\partial p^2} = p^T \nabla^2 f(x) p$$

A -нақты симметриялы матрица $n \times n$ өлшемі болсын. Егер $u \in R^n$ нөлдік емес векторы үшін теңсіздік орындалса, A оң анықталады деп айтамыз $u^T A u > 0$.

Егер барлық $u \in R^n$ үшін $u^T A u \geq 0$ болса, A матрицасы оң анықталады.

* x нүктесі жергілікті минималды нүкте $f(x)$ болуы үшін $\nabla f(x^*) = 0$ теңдігін орындау қажет. Сонымен қатар, жеткілікті шарт оң сенімділікті $\nabla^2 f(x^*)$ талап етеді, ал қажетті шарт – кем дегенде оң жартылай анықталу $\nabla^2 f(x^*)$.

Әрі қарай, біз $\nabla f(x)$, $\nabla^2 f(x)$ бар және үздіксіз деп санаймыз.

Төменде сипатталған барлық әдістер формулаға сәйкес жүзеге асырылатын итерациялық процедураға негізделген

$$x^{k+1} = x^k + \lambda^k s(x^k)$$

мұндағы x^k -шешімге ағымдағы жуықтау * x , λ^k -қадам ұзындығын сипаттайтын параметр, S^k – n өлшемді кеңістіктегі іздеу бағыты.

Алғашқы туындыларды қолданатын бірінші ретті әдісті қарастырыңыз.

4.1. Градиент әдістері

Кез келген x нүктесіндегі функцияның градиенті ең үлкен жергілікті үлкейту $f(x)$ бағытын көрсетеді. Сондықтан, минимумды іздегенде, сіз осы нүктеде градиентке қарама-қарсы бағытта, яғни ең жақын түсу бағытында қозғала аласыз. Бұл тәсіл градиентті түсіру әдісін сипаттайтын итерациялық формулаға әкеледі:

$$x^{k+1} = x^k - \lambda^k \nabla f(x^k) \text{ немесе}$$

$$x^{k+1} = x^k - \lambda^k \frac{\nabla f(x^k)}{\|\nabla f(x^k)\|} = x^k - \lambda^k S^k$$

$\|\nabla f(x^k)\|$ -градиент нормасы, S^k -бірлік векторы.

(u векторының нормасы ретінде сіз $\|u\| = \sqrt{u_1^2 + \dots + u_n^2}$ Гаусс нормасын таңдай аласыз.)

λ параметрін таңдауға байланысты түсу траекториясы айтарлықтай өзгереді.

λ мәні үлкен болған кезде траектория тербелмелі процесс болады, ал λ тым үлкен болған кезде процесс ауытқуы мүмкін.

Кішкентай λ тегіс болады, бірақ процесс баяу айналады.

λ^k параметрін тұрақты қабылдауға немесе әр итерацияда әр түрлі таңдауға болады. Кейде әр k -ші қадамда λ^k параметрі кез-келген бір өлшемді әдісті қолдана отырып, s^k бағыты бойынша бір өлшемді минимизация жасау арқылы таңдалады. Әдетте бұл процесс ең жақын түсу әдісі немесе Коши әдісі деп аталады.

Егер λ^k бір өлшемді минимизация нәтижесінде анықталса, яғни, $\lambda^k = \arg \min_{\lambda} f(x^k + \lambda s^k)$ онда келесі жақындау нүктесіндегі градиент алдыңғы түсу бағытына ортогоналды болады (сурет.19).

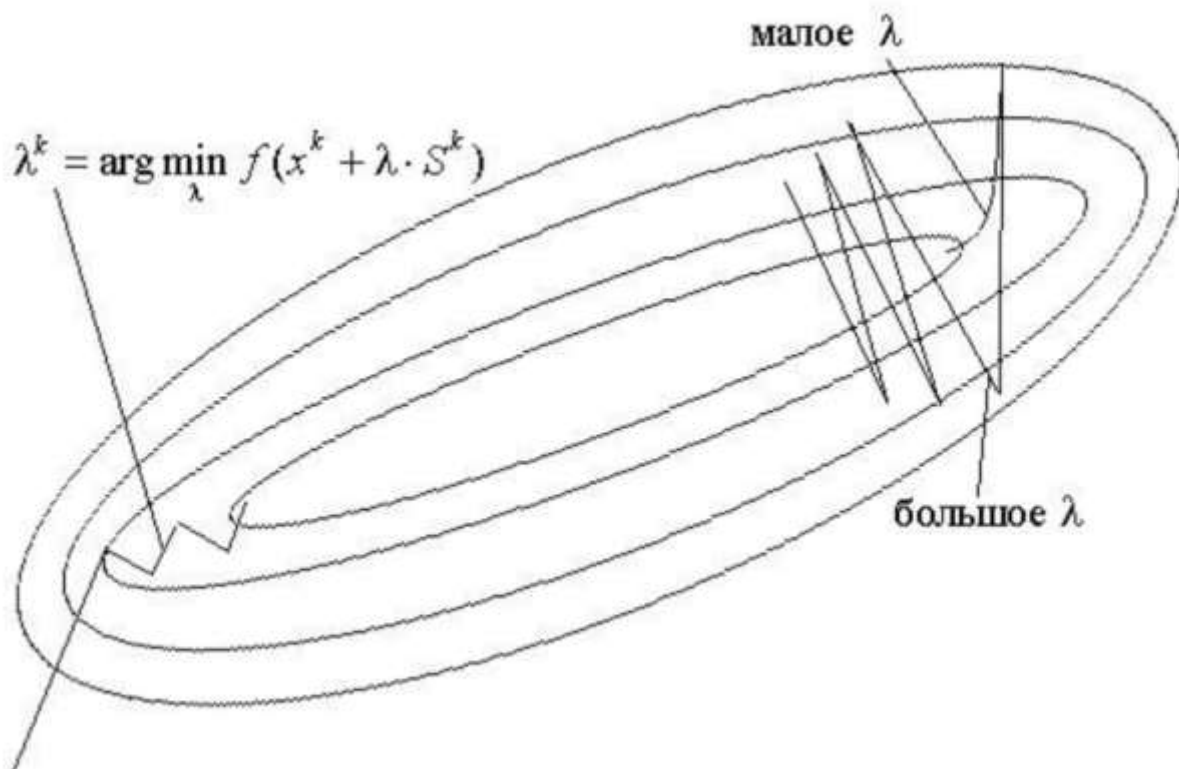


Рис. 19. Иллюстрация к градиентным методам

s^k бағыты бойынша бір өлшемді оңтайландыру әдістің конвергенциясын жақсартады, бірақ сонымен бірге әр итерацияда $f(x)$ функциясын есептеу саны артады. Сонымен қатар, экстремумға жақын жерде $\|\nabla f(x)\|$ нормасы нөлге жақын және конвергенция өте баяу болады.

Алгоритмнің тиімділігі азайтылатын функцияның түріне байланысты. Сонымен, $f(x) = x_1^2 + x_2^2$ функциясы үшін Коши әдісі кез-келген бастапқы жақындау кезінде кем дегенде бір итерацияға дейін азаяды, ал

$f(x) = x_1^2 + 100x_2^2$ функция үшін конвергенция өте баяу болады.

Жалпы, бұл әдістің шатқал рельефтеріндегі тиімділігі өте нашар.

Бұл әдіс өте сирек қолданылады.

4.2. Ньютон Әдісі

x^k нүктесінің төңірегіндегі функцияның квадраттық моделін құрамыз, оны Тейлор қатарына екінші ретті мүшелерге таратамыз:

$$f(x^k + p) = f(x^k) + \nabla f^T(x^k)p + \frac{1}{2}p^T \nabla^2 f(x^k)p$$

Квадраттық модельдің минималды жағдайынан $x^{k+1} = x^k + s^k$ нүктесін табыңыз (4.5). Бұл минимумның қажетті шарты $\nabla f(x^{k+1}) = 0$ болады.

Иеміз:

$$\nabla f(x^k + s^k) = 0 = \nabla f(x^k) + \nabla^2 f(x^k)s^k$$

бұл келесі алгоритмге әкеледі:

әрбір k итерациясында теңдеулер жүйесін шешіңіз

$$\nabla^2 f(x^k)s^k = -\nabla f(x^k)$$

қатысты s^k және қою

$$x^{k+1} = x^k + s^k$$

Алгоритм (4.6), (4.7) Ньютон әдісі деп аталады. Ньютонның оң жүз-
рондық әдісі, егер x^0 жергілікті минимумның x^* нүктесіне жақын болса, $f(x)$
функциясының Гесс дегенеративті матрицасы бар болса, онда дәлелдеуге
болады $\nabla^2 f(x^*)$ оң анықталған және алгоритмнен кейін пайда болған $\{x^k\}$
(4.6), (4.7) минимумға жақындайды x^* (және конвергенция жылдамдығы q-
квадраттық деп аталады).

Кемшіліктеріне әдісіне жатады:

1. Бұл әдіс жаһандық деңгейде жинақталмайды, яғни x^0 бастапқы жуықтауды қажет етеді;
2. Бірінші және екінші туындылардың аналитикалық тапсырмасы қажет;
3. Әр қадамда теңдеулер жүйесін шешу қажет (4.7);
4. Ньютон әдісінде оны максимумға немесе ер-тоқымға қарай жылжытудан сақтайтын ештеңе жоқ, мұнда ∇f нөлге тең.

Мұнда әр қадам жергілікті квадраттық модельдің стационарлық нүктесіне бағытталған, стационарлық нүкте минималды, максималды немесе ер-тоқымға қарамастан. Бұл қадам гессиан $\nabla^2 f(x^k)$ оң анықталған кезде ғана азайтуға негізделген.

Жалпы алғанда, Ньютон әдісі жоғары сенімділікке ие емес.

4.3. Марквардт Әдісі

Бұл әдіс градиентті түсіру әдістері мен Ньютонның үйлесімі болып табылады, ол екі әдістің де оң қасиеттерін сәтті үйлестіреді. Минимум x^*

нүктесінен едәуір қашықтықта орналасқан x^0 нүктесінен антиградиент бағытында қозғалыс, әдетте, мақсатты функцияның айтарлықтай төмендеуіне әкеледі. Екінші жағынан, ең төменгі нүктенің төңірегіндегі тиімді іздеу бағыты Ньютон әдісімен анықталады.

Марквардт әдісіне сәйкес s^k іздеу бағыты теңдікпен анықталады

$$(H^k + \lambda^k I)s^k = -\nabla f(x^k), \quad (4.8)$$

ал жаңа x^{k+1} нүктесі формуламен берілген

$$x^{k+1} = x^k + s^k. \quad (4.9)$$

Жүйеде (1) I–бірлік матрица, H^k -Гесс матрицасы, $\lambda^k \geq 0$.

Соңғы формулада s^k алдындағы коэффициент 1-ге тең алынады, өйткені λ^k в (4.8) параметрлері қадамның ұзындығын да, бағытын да өзгертуге мүмкіндік береді.

Іздеудің бастапқы кезеңінде λ^0 параметрі 10^4 сияқты үлкен мәнге ие болады, сондықтан марквардтың теңдігінің сол жағы (4.8) үлкен λ^0 түрінде болады

$$(H^0 + \lambda^0 I)s^k \approx (\lambda^0 I)s^k = \lambda^0 s^k. \quad (4.10)$$

Сонымен, λ^0 -ден (4.8) және (4.10) үлкен мән іздеу бағытына сәйкес келеді

$$s^k = -\frac{1}{\lambda^0} \nabla f(x^k),$$

яғни, ең жақын түсу бағыты.

(4.8) формуладан λ^k нөлге дейін төмендеген кезде s^k векторы градиентке қарама-қарсы бағыттан Ньютон анықтаған бағытқа ауысады деп қорытынды жасауға болады. Егер бірінші қадамнан кейін мақсатты функцияның төменгі мәні бар нүкте алынса, яғни $f(x^1) < f(x^0)$ болса, $\lambda^1 < \lambda^0$ таңдап, тағы бір қадамды орындау керек. Әйтпесе, сіз $\lambda^0 = \beta \lambda^0$ -ді $\beta - 1$ -ге қойып, алдыңғы қадамды қайта орындауыңыз керек.

Ньютон әдісінің кемшілігі, егер Гесс H^k матрицасы оң анықталмаса, онда s^k ньютондық қадамы функцияның төмендеуіне әкелмейді. Сондықтан Гессианның "түзету" формуласы бойынша $H^k + \lambda^k I$ матрицаны өзгертеді және тиісті λ^k таңдау кезінде оны оң анықтайды, өйткені бірлік матрицасы оң анықталады.

Енді әдіс алгоритмін келтірейік:

1. x^0 орнатыңыз - *х-ке бастапқы жуықтау, М-итерациялардың максималды рұқсат етілген саны және ε -конвергенция параметрі.

2. Қою $k = 0, \lambda^0 = 10^4$.

3. Есептеу $\nabla f(x^k)$.

4. Тексеру $\|\nabla f(x^k)\| < \varepsilon$?

$$(\text{алуға болады } \|\nabla f(x^k)\| = \sqrt{\left(\frac{\partial f(x^k)}{\partial x_1}\right)^2 + \dots + \left(\frac{\partial f(x^k)}{\partial x_n}\right)^2})$$

Егер иә болса, онда көшу п. 11.

5. Тексеру $k \geq M$? Егер иә болса, онда көшу п. 11.

6. Жүйені шешу арқылы s^k қадамын есептеңіз

$$(H^k + \lambda^k I)s^k = -\nabla f(x^k)$$

7. Қою $x^{k+1} = x^k + s^k$.

8. Тексеру $f(x^{k+1}) < f(x^k)$?

Егер иә болса, онда көшу п. 9, басқаша т. 10.

9. Қою $\lambda^{k+1} = \frac{1}{2}\lambda^k, k = k + 1$. П. 3 көшу.

10. Қою $\lambda^k = 2\lambda^k$. П. 7 көшу.

11. Нәтижелерді шығару:

$$x^k, f(x^k), \nabla f(x^k), \|\nabla f(x^k)\|, k.$$

Ньютон әдісінің әртүрлі модификацияларында көптеген есептеулер қажет екенін ескеріңіз, өйткені әр итерацияда алдымен $n \times n$ матрицасының элементтерін есептеп, содан кейін сызықтық теңдеулер жүйесін шешу керек. Әрине, бірінші және екінші туындылардың айырмашылықтарын қолдану жағдайды нашарлатады.

Сондықтан, жақында Гесс матрицасының градиентіи аналитикалық есептеуімен де, олардың әрине айырмашылыққа жақындауымен де квазиньютондық деп аталатын көптеген әдістер жасалды. Бұл әдістер сызықтық емес объективті функцияның қисықтығын оның Гесс матрицасын айқын қалыптастырусыз жақындату мүмкіндігіне негізделген. Қисықтық туралы деректер түсу кезінде градиенттің өзгеруін бақылау негізінде жинақталады.

Мұндай әдістердің әртүрлі формалары, көбінесе секциялық әдістер деп аталады, ғылыми және техникалық зерттеулерде жақсы нәтиже көрсетті.