18. Безусловная многомерная оптимизация . Метод Ньютона для безусловной оптимизации. Соотношение секущих для безусловной оптимизации. Формулы PSB, BFGS и DFP.

Методы Ньютона и секущих для безусловной оптимизации. Методы Ньютона и секущих можно применять и для оптимизации, использовав необходимые условия существования локального экстремума (как нелинейное уравнение). Это равносильно использованию для отыскания оптимума f(x) вместо самой функции её квадратичной модели.

$$f(x) \approx m_q(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2} f''(x^{(k)})(x - x^{(k)})^2$$

Тогда в одномерном случае метод Ньютона базируется на выражении:

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

(25) А метод секущих:

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})(x^{(k)} - x^{(k-1)})}{f'(x^{(k)}) - f'(x^{(k-1)})}$$

(26)

В многомерном случае применение методов Ньютона и секущих основано на решении системы нелинейных уравнений (СНУ):

$\nabla^T f(x) = 0,\ x \in R^n,\ f(x) \in R^1,\ \nabla f(x) \in R^n$

Предпосылками методов Ньютона и секущих может служить квадратичная модель в окрестности текущей точки:

$$Q(x) = f(x^{(k)}) + \nabla^T f(x^{(k)})(x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T H^{(k)}(x - x^{(k)}) \quad (27)$$

Дополнительное условие приводят к методам Ньютона или секущих.Во всех случаях квадратичная модель приводит к линейной модели для градиента ∇f(x).

$$\nabla^T Q(x) = \nabla^T f(x^{(k)}) + H^{(k)}(x - x^{(k)}) \approx \nabla^T f(x) = 0 \quad (28)$$

Формула (28) аналогична (11) при $F(x^{(k)}) = \nabla^T f(x^{(k)}),\ A^{(k)} = H^{(k)}$, следовательно, направление движения из $x^{(k)}$ в $x^{(k+1)}$ находим из уравнения:

$H^{(k)}S^{(k)} = -\nabla^T f(x^{(k)})$ (29) Метод Ньютона для оптимизации может быть применен, когда в текущей точке известна матрица Гессе: $\nabla^2 f(x^{(k)}) = G_f(x^{(k)})$

В этом случае квадратичное приближение строится как тейлоровское, т.е. для (29) можно использовать $H^{(k)} = \nabla^2 f(x^{(k)})$ и модель с такой матрицей удовлетворяет дополнительному условию: $\nabla^2 Q(x^{(k)}) = \nabla^2 f(x^{(k)})$

Поэтому направление ищется из условия: $\nabla^2 f(x^{(k)})S^{(k)} = -\nabla^T f(x^{(k)})$ (30)

Матрица Гессе обладает двумя важными свойствами: Симметричность ,Положительно определенность

Из этих предположений следует, что $\det \nabla^2 f(x^{(k)}) \neq 0$

$S^{(k)} = -G_f^{-1}\nabla^T f(x^{(k)})$ (31) $S^{(k)}$ – называется ньютоновским направлением поиска.

Предположим, что матрица Гессе неизвестна, тогда $H^{(k)}$ можно искать из условия линейной интерполяции для квадратичной модели.

$H^{(k)} : \nabla^T Q(x^{(k-1)}) = \nabla^T f(x^{(k-1)})$

$\nabla^T f(x^{(k)}) + H^{(k)}(x^{(k-1)} - x^{(k)}) = \nabla^T f(x^{(k-1)})$

Если $S^{(k-1)} = x^{(k-1)} - x^{(k)}$ и $g^{(k-1)} = \nabla^T f(x^{(k)}) - \nabla^T f(x^{(k-1)})$, тогда соотношение секущих (квазиньютоновское условие) для оптимизации имеет вид:

$H^{(k)}S^{(k-1)} = g^{(k-1)}$

$H^{(k+1)}S^{(k)} = g^{(k)}$

PSB $S^{(k)} = -G_f^{-1}\nabla^T f(x^{(k)})$ (31)

$S^{(k)}$ – называется ньютоновским направлением поиска.

Предположим, что матрица Гессе неизвестна, тогда $H^{(k)}$ можно искать из условия линейной интерполяции для квадратичной модели.

$H^{(k)} : \nabla^T Q(x^{(k-1)}) = \nabla^T f(x^{(k-1)})$

$\nabla^T f(x^{(k)}) + H^{(k)}(x^{(k-1)} - x^{(k)}) = \nabla^T f(x^{(k-1)})$

Если $S^{(k-1)} = x^{(k-1)} - x^{(k)}$ и $g^{(k-1)} = \nabla^T f(x^{(k)}) - \nabla^T f(x^{(k-1)})$, тогда **соотношение секущих** (квазиньютоновское условие) для оптимизации имеет вид:

$H^{(k)}S^{(k-1)} = g^{(k-1)}$

$H^{(k+1)}S^{(k)} = g^{(k)}$

Матрицы $H^{(k)}$, $H^{(k+1)}$… не гарантируют симметричность и положительно-определенность.

Легко проверить, что если $H^{(k)T} = H^{(k)}$, то (34) не дает гарантии, что $H^{(k+1)} = H^{(k+1)T}$.

Общеизвестный **метод симметризации матрицы** заключается в следующем:

$$(H^{(k+1)})_2 = \frac{1}{2}[(H^{(k+1)})_1^T + (H^{(k+1)})_1]$$

Рассмотрим продолжение процесса, генерируя матрицы $(H^{(k+1)})_3, (H^{(k+1)})_4 \cdots$ по формулам:

$$\begin{cases} (H^{(k+1)})_{2m+1} = (H^{(k+1)})_{2m} + \dfrac{[g^{(k)} - (H^{(k+1)})_{2m}S^{(k)}]}{S^{(k)T}S^{(k)}} S^{(k)T} \\ (H^{(k+1)})_{2m+2} = \dfrac{1}{2}[(H^{(k+1)})_{2m+1} + (H^{(k+1)})_{2m+1}^T] \end{cases}$$

Здесь для каждой $(H^{(k+1)})_{2m+1}$ выполняется соотношение секущих, а для $(H^{(k+1)})_{2m+2}$ - симметричность матрицы. В пределе эти матрицы сходятся к одной формуле.

Доказано, что эта последовательность матриц сходится к формуле:

$\mathrm{H}^{(k+1)} = H^{(k)} + \frac{(g^{(k)} - H^{(k)}S^{(k)})S^{(k)T} + S^{(k)}(g^{(k)} - H^{(k)}S^{(k)})^T}{S^{(k)T}S^{(k)}} - \frac{(g^{(k)} - H^{(k)}S^{(k)})^T S^{(k)}S^{(k)}S^{(k)T}}{(S^{(k)T}S^{(k)})^2}$ (35) известна как симметричная формула пересчета матриц **Пауэлла-Бройдена**. В англоязычной литературе называется PSB. (35) не гарантирует положительно-определенность матрицы, что на практике отрицательно сказывается на работоспособности алгоритма.

BFGS $S^{(k)} = -G_f^{-1}\nabla^T f(x^{(k)})$ (31)

$S^{(k)}$ – называется ньютоновским направлением поиска.
$H^{(k)}$ можно искать из условия линейной интерполяции

$\mathrm{H}^{(k)} : \nabla^T Q(x^{(k-1)}) = \nabla^T f(x^{(k-1)})$

$\nabla^T f(x^{(k)}) + \mathrm{H}^{(k)}(x^{(k-1)} - x^{(k)}) = \nabla^T f(x^{(k-1)})$

Если $\mathrm{S}^{(k-1)} = x^{(k-1)} - x^{(k)}$ и $g^{(k-1)} = \nabla^T f(x^{(k)}) - \nabla^T f(x^{(k-1)})$, тогда **соотношение секущих** (квазиньютоновское условие) для оптимизации имеет вид:

$H^{(k)}S^{(k-1)} = g^{(k-1)}$

$H^{(k+1)}S^{(k)} = g^{(k)}$

Матрицы $\mathrm{H}^{(k)}$, $\mathrm{H}^{(k+1)}$… не гарантируют симметричность и положительно-определенность. Для обеспечения симметричности и положительно-определенности матрицы можно воспользоваться следующим алгебраическим фактом:
матрица H симметрична и положительно-определенна тогда и только тогда, когда ее можно представить в виде:
$H = LL^T$, где $\mathrm{L} \in \mathrm{R}^{n\times n}$ L – некоторая невырожденная матрица (представление Холесского). Предположим, что в текущей точке имеет место соотношение:

$H^{(k)} = L^{(k)}L^{(k)T}$

Тогда возникает задача о получении $\mathrm{H}^{(k+1)}$: $H^{(k+1)} = L^{(k+1)}L^{(k+1)T}$, причем должно выполнятся квазиньютоновское условие: $L^{(k+1)}L^{(k+1)T}S^{(k)} = g^{(k)}$ От матрицы $\mathrm{L}^{(\kappa+1)}$ не требуется симметричность и положительно-определенность, но доказано, что такая $\mathrm{L}^{(\kappa+1)}$ существует тогда и только тогда, когда выполняется условие:
$S^{(k)T}g^{(k)} > 0$ Обозначим: $1.\, \upsilon^{(k)} = L^{(k+1)T}S^{(k)}$, тогда $2.\, L^{(k+1)}\upsilon^{(k)} = g^{(k)}$. Для решения подзадачи 2 можно воспользоваться формулой Бройдена:

$L^{(k+1)} = L^{(k)} + (g^{(k)} - L^{(k)}\upsilon^{(k)})\upsilon^{(k)+}$

Для решения подзадачи 1 запишем:

$\upsilon^{(k)} = L^{(k)T}S^{(k)} + \upsilon^{(k)}\frac{(g^{(k)T} - \upsilon^{(k)T}L^{(k)T})}{\upsilon^{(k)T}\upsilon^{(k)}}S^{(k)}$ (36) Обозначим $\alpha^{(k)} = (1-\gamma^{(k)})^{-1},\ \alpha^{(k)}$ и $\gamma^{(k)} - \textit{скалярные}$.

$(1-\gamma^{(k)})\upsilon^{(k)} = L^{(k)T}S^{(k)}$

$\upsilon^{(k)} = \alpha^{(k)}L^{(k)T}S^{(k)}$ (37)

Подставим (37) в (36) и после некоторого упрощения получим:

$\alpha^{(k)2} = \frac{g^{(k)T}S^{(k)}}{S^{(k)T}L^{(k)}L^{(k)T}S^{(k)}} = \frac{g^{(k)T}S^{(k)}}{S^{(k)T}H^{(k)}S^{(k)}}$

Данное выражение корректно, т.к. предполагается, что:
$S^{(k)T}g^{(k)} = g^{(k)T}S^{(k)} > 0$, а квадратичная форма $S^{(k)T}H^{(k)}S^{(k)} > 0$ для положительно-определенной матрицы, следовательно $\alpha^{(k)2} > 0$.
Выбираем положительные значения квадратного корня:

$\alpha^{(k)} = \left(\frac{g^{(k)T}S^{(k)}}{S^{(k)T}H^{(k)}S^{(k)}}\right)^{\frac{1}{2}}$ (38)

Тогда из (37) получим: $\upsilon^{(k)} = \left(\frac{g^{(k)T}S^{(k)}}{S^{(k)T}H^{(k)}S^{(k)}}\right)^{\frac{1}{2}}\mathrm{L}^{(k)}S^{(k)}$ (39)

Подставим (39) в формулу Бройдена для $\mathrm{L}^{(k+1)}$ и учтя

$\begin{cases} H^{(k)} = L^{(k)}L^{(k)T} \\ H^{(k+1)} = L^{(k+1)}L^{(k+1)T} \end{cases}$ получим: $\mathrm{H}^{(k+1)} = H^{(k)} + \frac{g^{(k)}g^{(k)T}}{g^{(k)T}S^{(k)}} - \frac{H^{(k)}S^{(k)}S^{(k)T}H^{(k)}}{S^{(k)T}H^{(k)}S^{(k)}}$ (40)

(40) гарантирует симметричность и положительно-определенность матриц секущих на каждом шаге итерации.
(40) называется матрицей **Бройдена-Флетчера-Голдфарва-Шанно** (**BFGS**)

DFP $S^{(k)} = -G_f^{-1}\nabla^T f(x^{(k)})$ (31)

$S^{(k)}$ – называется ньютоновским направлением поиска.
$H^{(k)}$ можно искать из условия линейной интерполяции

$\mathrm{H}^{(k)} : \nabla^T Q(x^{(k-1)}) = \nabla^T f(x^{(k-1)})$

$\nabla^T f(x^{(k)}) + \mathrm{H}^{(k)}(x^{(k-1)} - x^{(k)}) = \nabla^T f(x^{(k-1)})$

Если $\mathrm{S}^{(k-1)} = x^{(k-1)} - x^{(k)}$ и $g^{(k-1)} = \nabla^T f(x^{(k)}) - \nabla^T f(x^{(k-1)})$, тогда **соотношение секущих** (квазиньютоновское условие) для оптимизации имеет вид:

$H^{(k)}S^{(k-1)} = g^{(k-1)}$

$H^{(k+1)}S^{(k)} = g^{(k)}$

Матрицы $\mathrm{H}^{(k)}$, $\mathrm{H}^{(k+1)}$… не гарантируют симметричность и положительно-определенность.
Рассмотрим матрицу $H^{(k+1)-1} = L^{(k+1)-T}L^{(k+1)-1}$, тогда квазиньютоновское условие запишется в виде:

$L^{(k+1)-T}L^{(k+1)-1}g^{(k)} = S^{(k)}$ (41)

Аналогично рассуждая, результатом (41) является:

$L^{(k+1)T} = L^{(k+1)-T} + (S^{(k)} - L^{(k)-T}\omega^{(k)})\omega^{(k)+}$ Подставим (42) в формулу Бройдена и

$\omega^{(k)} = \left(\frac{g^{(k)T}S^{(k)}}{g^{(k)T}H^{(k)-1}g^{(k)}}\right)^{\frac{1}{2}}\mathrm{L}^{(k)-1}g^{(k)}$ (42)

получим: $$H^{(k+1)-1} = H^{(k)-1} + \frac{S^{(k)}S^{(k)T}}{g^{(k)}S^{(k)}} - \frac{H^{(k)-1}g^{(k)}g^{(k)T}H^{(k)-1}}{g^{(k)T}H^{(k)-1}g^{(k)}} \qquad (43)$$

(43) называется матрицей **Давидона-Флетчера-Пауэлла (DFP)**.(40) и (43) - симметричные и положительно-определенные матрицы и обратные к ним матрицы также будут обладать этими свойствами.

Прямая формула для **DFP**:

$$H^{(k+1)} = H^{(k)} + \frac{(g^{(k)} - H^{(k)}S^{(k)})g^{(k)T} + g^{(k)}(g^{(k)} - H^{(k)}S^{(k)})^T}{g^{(k)T}S^{(k)}} - \frac{(g^{(k)} - H^{(k)}S^{(k)})^T S^{(k)}g^{(k)}g^{(k)T}}{(g^{(k)T}S^{(k)})^2} \qquad (45)$$