Квазиньютоновские методы оптимизации
Метод Modified Symmetric Rank-One(+MSR1)

Сергей Шилин

Описание Метод SR1 - безусловный квазиньютоновский метод, который за каждую итерацию обновляет матрицу Гессе вторых частных производных, рассчитанных на двух дочках. Как КН-метод, основывается на накоплении информации о кривизне целевой функции по наблюдениям за изменением градиента. Этот метод исключает явное форматирование матрицы Гессе, заменяя её некоторым приближением.

Задача Рассмотрим задачу безусловной оптимизации

$$\min_{x\in\mathbb{R}} f(x) \ . \tag{1}$$

Функция $f : \mathbb{R}^n \to \mathbb{R}$ является гладкой, её градиент в точке x_k $\nabla f(x_k)$. Обозначим

$$g_k = \nabla f(x_k)$$

Функция f должна быть непрерывной и дважды дифференцируемой. Данный метод разрабатывался, основываясь на методе Ньютона, в котором матрица Гессе H_k функции f в точке x_k заменяется некоторой матрицей B_k для избежания перерасчета матрицы Гессе. Для решения уравнения (1) используется следующий алгоритм.

Алгоритм Рассматривая k-ую итерацию и, именно, градиент $\nabla f(x_k)$ в точке x_k, мы определяем направление спуска d_k из уравнения

$$d_k = -\frac{\nabla f(x_k)}{B_k} \ , \tag{2}$$

где $B_k = H_k^-1$. На следующей итерации мы получаем приближение

$$x_{k+1} = x_k + \alpha_k d_k \ , \tag{3}$$

где α – длина шага приближения. Алгоритм нахождения α будет описан позже. Далее следует обновить матрицу B, используя следующее равенство

$$B_{k+1} s_k = y_k \ ,$$

где $s_k = x_{k+1} - x_k$ и $y_k = g_{k+1} - g_k$. Однако уравнение секущей использует только информацию о градиенте и игнорирует информацию о функции.

Улучшенное уровнение секущей будет иметь вид

$$B_{k+1}s_k = y_k^* \ , \tag{4}$$

где

$$y_k^* = y_k + A_k s_k \ , \tag{5}$$

и A_k является простой симметричной положительно определенной матрицей.

Пошаговый алгоритм

Шаг 1. $k = 0$. Имеем начальную точку x_0 и матрицу Гессе, равную единичной $H_0 = E$.

Шаг 2. Если выполняется критерий сходимости (6), то останавливаемся. Минимум достигнут.

$$\|\nabla f(x_k)\| \leqslant \varepsilon \times \max(1, \|x_k\|) \tag{6}$$

Шаг 3. По формуле (2) находим направление спуска d_k

Шаг 4. Находим допустимое значение α величины шага приближения, исходя из условий Вольфа

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \delta_1 \alpha_k \nabla f(x_k)^T d_k, \tag{7a}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geqslant \delta_2 \nabla f(x_k)^T d_k. \tag{7b}$$

$\delta_1 = 10^{-4}$, $\delta_2 = 0.9$, всегда сперва пробуем $\alpha = 1$.

Шаг 5. Задаем следующее приблежение x_{k+1} по формуле (3).

Шаг 6. Проверяем условия для обновления матрицы Гессе H_k и направления d_k . Если

$$s_k^T y_k - y_k^T H_k y_k < 0, \tag{8}$$

или

$$|\tilde{y}_k^T (s_k - H_k \tilde{y}_k)| < r \|\tilde{y}_k\| \|s_k - H_k \tilde{y}_k\|, \tag{9}$$

где $r \in (0, 1)$

или

$$\|H_k\|_\infty > L, \tag{10}$$

где L - заданная константа,

то задаем $H_{k+1} = \tilde{\lambda}_k E$, где $\tilde{\lambda}_k$ выходит из формулы

$$\tilde{\lambda}_k = \frac{s_k^T s_k}{\tilde{y}_k^T s_k} - \left\{ \frac{(s_k^T s_k)^2}{(\tilde{y}_k^T s_k)^2} - \frac{s_k^T s_k}{\tilde{y}_k^T y_k} \right\}^{1/2} \tag{11}$$

и в последствии $d_{k+1} = -\tilde{\lambda}_k \nabla f(x_k)$.

Шаг 7. Рассчитаем $\tilde{y}_k$, используя равенство

$$\tilde{y}_k = y_k + sgn(\psi_k)\frac{\psi_k}{s_k^T s_k} s_k,$$

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

Шаг 8. Снова обновим матрицу Гессе H_k

$$H_{k+1} = H_k + \frac{(s_k - H_k\tilde{y}_k)(s_k - H_k\tilde{y}_k)^T}{(s_k - H_k\tilde{y}_k)^T\tilde{y}_k} \tag{12}$$

Шаг 9. $k = k + 1$ и переходим на *Шаг 1.*

Данный материал взят из Journal of Mathematics Research, Vol. 2, No. 3, August 2010, и переведен на русский язык. Автором этого материала была написана реализация алгоритма +MSR1 на языке программирования Java.