

19. Безусловная многомерная оптимизация. Нелинейная задача наименьших квадратов (НЗНК). Метод Ньютона для НЗНК. Метод Левенберга-Маркварута. Методы квазиньютоновского типа.

Среди задач на поиск безусловного оптимума особое место занимает задача следующего вида:

$$\text{Найти } \min f(x) = \frac{1}{2} \sum_{i=1}^m r_i^2(x) = \frac{1}{2} \|R(x)\|_2^2 = \frac{1}{2} R^T(x) R(x),$$

где $R(x) = [r_1(x) \dots r_m(x)]^T \in R^m$ –

нелинейная векторная функция векторного аргумента (7)

(7) возникает, например, при настройке математической модели на реальные данные.

Под математической моделью понимается некоторая функция $\varphi(x, t)$ с независимым аргументом t и векторным параметром x . Пусть получена таблица экспериментальных данных. Тогда согласование модели с реальностью будет состоять в том, чтобы подобрать параметр x коэффициента регрессии, при

котором $r_i(x) = \varphi(x, t_i) - y_i$

Если $R(x)$ линейна, то (7) представляет линейную задачу о наименьших квадратах (ЛЗНК).

Фактически задача регрессии формируется в виде:

$$\text{Найти } \min f(x) = \frac{1}{2} \sum_{i=1}^m (\varphi(x, t_i) - y_i)^2$$

Хотя для решения (7) обращаются к специальным алгоритмам.

Матрица первых производных $R(x)$ это матрица Якоби

$$J(x) = \begin{bmatrix} \nabla^T r_1 \\ \vdots \\ \nabla^T r_m \end{bmatrix}, \quad \nabla r_i = \begin{bmatrix} \partial r_i / \partial x_1 \\ \vdots \\ \partial r_i / \partial x_n \end{bmatrix}, \quad i \in 1 \dots m$$

размерностью $m \times n$.

Т.о. аффинной моделью функции $R(x)$ в окрестности точки $x^{(k)}$

будет: $M(x) = R(x^{(k)}) + J(x^{(k)})(x - x^{(k)})$ (8)

Первой производной от $f(x)$ из (8) является:

$$\nabla f(x) = \sum_{i=1}^m r_i(x) \nabla r_i(x) = J^T(x) R(x)$$

Аналогично вторая производная:

$$\nabla^2 f(x) = \sum_{i=1}^m (\nabla r_i(x) \nabla^T r_i(x) + r_i(x) \nabla^2 r_i(x)) = J^T(x) J(x) + \theta(x),$$

$$\text{где } \theta = \sum_{i=1}^m r_i(x) \nabla^2 r_i(x)$$

Тогда, используя ньютоновское соотношение

$$\nabla^2 f(x^{(k)}) S^{(k)} = -\nabla f(x^{(k)}), \text{ уравнение для определения}$$

направления $S^{(k)}$ запишется в виде:

$$[J^T(x) J(x) + \theta(x)] S^{(k)} = -J^T(x) R(x)$$

Тогда модельная схема итерационных процедур для решения

НЗНК будет базироваться на формуле:

$$x^{(k+1)} = x^{(k)} - \left(J^T(x^{(k)}) J(x^{(k)}) + \theta(x^{(k)}) \right)^{-1} J^T(x^{(k)}) R(x^{(k)}) \quad (9)$$

Метод, использующий (9) называется методом Ньютона. Трудность

заключается в том, что $\theta(x)$ в явном виде не используется.

Способы аппроксимации $\theta(x)$ породили ряд методов для НЗНК.

Метод Гаусса-Ньютона.

Данный метод относится к методам первого порядка, т.к.

основан на пренебрежении $\theta(x)$. В этом случае (9) имеет вид:

$$x^{(k+1)} = x^{(k)} - \left(J^T(x^{(k)}) J(x^{(k)}) \right)^{-1} J^T(x^{(k)}) R(x^{(k)})$$

$$\text{или } x^{(k+1)} = x^{(k)} - J^T(x^{(k)}) R(x^{(k)}) \quad (10)$$

Очевидно, что успех применения метода Гаусса-Ньютона будет

зависеть от того, насколько весомо окажется $\theta(x)$. Если $\theta(x)$

мало по сравнению с $J^T(x) J(x)$, то метод Гаусса-Ньютона является быстро локально сходящимся. В противном случае метод может вообще не сойтись.

Особенностью метода является то, что он в большинстве случаев делает плохие шаги в окрестности точки оптимума, выбирая их слишком большими по длине, но верными по направлению. Это обстоятельство приводит к использованию процедур изменения $\lambda^{(k)}$ и тогда имеет место формула:

$$x^{(k+1)} = x^{(k)} - \lambda^{(k)} J^T(x^{(k)}) R(x^{(k)}) \quad (11)$$

где $\lambda^{(k)}$ может выбираться известными методами одномерного поиска.

На (11) базируется демпфированный метод Гаусса-Ньютона.

Достоинства: этот метод локально сходится почти для всех НЗНК, включая задачи с большой невязкой, хотя в некоторых ситуациях это делается довольно медленно.

Метод Левенберга-Марквардта.

$$S^{(k)} = - \left(J^{(k)T} J^{(k)} + \theta^{(k)} \right)^{-1} J^{T(k)} R(x^{(k)})$$

Основан на замене $\theta(x^{(k)})$ в (9) выражением $\mu^{(k)} \mathbb{I}$.

$\mu^{(k)} \in R^n$ - параметр регуляризации, который регулирует не

только длину шага, но и направление поиска. Тогда

итерационная формула будет иметь вид:

$$x^{(k+1)} = x^{(k)} - \left(J^T(x^{(k)})J(x^{(k)}) + \mu^{(k)}I \right)^{-1} J^T(x^{(k)})R(x^{(k)}) \quad (12)$$

Самый простой алгоритм метода Л.-М. можно описать следующими шагами:

Шаг 1: Задать $x^{(0)}$, τ , $k=0$, $\mu^{(0)}$ (например: $\mu^{(0)}=10^4$)

Шаг 2: Проверить условия остановки.

Например: $\|J^T(x^{(k)})R(x^{(k)})\| \leq \varepsilon$

если нет, то переходим на шаг 3

Шаг 3: $S(x^{(k)}) = -\left(J^T(x^{(k)})J(x^{(k)}) + \mu^{(k)}I \right)^{-1} J^T(x^{(k)})R(x^{(k)})$

Шаг 4: $x^{(k+1)} = x^{(k)} + S(x^{(k)})$

Шаг 5: Проверить условие

$f(x^{(k+1)}) < f(x^{(k)})$

если выполняется, то переходим на шаг 6

если нет, то переходим на шаг 7

Шаг 6: $\mu^{(k+1)} = 0.5\mu^{(k)}$, $k \leftarrow k+1$ и переходим на шаг 2

Шаг 7: $\mu^{(k)} = 2\mu^{(k)}$ и переходим на шаг 3

Существуют также другие алгоритмы выбора параметра μ .

Наиболее удачная стратегия была предложена Морэ в 1977 году. Подход использует алгоритм доверительной области.

Пусть есть $x^{(k)}$, оценка сверху $\delta^{(k)}$ максимальной длины успешного шага из $x^{(k)}$. Тогда $x^{(k+1)}$ можно получить исходя из задачи:

Найти $\min \|M(x^{(k+1)})\| = \|R(x^{(k)}) + J(x^{(k)})(x^{(k+1)} - x^{(k)})\|$

при условии $\|x^{(k+1)} - x^{(k)}\| \leq \delta^{(k)}$

где $\mu \geq 0$, такое что:

$$\mu^{(k)} = \begin{cases} 0, & \text{если } \delta^{(k)} \geq \left\| \left(J^T(x^{(k)})J(x^{(k)}) \right)^{-1} J^T(x^{(k)})R(x^{(k)}) \right\| \\ > 0, & \text{в остальных случаях} \end{cases}$$

Это означает, что $\delta^{(k)}$ порождает область в которой аффинная модель $M(x)$ адекватно моделирует $f(x)$. Тогда направлением в (12) является:

$$S(\mu) = -\left(J^T(x)J(x) + \mu I \right)^{-1} J^T(x)R(x)$$

Значение $f(x)$ в доверительной области улучшается.

Для единственного $\mu \geq 0$, такое что:

$$\mu^{(k)} = \begin{cases} 0, & \text{если } \delta^{(k)} \geq \left\| \left(J^T(x^{(k)})J(x^{(k)}) \right)^{-1} J^T(x^{(k)})R(x^{(k)}) \right\| \\ > 0, & \text{в остальных случаях} \end{cases}$$

Алгоритм Л.-М. на многих задачах является предпочтительнее, чем Гаусса-Ньютона и достоинством является то, что он корректно определен, если матрица Якоби не имеет полного столбцового ранга.

Методы квазиньютоновского типа.

Являются более надежными, но и более сложными.

Строятся на замене $\theta(x^{(k)})$ квазиньютоновскими приближениями $A(x^{(k)})$. В этом случае:

$$x^{(k+1)} = x^{(k)} - \left(J^T(x^{(k)})J(x^{(k)}) + A(x^{(k)}) \right)^{-1} J^T(x^{(k)})R(x^{(k)}) \quad (13)$$

Здесь $A(x^{(k)})$ -аппроксимация $\theta(x)$ секущими.

Введем обозначения:

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

$$S^{(k)} = x^{(k+1)} - x^{(k)}$$

Используем для аппроксимации вторых производных конечно-разностное соотношение секущих.

$$(D_i)^{(k+1)}(x^{(k+1)} - x^{(k)}) = \nabla r_i(x^{(k+1)}) - \nabla r_i(x^{(k)}) =$$

$$= (i\text{-ая строка } J(x^{(k+1)}))^T - (i\text{-ая строка } J(x^{(k)}))^T$$

$$(D_i)^{(k+1)} - \text{аппроксимация } \nabla^2 r_i(x^{(k+1)})$$

$$\theta(x^{(k+1)}) \approx A(x^{(k+1)}) = \sum_{i=1}^m r_i(x^{(k+1)})(D_i)^{(k+1)} \text{ или}$$

$$A^{(k+1)}(x^{(k+1)} - x^{(k)}) = \sum_{i=1}^m r_i(x^{(k+1)})(D_i)^{(k+1)}(x^{(k+1)} - x^{(k)}) =$$

$$J^T(x^{(k+1)})R(x^{(k+1)}) - J^T(x^{(k)})R(x^{(k+1)}) = \tilde{g}^{(k)}$$

Если далее использовать рассуждения, аналогичные при выводе квазиньютоновских формул безусловной оптимизации с соответствующими обозначениями, то получим:

DFF:

$$A^{(k+1)} = A^{(k)} + \frac{(\tilde{g}^{(k)} - A^{(k)}S^{(k)})g^{(k)T} + g^{(k)}(\tilde{g}^{(k)} - A^{(k)}S^{(k)})^T}{g^{(k)T}S^{(k)}} - \quad (14)$$

$$- \frac{(\tilde{g}^{(k)} - A^{(k)}S^{(k)})^T S^{(k)} g^{(k)} g^{(k)T}}{(g^{(k)T}S^{(k)})^2}$$

BFGS:

$$A^{(k+1)} = A^{(k)} + \frac{g^{(k)} g^{(k)T}}{g^{(k)T} S^{(k)}} - \frac{\tilde{A}^{(k)} S^{(k)} S^{(k)T} \tilde{A}^{(k)}}{S^{(k)T} \tilde{A}^{(k)} S^{(k)}} \quad (15)$$
$$\tilde{A}^{(k)} = J^T(x^{(k+1)}) J(x^{(k+1)}) + A^{(k)}$$

На практике на первых итерациях независимо от типа используется метод Гаусса-Ньютона в чистом виде (с $\lambda=1$).

А в окрестности точки оптимума можно использовать либо регуляризатор Л.-М., либо формулы секущих.