МИСиС
Национальный исследовательский
технологический университет

Методы оптимизации

Лекция 3

17 марта 2022 г.

МЕТОДЫ МНОГОМЕРНОЙ МИНИМИЗАЦИИ ВТОРОГО ПОРЯДКА

Предположим, что функция f выпукла и дважды дифференцируема на $\mathbb{R}^n$, причем матрица $\nabla^2 f(x)$ невырождена на $\mathbb{R}^n$. В методе Ньютона последовательность x^0, x^1, x^2, ... генерируется, исходя из следующих соображений. По определению дважды дифференцируемой функции, для очередной точки x^k имеем

$$f(x) - f(x^k) = \left(\nabla f(x^k), x - x^k\right) + \\ + \frac{1}{2}\left(\nabla^2 f(x^k)(x - x^k), x - x^k\right) + o\left(\|x - x^k\|^2\right)$$

Для определения следующей точки x^{k+1} минимизируется квадратичная часть приращения $f(x) - f(x^k)$, т. е. решается задача

$$f_k(x) = \left(\nabla f(x^k), x - x^k\right) + \frac{1}{2}\left(\nabla^2 f(x^k)(x - x^k), x - x^k\right) \to \min, \quad x \in \mathbb{R}^n$$

Необходимое и достаточное условие минимума имеет вид

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

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

Решая полученную систему линейных уравнений и принимая найденную точку минимума за x^{k+1}, получаем

$$x^{k+1} = x^k + h^k, \quad h^k = -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$$

Теорема 1.

Пусть $f(x)$ дважды дифференцируема, $\nabla^2 f(x)$ удовлетворяет условию Липшица с константой L, $f(x)$ сильно выпукла с константой l и начальное приближение удовлетворяет условию

$$q = \left(Ll^{-2}/2\right)\left\|\nabla f(x^0)\right\| < 1$$

Тогда метод Ньютона сходится к точке минимума x^* с квадратичной скоростью

$$\left\|x^k - x^*\right\| \leqslant (2l/L)q^{2^k}.$$

Метод Ньютона

- Сходимость метода Ньютона доказана лишь для достаточно хорошего начального приближения x^0. При этом условие, гарантирующее сходимость для данного начального приближения, труднопроверяемо, так как фигурирующие в нём константы, как правило, неизвестны.
- Сложность отыскания нужного начального приближения является недостатком метода Ньютона. Ещё более существенным недостатком является высокая трудоёмкость метода, обусловленная необходимостью вычисления и обращения на каждом шаге матрицы вторых производных минимизируемой функции.

Достаточное условие локальной оптимальности

Пусть $f(x)$ дважды дифференцируема в точке $x^* \in \mathbb{R}$, причём $\nabla f(x^*) = 0$, т. е. x^* — стационарная точка.

Тогда, если матрица $\nabla^2 f(x^*)$ является положительно (отрицательно) определённой, то x^* — точка локального минимума (максимума); если матрица $\nabla^2 f(x^*)$ является неопределённой, то x^* — седловая точка.

Для проверки знакоопределённости матрицы, используется

Критерий Сильвестра

Симметричная матрица является положительно определённой в том и только том случае, если все её угловые миноры положительны. Для отрицательной определённости квадратичной формы необходимо и достаточно, чтобы угловые миноры чётного порядка её матрицы были положительны, а нечётного порядка — отрицательны.

Угловые миноры

Пусть матрица A $(a_{ij} = a_{ji};\ i, j = 1, \ldots, n)$ имеет вид

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Угловыми минорами этой матрицы являются:

$$\Delta_1 = a_{11}, \quad \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \ldots, \quad \Delta_n = \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}.$$

Как видим, угловой минор порядка k расположен на пересечении первых k строк и первых k столбцов матрицы. Угловой минор максимального, n-го порядка представляет собой определитель матрицы.

Метод Ньютона

Алгоритм.

1. Задать начальную точку x^0, погрешности расчёта $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M — предельное число итераций. Найти в произвольной точке градиент функции $\nabla f(x)$ и матрицу Гессе $H(x) = \nabla^2 f(x)$.

2. Принять $k := 0$.

3. Вычислить $\nabla f(x^k)$.

4. Проверить выполнение критерия окончания $\left\|\nabla f(x^k)\right\| < \varepsilon_1$:

 - если критерий выполнен, расчёт закончен $x^* := x^k$;
 - если критерий не выполнен, то перейти к п. 5;

5. Проверить выполнение неравенства $k \geqslant M$:

 - если неравенство выполнено, то расчёт окончен: $x^* := x^k$;
 - если нет, то перейти к п. 6.

6. Вычислить матрицу $H(x^k)$.

7. Вычислить матрицу $H^{-1}(x^k)$.

8. Проверить выполнение условия $H^{-1}(x^k) > 0$ (матрица положительно определена). Если $H^{-1}(x^k) > 0$, то перейти к п. 9; иначе перейти к п. 10, приняв $d^k := -\nabla f(x^k)$.

9. Определить $d^k := -H^{-1}(x^k)\nabla f(x^k)$.

10. Вычислить $x^{k+1} := x^k + t_k d^k$, приняв $t_k := 1$, если $d^k = -H^{-1}(x^k)\nabla f(x^k)$ или выбрав t_k из условия $f(x^{k+1}) < f(x^k)$, если $d^k = -\nabla f(x^k)$.

11. Проверить выполнение условий

$$\left\|x^{k+1} - x^k\right\| \leqslant \varepsilon_2, \quad \left|f(x^{k+1}) - f(x^k)\right| \leqslant \varepsilon_2$$

- если оба условия выполнены при текущем значении k и при $k-1$, то расчёт окончен и $x^* := x^{k+1}$;
- если хотя бы одно из условий не выполнено, то принять $k := k+1$ и перейти к п. 3.

Метод Ньютона с регулировкой шага

Придать методу Ньютона свойство глобальной сходимости можно различными способами. Один из них связан с регулировкой длины шага:

$$x^{k+1} = x^k - \gamma_k \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$$

Параметр γ_k может выбираться по-разному, например

$$\gamma_k = \arg\min_{\gamma \geqslant 0} f\left(x^k - \gamma \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)\right)$$

или γ дробится (умножается на $0 < a < 1$), начиная с $\gamma = 1$, до выполнения условия

$$f(x^{k+1}) \leqslant f(x^k) - \gamma q \left(\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k), \nabla f(x^k)\right), \quad 0 < q < 1$$

или условия

$$\left\|\nabla f(x^{k+1})\right\|^2 \leqslant (1 - \gamma q)\left\|\nabla f(x^k)\right\|^2, \quad 0 < q < 1$$

Метод Ньютона с регулировкой шага

– Для гладких сильно выпуклых функций метод Ньютона с регулировкой шага глобально сходится
– Что касается скорости сходимости, то на начальных итерациях можно утверждать лишь сходимость со скоростью геометрической прогрессии.
– При попадании в окрестность x^*, в которой выполняются условия теоремы 1, будет иметь место квадратичная сходимость.

МИСиС

Метод Ньютона с регулировкой шага

Алгоритм.

1. Задать начальную точку x^0, погрешности расчёта $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M — предельное число итераций. Найти в произвольной точке градиент функции $\nabla f(x)$ и матрицу Гессе $H(x) = \nabla^2 f(x)$.

2. Принять $k := 0$.

3. Вычислить $\nabla f(x^k)$.

4. Проверить выполнение критерия окончания $\left\|\nabla f(x^k)\right\| < \varepsilon_1$:

   - если критерий выполнен, расчёт закончен $x^* := x^k$;
   - если критерий не выполнен, то перейти к п. 5;

5. Проверить выполнение неравенства $k \geqslant M$:

   - если неравенство выполнено, то расчёт окончен: $x^* := x^k$;
   - если нет, то перейти к п. 6.

6. Вычислить матрицу $H(x^k)$.

Метод Ньютона с регулировкой шага

Алгоритм.

7. Вычислить матрицу $H^{-1}(x^k)$.

8. Проверить выполнение условия $H^{-1}(x^k) > 0$:

   – если $H^{-1}(x^k) > 0$, то найти $d^k := -H^{-1}(x^k)\nabla f(x^k)$;
   – если $H^{-1}(x^k) \leqslant 0$, то принять $d^k := -\nabla f(x^k)$.

9. Найти шаг t_k^* из условия $\varphi(t_k) = f(x^k + t_k d^k) \to \min\limits_{t_k}$

10. Вычислить $x^{k+1} := x^k + t_k^* d^k$.

11. Проверить выполнение условий

$$\left\|x^{k+1} - x^k\right\| \leqslant \varepsilon_2, \quad \left|f(x^{k+1}) - f(x^k)\right| \leqslant \varepsilon_2$$

   – если оба условия выполнены при текущем значении k и при $k-1$, то расчёт окончен и $x^* := x^{k+1}$;
   – если хотя бы одно из условий не выполнено, то принять $k := k+1$ и перейти к п. 3.

Метод Левенберга-Марквардта

В этом методе само направление движения отличается от задаваемого методом Ньютона.

Добавим к аппроксимирующей функции квадратичный штраф за отклонение от точки x^k, т. е. будем искать x^{k+1} из условия

$$f_k(x) + \frac{\alpha_k}{2}\left\|x - x^k\right\|^2 \to \min$$

при

$$f_k(x) = \left(\nabla f(x^k), x - x^k\right) + \frac{1}{2}\left(\nabla^2 f(x^k)(x - x^k), x - x^k\right) \to \min, \quad x \in \mathbb{R}^n$$

Тогда приходим к методу

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) + \alpha_k I\right)^{-1} \nabla f(x^k).$$

При $\alpha_k = 0$ метод переходит в метод Ньютона, при $\alpha_k \to \infty$ направление движения стремится к антиградиенту. Таким образом, метод представляет собой компромисс между этими двумя методами.

Метод Левенберга-Марквардта

Алгоритм.

1. Задать начальную точку x^0, погрешности расчёта $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M — предельное число итераций. Найти в произвольной точке градиент функции $\nabla f(x)$ и матрицу Гессе $H(x) = \nabla^2 f(x)$.

2. Принять $k := 0$, $\mu^k := \mu^0$.

3. Вычислить $\nabla f(x^k)$.

4. Проверить выполнение критерия окончания $\left\|\nabla f(x^k)\right\| < \varepsilon_1$:

   - если критерий выполнен, расчёт закончен $x^* := x^k$;
   - если критерий не выполнен, то перейти к п. 5;

5. Проверить выполнение неравенства $k \geqslant M$:

   - если неравенство выполнено, то расчёт окончен: $x^* := x^k$;
   - если нет, то перейти к п. 6.

МИСиС

Алгоритм.

6. Вычислить матрицу $H(x^k)$.
7. Вычислить матрицу $H(x^k) + \mu^k I$.
8. Вычислить матрицу $[H(x^k) + \mu^k I]^{-1}$.
9. Вычислить $d^k := -[H(x^k) + \mu^k I]^{-1} \nabla f(x^k)$.
10. Вычислить $x^{k+1} := x^k - [H(x^k) + \mu^k I]^{-1} \nabla f(x^k)$.
11. Проверить выполнение условия $f(x^{k+1}) < f(x^k)$,
    - если условие выполняется, то перейти к п. 12;
    - если нет, то перейти к п. 13.
12. Принять $k := k + 1$, $\mu^{k+1} := \dfrac{\mu^k}{2}$ и перейти к п. 3.
13. Принять $\mu^k := 2\mu^k$ и перейти к п. 7.

МНОГОШАГОВЫЕ МЕТОДЫ

В градиентном методе на каждом шаге никак не используется информация, полученная на предыдущих итерациях. Естественно попытаться учесть «предысторию» процесса для ускорения сходимости. Такого рода методы, в которых новое приближение зависит от s предыдущих:

$$x^{k+1} = \varphi_k(x^k, \dots, x^{k-s+1})$$

называются s-*шаговыми*. Градиентный метод и метод Ньютона были одношаговыми, теперь рассмотрим *многошаговые* $(s > 1)$ методы.

Одним из простейших многошаговых методов является двухшаговый метод тяжёлого шарика

$$x^{k+1} = x^k - \alpha \nabla f(x^k) + \beta(x^k - x^{k-1}), \tag{1}$$

где $\alpha > 0$, $\beta \geqslant 0$ — некоторые параметры.

При $\beta = 0$ метод переходит в градиентный.

Метод тяжёлого шарика

Своё название метод получил из-за следующей физической аналогии. Движение тела («тяжёлого шарика») в потенциальном поле при наличии силы трения (или вязкости) описывается дифференциальным уравнением второго порядка

$$\mu \frac{d^2 x(t)}{dt^2} = -\nabla f(x(t)) - p \frac{dx(t)}{dt} \tag{2}$$

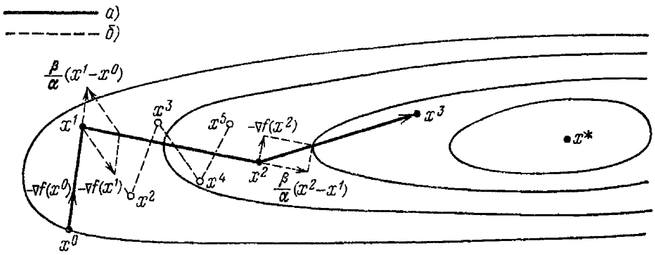
Из-за потери энергии на трение тело в конце концов окажется в точке минимума потенциала $f(x)$. Таким образом, тяжёлый шарик «решает» соответствующую задачу минимизации.

Если рассмотреть разностный аналог уравнения (2), то придём к итерационному методу (1).

Метод тяжёлого шарика

Введение инерции движения (слагаемое $\beta(x^k - x^{k-1})$)
в итерационный процесс может привести к ускорению сходимости.

Вместо зигзагообразного движения в градиентном методе в данном случае получается более плавная траектория по «дну оврага».



Метод тяжёлого шарика (а) и градиентный метод (б)

Эти эвристические соображения подкрепляются следующей теоремой.

Теорема 2.

Пусть x^* — невырожденная точка минимума $f(x)$, $x \in \mathbb{R}^n$. Тогда при

$$0 \leqslant \beta < 1, \quad 0 < \alpha < 2(1+\beta)/L, \quad lI \leqslant \nabla^2 f(x^*) \leqslant LI$$

найдётся $\varepsilon > 0$ такое, что при любых x^0, x^1, $\left\|x^0 - x^*\right\| \leqslant \varepsilon$, $\left\|x^1 - x^*\right\| \leqslant \varepsilon$ метод (1) сходится к x^* со скоростью геометрической прогрессии:

$$\left\|x^k - x^*\right\| \leqslant c(\delta)(q+\delta)^k, \quad 0 \leqslant q < 1, \quad 0 < \delta < 1 - q.$$

Величина q минимальна и равна

$$q^* = \frac{\sqrt{L} - \sqrt{l}}{\sqrt{L} + \sqrt{l}} \quad \text{при} \quad \alpha^* = \frac{4}{\left(\sqrt{L} + \sqrt{l}\right)^2}, \quad \beta^* = \left(\frac{\sqrt{L} - \sqrt{l}}{\sqrt{L} + \sqrt{l}}\right)^2.$$

Сравним скорость сходимости, даваемую одношаговым и двухшаговым методами при оптимальном выборе параметров. И в том, и в другом случаях имеем сходимость со скоростью геометрической прогрессии, но знаменатель прогрессии для одношагового метода равен

$$q_1 = (L - l)/(L + l),$$

а для двухшагового

$$q_2 = (\sqrt{L} - \sqrt{l})/(\sqrt{L} + \sqrt{l}).$$

Для больших значений числа обусловленности $\mu = L/l$

$$q_1 \approx 1 - 2/\mu, \quad q_2 \approx 1 - 2/\sqrt{\mu}$$

Для плохо обусловленных задач метод тяжёлого шарика даёт выигрыш в $\sqrt{\mu}$ раз по сравнению с градиентным.

С вычислительной же точки зрения метод (1) немногим сложнее одношагового.

Метод сопряжённых градиентов

Высокая скорость сходимости метода Ньютона обусловлена тем, что он минимизирует квадратичную функцию

$$f(x) = \frac{1}{2}(Ax, x) + (b, x) \tag{3}$$

где A — симметрическая положительно определённая матрица размера $n \times n$, за один шаг.

Идея методов сопряжённых градиентов основана на стремлении минимизировать квадратичную функцию за конечное число шагов. Точнее говоря, в методах сопряжённых градиентов требуется найти направления $h^0, h^1, \ldots, h^{n-1}$ такие, что последовательность n одномерных минимизаций вдоль этих направлений приводит к отысканию минимума квадратичной функции $f(x^n) = \min\limits_{x \in \mathbb{R}^n} f(x)$ при любом $x^0 \in \mathbb{R}^n$, где

$$x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \arg\min_{\alpha \geqslant 0} f(x^k + \alpha p^k) \tag{4}$$

Определение.

Векторы (направления) p' и p'' называются сопряжёнными (относительно матрицы A), если они отличны от нуля и $(Ap', p'') = 0$. Векторы (направления) $p^0, p^1, \ldots, p^k$ называются взаимно сопряжёнными (относительно матрицы A), если все они отличны от нуля и $(Ap^i, p^j) = 0$, $i \neq j$, $0 \leqslant i, j \leqslant k$.

Теорема 3.

Если векторы p^k в методе (4) взаимно сопряжены, $k = 0, 1, \ldots, n-1$, то для функции f, заданной формулой (3), и произвольной точки $x^0 \in \mathbb{R}^n$

$$f(x^n) = \min_{x \in \mathbb{R}^n} f(x)$$

Таким образом, метод (4) позволяет найти точку минимума квадратичной функции (3) не более чем за n шагов.

Метод сопряжённых градиентов

Метод Флетчера-Ривза

$$
\begin{aligned}
&x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \arg\min_{\alpha \geqslant 0} f(x^k + \alpha p^k) \\
&p^k = -\nabla f(x^k) + \beta_k p^{k-1}, \\
&\beta_0 = 0, \quad \beta_k = \frac{\left\|\nabla f(x^k)\right\|^2}{\left\|\nabla f(x^{k-1})\right\|^2}
\end{aligned}
$$

Метод Полака-Рибьера

$$
\begin{aligned}
&x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \arg\min_{\alpha \geqslant 0} f(x^k + \alpha p^k) \\
&p^k = -\nabla f(x^k) + \beta_k p^{k-1}, \\
&\beta_0 = 0, \quad \beta_k = \frac{\left(\nabla f(x^k), \nabla f(x^k) - \nabla f(x^{k-1})\right)}{\left\|\nabla f(x^{k-1})\right\|^2}
\end{aligned}
$$

Метод сопряжённых градиентов можно интерпретировать как градиентный метод с памятью.

Метод сопряжённых градиентов

Из теоремы 3 следует, что методы дают точку минимума квадратичной функции в $\mathbb{R}^n$ за число итераций, не превосходящее n.

Иначе говоря, по скорости сходимости n шагов метода сопряжённых градиентов эквивалентны одному шагу метода Ньютона.

Разумеется, в неквадратичном случае теряются свойства конечности методов и они превращаются в, вообще говоря, бесконечные итерационные двухшаговые методы.

Как показывает опыт вычислений, для неквадратичного случая несколько более быструю сходимость обычно даёт метод Полака-Рибьера.

Оказывается, что в то же время в окрестности минимума он сходится с квадратичной скоростью.

МИСиС

Метод сопряжённых градиентов

Обычно для неквадратичных задач метод сопряжённых градиентов применяется в несколько иной форме. В него вводится процедура обновления — время от времени шаг делается как в начальной точке, т. е. по градиенту.

Наиболее естественно производить обновление через число итераций, равное размерности пространства:

$$x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \arg\min_{\alpha \geqslant 0} f(x^k + \alpha p^k)$$

$$p^k = -\nabla f(x^k) + \beta_k p^{k-1},$$

$$\beta_k = \begin{cases} 0, & k = 0, n, 2n, \dots \\ \left\|\nabla f(x^k)\right\|^2 / \left\|\nabla f(x^{k-1})\right\|^2, & k \neq 0, n, 2n, \dots \end{cases}$$

МИСиС

Алгоритм.

1. Задать начальную точку x^0, погрешности расчёта $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M — предельное число итераций. Найти градиент функции в произвольной точке $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)^{\mathsf{T}}$.

2. Принять $k := 0$.

3. Вычислить $\nabla f(x^k)$.

4. Проверить выполнение критерия окончания $\left\|\nabla f(x^k)\right\| < \varepsilon_1$. Если критерий выполнен, расчёт закончен $x^* := x^k$.

5. Проверить выполнение неравенства $k \geqslant M$. Если неравенство выполнено, то расчёт окончен.

6. Вычислить

$$\beta_k := \begin{cases} 0, & k = 0, n, 2n, \ldots \\ \left\|\nabla f(x^k)\right\|^2 / \left\|\nabla f(x^{k-1})\right\|^2, & k \neq 0, n, 2n, \ldots \end{cases}$$

Алгоритм.

7. Определить $d^0 := -\nabla f(x^0)$ и $d^k := -\nabla f(x^k) + \beta_k d^{k-1}$ при $k \geqslant 1$.

8. Определить величину шага:

$$t_k^* = \arg\min_{t_k \geqslant 0} f(x^k + t_k d^k)).$$

9. Вычислить $x^{k+1} := x^k + t_k^* d^k$.

10. Проверить выполнение условий

$$\left\|x^{k+1} - x^k\right\| \leqslant \varepsilon_2, \quad \left|f(x^{k+1}) - f(x^k)\right| \leqslant \varepsilon_2$$

- если оба условия выполнены при текущем значении k и при $k-1$, то расчёт окончен и $x^* := x^k$;
- если хотя бы одно из условий не выполнено, то принять $k := k+1$ и перейти к п. 3.

КВАЗИНЬЮТОНОВСКИЕ МЕТОДЫ

Общая идея методов

Сам метод Ньютона является методом второго порядка, т. е. использует вычисление вторых производных минимизируемой на $\mathbb{R}^n$ функции $f(x)$.

В основе квазиньютоновских методов лежит идея восстановления квадратичной аппроксимации функции по значениям её градиентов в ряде точек. Тем самым методы объединяют достоинства градиентного метода (не требуется вычисление матрицы вторых производных) и метода Ньютона (быстрая сходимость вследствие использования квадратичной аппроксимации).

В квазиньютоновских методах матрица вторых производных аппроксимируется с помощью информации о значениях градиентов функции $f(x)$, и эти методы, таким образом, являются методами первого порядка.

Квазиньютоновские методы имеют общую структуру:

$$x^{k+1} = x^k - \gamma_k H_k \nabla f(x^k),$$

где матрица H_k пересчитывается рекуррентным способом на основе информации, полученной на k-й итерации, так что $H_k - [\nabla^2 f(x^k)]^{-1} \to 0$.

Таким образом, методы в пределе переходят в ньютоновский, что и объясняет их название.

Квазиньютоновские методы

Перейдём к вопросу о способах построения матриц H_k, аппроксимирующих $[\nabla^2 f(x^k)]^{-1}$. В принципе их можно формировать с помощью конечно-разностной аппроксимации. Из точки x^k можно сделать n «пробных шагов» длины α_k по координатным осям и вычислить в этих точках градиенты.

Однако такой прямолинейный способ аппроксимации неэкономен — в нём делается n пробных вычислений градиента на каждой итерации и никак не используются градиенты, найденные на предыдущих итерациях. Кроме того, в нём требуется обращать матрицу.

Основная идея квазиньютоновских методов заключается:

- в том, чтобы не делать специальных пробных шагов, а использовать найденные градиенты в предыдущих точках (поскольку они близки к x^k);
- в том, чтобы строить аппроксимацию непосредственно для обратной матрицы $[\nabla^2 f(x^k)]^{-1}$.

Итак, рассмотрим метод:

$$x^{k+1} = x^k - \gamma_k H_k \nabla f(x^k).$$

Матрицу H_k будем выбирать таким образом, чтобы она в некотором смысле аппроксимировала матрицу $[\nabla^2 f(x^k)]^{-1}$. Заметим, что

$$\nabla f(x^k) - \nabla f(x^{k+1}) = \nabla^2 f(x^{k+1})(x^k - x^{k+1}) + o\left(\left\|x^k - x^{k+1}\right\|\right)$$

Предполагая невырожденной матрицу $\nabla^2 f(x^{k+1})$, отсюда с точностью до членов более высокого порядка малости по сравнению с $\left\|x^k - x^{k+1}\right\|$ получим

$$[\nabla^2 f(x^{k+1})]^{-1} \left(\nabla f(x^{k+1}) - \nabla f(x^k)\right) \approx x^{k+1} - x^k$$

МИСиС

При этом, если $f(x) = (Ax, x)/2 + (b, x)$ — квадратичная функция, A — симметрическая положительно определённая матрица, то $\nabla f(x) = Ax + b$, $\nabla^2 f(x) = A$ и приближенное равенство обращается в точное

$$[\nabla^2 f(x^{k+1})]^{-1} \Delta y^k = \Delta x^k,$$

где $\Delta x^k = x^{k+1} - x^k$, $\Delta y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$. Поэтому естественно потребовать, чтобы для матрицы H_{k+1}, приближающей $[\nabla^2 f(x^{k+1})]^{-1}$, выполнялось условие

$$H_{k+1} \Delta y^k = \Delta x^k \tag{5}$$

Это условие носит название квазиньютоновского. Оно лежит в основе целого ряда методов аппроксимации $[\nabla^2 f(x)]^{-1}$. Соответствующие методы минимизации, для которых на каждом шаге выполняется квазиньютоновское условие, также называются квазиньютоновскими.

Кроме того, удобно получать H_{k+1} как поправку к H_k с помощью матриц первого или второго ранга. Наконец, эти поправки должны быть такими, чтобы для квадратичного случая оказалось $H_k = A^{-1}$. Пусть

$$H_{k+1} = H_k + \Delta H_k.$$

Матрица ΔH_k, обеспечивающая выполнение (5) определяется неоднозначно и возникает семейство методов.

В качестве H_0 можно выбрать любую положительно определённую симметрическую матрицу. На практике часто выбирается единичная матрица.

Квазиньютоновские методы

Метод Давидона-Флетчера-Пауэлла (DFP):

$$H_{k+1} = H_k + \frac{\Delta x_k \, \Delta x_k^\top}{\Delta x_k^\top \, \Delta y_k} - \frac{H_k \, \Delta y_k \, \Delta y_k^\top \, H_k}{\Delta y_k^\top \, H_k \, \Delta y_k}$$

Метод Бройдена:

$$H_{k+1} = H_k + \frac{(\Delta x_k - H_k \, \Delta y_k) \Delta x_k^\top \, H_k}{\Delta x_k^\top \, H_k \, \Delta y_k}$$

Симметричная формула ранга 1 (SR1):

$$H_{k+1} = H_k + \frac{(\Delta x_k - H_k \, \Delta y_k)(\Delta x_k - H_k \, \Delta y_k)^\top}{(\Delta x_k - H_k \, \Delta y_k)^\top \, \Delta y_k}$$

Метод Бройдена-Флетчера-Гольдфарба-Шанно (BFGS):

$$H_{k+1} = \left(I - \frac{\Delta x_k \, \Delta y_k^\mathsf{T}}{\Delta y_k^\mathsf{T} \, \Delta x_k} \right) H_k \left(I - \frac{\Delta y_k \, \Delta x_k^\mathsf{T}}{\Delta y_k^\mathsf{T} \, \Delta x_k} \right) + \frac{\Delta x_k \, \Delta x_k^\mathsf{T}}{\Delta y_k^\mathsf{T} \, \Delta x_k}$$

или

$$H_{k+1} = H_k + \frac{(\Delta x_k^\mathsf{T} \, \Delta y_k + \Delta y_k^\mathsf{T} \, H_k \, \Delta y_k)(\Delta x_k \, \Delta x_k^\mathsf{T})}{(\Delta x_k^\mathsf{T} \, \Delta y_k)^2} - \\ - \frac{H_k \, \Delta y_k \, \Delta x_k^\mathsf{T} + \Delta x_k \, \Delta y_k^\mathsf{T} \, H_k}{\Delta x_k^\mathsf{T} \, \Delta y_k}$$

BFGS — один из наиболее широко применяемых квазиньютоновских методов.

Квазиньютоновские методы

Для квадратичной функции любой квазиньютоновский метод находит минимум за конечное число итераций. Для неквадратичных функций квазиньютоновские методы, естественно, применимы, но перестают быть конечными. В связи с этим при $k > n$ можно либо продолжать счёт по этим же формулам, либо ввести процедуру обновления (заменять матрицу H_k на H_0 через каждые n итераций).

Квазиньютоновские методы чрезвычайно популярны, это объясняется упоминавшимися выше достоинствами методов — они требуют лишь одного вычисления градиента на каждом шаге, в них не нужно обращать матрицу или решать систему линейных уравнений, они обладают глобальной сходимостью, в окрестности решения скорость сходимости высока (часто квадратична) и т. п.

При численной проверке методов обычно наилучшие результаты даёт метод BFGS.

Квазиньютоновские методы

Алгоритм.

1. Задать точность вычислений $\varepsilon > 0$, выбрать начальное приближение x^0 и матрицу $H_0 := I$.

2. Положить $k := 0$.

3. Определить направление поиска $d^k := -H_k \nabla f(x^k)$.

4. Вычислить значение шага $\gamma_k := \arg\min\limits_{\gamma \geqslant 0} f(x^k + \gamma d^k)$.

5. Вычислить точку $x^{k+1} := x^k + \gamma_k d^k$ и градиент $\nabla f(x^{k+1})$.

6. Проверить критерий окончания поиска $\left\|\nabla f(x^{k+1})\right\| < \varepsilon$. Если критерий выполнен, перейти к п. 8.

7. Вычислить $\Delta x^k := x^{k+1} - x^k = \gamma_k d^k$, $\Delta y^k := \nabla f(x^{k+1}) - \nabla f(x^k)$. Вычислить матрицу H_{k+1} по формуле для выбранного метода. Положить $k := k + 1$ и перейти к п. 3.

8. Выбрать приближённо $x^* := x^{k+1}$, $f(x^*) := f(x^{k+1})$. Поиск завершён.

Вопросы?