Сычёв ПМИ-63

Лекция 7.

Многомерная оптимизация. Градиентные методы

7.1. Основные определения

Пусть ф-ия $f(u)$, $u \in U \subseteq R^n$ непрерывно дифф-ма в об. U, то есть ф-ия $f(u)$ диф-ма и все ее частные производные непр. в каждой точке U. Необходимым условием экстремума ф-ии в точке является равенство нулю градиента ф-ии. Точки, в котор. градиент ф-ии равен нулю наз-ся стационарными. Неравный нулю градиент указывает направление наискорейшего возрастания ф-ии. Известно также, что если $f(u)$ - выпукл. ф-ия, то в стац. точке всегда достигается минимум. Рассмотрим методы оптимизации ф-ии неогр. перем., связанные с этим св-ом градиента. Будем считать известно, что в некоторой области ф-ия имеет единств. стац. точку и эта точка явл. точкой минимума.

Опр 7.1.1 Вектор $-grad(f(u)) = -f'(u)$, противоположный градиенту ф-ии в данной точке, наз-ся антиградиентом.

Зам 7.1.2 Ясно, что ненулевой антиградиент указывает напр. наибыстрейшего убывания ф-ии в точке.

Будем теперь искать приближенное значение точки миним. данной ф-ии. Возьмем нач. приближ. - точку u_0. К сожал. не существ. общих правил выбора этой точки. Если из каких-либо соображений можно заключить, что точка минимума находится в некой области, то нач. точку стараются выбрать поближе к этой области.

Будем строить последовательные приближения по правилу:

$$u_{k+1} = u_k - \alpha_k \cdot f'(u_k), \quad k = 0, 1, \dots$$

Число $\alpha_k > 0$ наз-ют длиной шага или шагом градиентного метода. Если на определенном шаге $f'(u_k)=0$, то это стац. точка. Если же $f'(u_k) \neq 0$, то можно выбрать α_k так, чтобы $f(u_{k+1}) < f(u_k)$. Если стац. точка не будет найдена точно, то процесс продолжается, пока не будет выполнено условие:

$$|f'(u_{k+1})| < \varepsilon$$

для наперед заданной константы $\varepsilon > 0$. В зависимости от способа выбора константы α_k существуют различные вар-ты градиентного метода.

Соколов Ананий 911 н-65
мат. анл
стр 2

7.2 Градиентный метод дробления шага

Самым простым вариантом для градиентного метода был бы выбор постоянного шага. Здесь, однако простая проблема: если шаг будет очень мал, то придется сделать достаточно большое число итераций, если же большим взять, то мы рискуем "проскочить" точку минимума. Выходом яв-ся постепенное дробление шага.

А именно, выберем нач знач $\alpha_0 > 0$. На $(k+1)$-ом шаге: вычислим u_{k+1}, проверим условие монотонности $f(u_{k+1}) < f(u_k)$. Если данное усл. выпол-ся, то продолжим вычисление. Если условие не выпол-ся, то примем $\frac{\alpha_k}{m}$ за новое знач-е α_k для некоторого $m \in N$, и еще раз проверим усл-ие монотонности.

Пример:

$$f(x, y) = x^2 + 2y^2; \quad u_0 = (1, 2); \quad \alpha_0 = 0.1$$

$$f(u_0) = 9; \quad f'(u) = (2x; 4y); \quad f'(u_0) = (2; 8);$$

$$u_1 = (1, 2) - 0.1 \cdot (2, 8) = (0.8; 1.2)$$

Получаем:

$$f(u_1) = 3.52 < 9 = f(u_0)$$

Далее:

$$f'(u_1) = (1.6; 4.8) \neq (0, 0); \quad u_2 = (0.64, 0.72)$$

$$f(u_2) = 1.45 < 3.52 = f(u_1)$$

Приблиз. знач точки минимума возьмем $(0.64; 0.72)$, минимум $m_* \approx 1.45$.

7.3 Градиентный метод наискорейшего спуска.

Опишем $(k+1)$-й шаг. На луче $\{u \in R^n \mid u = u_k - \alpha \cdot f'(u_k)$ направленном по анти-градиенту, введем ф-цию 1 перем. $\alpha \geq 0\}$.

$$\varphi_k(\alpha) = f(u_k - \alpha \cdot f'(u_k)), \quad \alpha \geq 0$$

Определим α_k из условий:

$$\varphi_k(\alpha_k) = \inf_{\alpha \geq 0} \varphi_k(\alpha), \quad \alpha_k \geq 0$$

Семёнов 8М-б МА34 лек сг

Если точно вычислять значение пар-ра α невозможно, то можно брать его приближенное значение. Заметим, что минимум ф-ции $\varphi_k(\alpha)$ нужно достигать при $\alpha > 0$. Далее вычисляем u_{k+1}, $f(u_{k+1})$

Заметим, что градиентный метод наискорейшего спуска действительно является эффективным методом нахождения минимума ф-ции. Однако, вычисление значения пар-ра α_k может представлять наибольшую сложность.

Заметим также, что чем ближе поверхности (линии) уровня ф-ции $f(x)$ к сферам (окружностям), тем лучше сходится метод наискорейшего спуска. Однако этот метод и другие вар-ты градиентного спуска медленно сходятся, если поверхности уровня и ~~другие варианты~~ ф-ции сильно вытянуты и функция имеет так назыв. "овражный" хар-р.

Пример:

$f(x,y) = x^2 + 2y^2$; $u_0 = (1,2)$;

$\varphi_0(\alpha) = f(u_0 - \alpha \cdot f'(u_0)) = f((1,2) - \alpha(2,8)) =$

$= (1-2\alpha)^2 + 2(2-8\alpha)^2 = 132\alpha^2 - 68\alpha + 9$

Минимум ф-ции $\varphi_0(\alpha)$ достигается при $\alpha = \frac{68}{2 \cdot 132} \approx 0.26$

Тогда:

$u_1 = (1,2) - 0.26(2,8) = (0.48; -0.08)$;

$f(u_1) \approx 0.24$; $f'(u_1) = (0.96; -0.32)$

Далее:

$\varphi_1(\alpha) = f(u_1 - \alpha \cdot f'(u_1)) = f((0.48; -0.08) - \alpha(0.96; -0.32)) \approx$

$\approx 1.13\alpha^2 - 1.02\alpha + 0.24$

Минимум ф-ции $\varphi_1(\alpha)$ достигается при $\alpha = \frac{1.02}{2 \cdot 1.13} \approx 0.45$. Отсюда

$u_2 = (0.48; -0.08) - 0.45 \cdot (0.96; -0.32) \approx (0.05; 0.06) \approx$

≈ 0.01

$f'(u_2) \approx (0.1; 0.24)$

Приблизительное знач. т. минимума вершины $(0.05; 0.06)$

$m_1 \approx 0.01$.