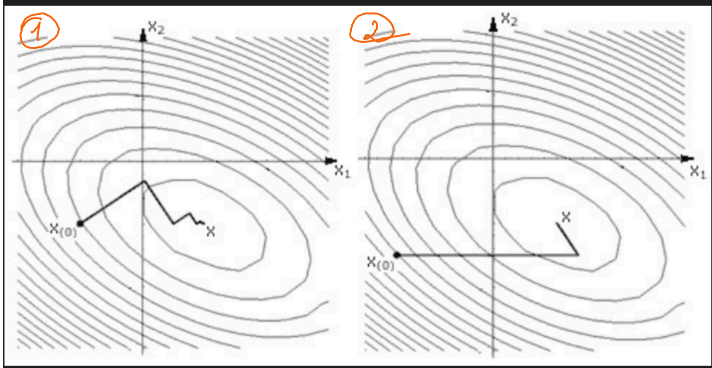
hw_markevich_22

15 декабря 2024 г. 20:20

4. Градиенты градиентами

На этих рисунках изображены траектории движения в точку минимума методами сопряжённых градиентов и наискорейшего спуска:



1 – наискорейший спуск (направления ⊥, сходится дольше)

2 – сопряж. градиент (направления не ⊥, быстро сх-ся)

Определите, на каком из них какой из этих двух методов проиллюстрирован. Ответ аргументируйте. Что изображают замкнутые линии? Докажите(*), что один из этих двух методов (какой?) позволяет решать квадратичные задачи за конечное число шагов. Каково максимально возможное количество этих шагов для матрицы размера n?

Замкнутые линии – это линии уровней функции

Метод сопр. град позволяет решать квадр. задачи за конечное число шагов (∀ задачу приближает квадратичной)

В методе ф-ия приближается квадр формой – матрицей Гессе ⇒ в квадр. случае все просто. Т.к матрица из чисел, то она задает ск. произведение и на каждом шаге мы ищем направление, ⊥-ное всем пред. и min на нем. Значит в $\mathbb{R}^n$ таких не больше n штук ⇒ не более n шагов для квадр. задачи.

По индукции, ∃ для всех n-1 доказано, что $\bar{S}_i^T H \bar{S}_j = 0 \quad \forall i \neq j$

$$S_n = \bar{r}_n + \beta_n \bar{S}_{n-1} \;;\; \bar{r}_k = -\nabla f(\bar{x}_n) \Rightarrow \beta_n = \frac{\|\nabla f(\bar{x}_k)\|^2}{\|\nabla f(\bar{x}_{k-1})\|^2}$$

$$S_n = \bar{r}_n + \frac{\|\nabla f(\bar{x}_n)\|^2}{\|\nabla f(\bar{x}_{n-1})\|^2}\left(\bar{r}_{n-1} + \frac{\|\nabla f(\bar{x}_{n-1})\|^2}{\|\nabla f(\bar{x}_{n-2})\|^2}\bar{S}_{n-2}\right) = -\nabla f(\bar{x}_n) - \|\nabla f(\bar{x}_n)\|^2\left(\frac{\nabla f(\bar{x}_{n-1})}{\|\nabla f(\bar{x}_{n-1})\|^2} + \frac{\bar{S}_{n-2}}{\|\nabla f(\bar{x}_{n-2})\|^2}\right) =$$

$$= -\nabla f(\bar{x}_n) - \|\nabla f(\bar{x}_n)\|^2\left(\frac{\nabla f(\bar{x}_{n-1})}{\|\nabla f(\bar{x}_{n-1})\|^2} + \dots + \frac{f(\bar{x}_0)}{\|\nabla f(\bar{x}_0)\|^2}\right). \quad \exists\; \bar{x}_1 = \bar{x}_0 + \lambda_1 \bar{S}_0$$

$$\nabla f(\bar{x})\big|_{x=x_0} = \nabla f(\bar{x}_0) + H(\bar{x} - \bar{x}_0). \quad \exists\; x = x_1$$

$$\nabla f(\bar{x}_1) - \nabla f(\bar{x}_0) = H(\bar{x}_1 - \bar{x}_0) = \lambda_1 H \bar{S}_0 \xrightarrow[\text{алгоритма}]{\text{исходя из}} (\nabla f(\bar{x}_0), \nabla f(\bar{x}_1)) = 0 \Rightarrow$$

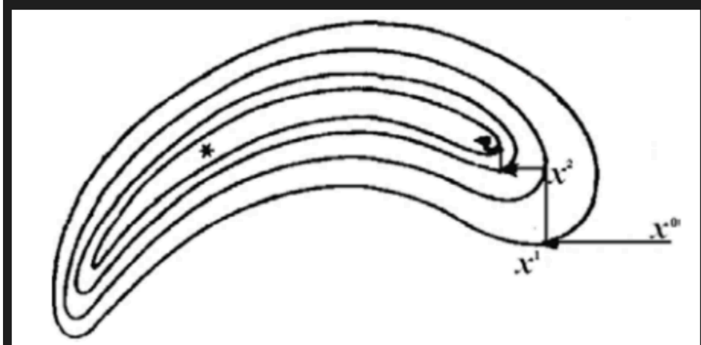
$$\Rightarrow (\nabla f(\bar{x}_1) - \nabla f(\bar{x}_0))^T(\underbrace{-\nabla f(\bar{x}_1)}_{S_0} - \beta_1 \nabla f(\bar{x}_0)) = 0$$

$$\frac{(\nabla f(x_1) - \nabla f(\bar{x}_0))^T}{\lambda_1} = \bar{S}_0^T H^T \qquad (\text{тут } H^T = H \text{ – матрица Гессе})$$

Значит $\bar{S}_0^T H \bar{S}_1 = 0$. Остальные аналогично

5. Овальный

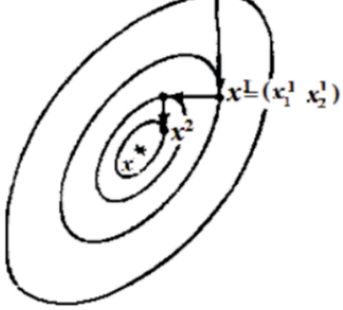
Градиентный метод сходится достаточно быстро, если для минимизируемой функции f (x) поверхности уровня близки к сферам (при n = 2 – к окружностям). Если же линии уровня сильно вытянуты в каком-то направлении, то по нормали к этому направлению целевая функция меняется значительно быстрее, чем вдоль направления. Такой характер целевой функции называется овражным. Исходя из рисунка, объясните, почему в этих случаях градиентный метод сходится хуже.

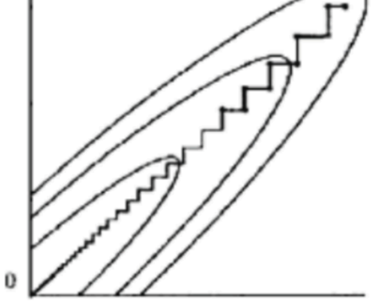


Как связан овражный характер функции с величиной наименьшего сингулярного числа матрицы квадратичной формы? А с величиной числа обусловленности матрицы?

На рисунке ниже изображена работа метода покоординатного спуска (поочередно минимизируем функцию вдоль каждой координаты однопараметрическим методом - совсем убого, поэтому не выдавали) для функций овражного характера. Какой вывод можно сделать из этого рисунка?

Шаг либо const, либо пропорци-





малым градиенту. В I случае вдоль вытянутого направления ф-ции больше пути

Вдоль одного из направлений скорость изменения производной мала $\Longleftrightarrow$ вторая производная мала $\Longleftrightarrow$ наименьшее сингулярное число значительно меньше всех остальных $(\sigma_{max}(H) \gg \sigma_{min}(H))$; $\mu_c(H) = \frac{\sigma_{max}(H)}{\sigma_{min}(H)} \gg 1$

Вывод из рисунка: метод плохой!