Методы оптимизации, HW7

Иванов Вячеслав, группа 699

28 октября 2018 г.

Оглавление

1 Задача 1

Проверьте, что направления в методе сопряжённых градиентов для квадратичной целевой функции и в методе Флетчера-Ривса являются направлениями убывания. Для любой ли стратегии линейного поиска шага в методе Флетчера-Ривса полученное направление будет направлением убывания? Почему?

Даёт ли процедура дробления шага шаг, удовлествoрябщий условию Вольфа? Если нет, то почему и как её нужно модифицировать, чтобы найти шаг, удовлетворяющий условию Вольфа?

Решение. Вспомним, как задаются направления в этих методах:

1. **Метод сопряжённых градиентов для квадратичной целевой функции:**

$$\frac{1}{2}x^\top Ax - b^T x \to \min_x$$
$$x_{k+1} := x_k + \alpha_k p_k$$
$$r_0 := Ax_0 - b,\ r_{k+1} := r_k + \alpha_k A p_k$$
$$p_0 := r_0,\ p_{k+1} := -r_{k+1} + \beta_k p_k$$
$$\alpha_k := \frac{r_k^\top r_k}{p_k^\top A p_k},\ \beta_k := \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}$$

2. **Метод Флетчера-Ривза:** Полностью аналогично, только минимизируется произвольная дифференцируемая функция, а $r_k := \nabla f(x_k)$ и α_k определяется линейным поиском.

$$f(x) \to \min_x$$
$$x_{k+1} := x_k + \alpha_k p_k$$
$$p_0 := \nabla f(x_0),\ p_{k+1} := -\nabla f(x_{k+1}) + \beta_k^{FR} p_k$$
$$\beta_k^{FR} := \frac{\nabla f(x_{k+1})^\top \nabla f(x_{k+1})}{\nabla f(x_k)^\top \nabla f(x_k)}$$

Нужно доказать, что $\forall k:\ f(x_{k+1}) \leqslant f(x_k)$. Подставим выражение для x_{k+1}.

1. **Метод сопряжённых градиентов для квадратичной целевой функции:**

$$\begin{aligned}
f(x_{k+1}) &= \frac{1}{2}x_{k+1}^\top A x_{k+1} - b^T x_{k+1} = \\
&= \frac{1}{2}(x_k + \alpha_k p_k)^\top A(x_k + \alpha_k p_k) - b^T(x_k + \alpha_k p_k) = \\
&= \left(\frac{1}{2}x_k^\top A x_k - b^T x_k\right) + \alpha_k x_k^\top A p_k + \frac{1}{2}\alpha_k^2 p_k^\top A p_k - \alpha_k b^\top p_k = \\
&= f(x_k) + \frac{1}{2}\alpha_k^2 p_k^\top A p_k + \alpha_k r_k^\top p_k = \\
&= f(x_k) + \frac{1}{2}\alpha_k^2 p_k^\top A p_k + \alpha_k r_k^\top(-r_k + \beta_{k-1} p_{k-1}) = \\
&= f(x_k) + \frac{1}{2}\alpha_k^2 p_k^\top A p_k - \alpha_k r_k^\top r_k = \\
&= f(x_k) + \frac{1}{2}\alpha_k \cdot \frac{r_k^\top r_k}{p_k^\top A p_k} p_k^\top A p_k - \alpha_k r_k^\top r_k = \\
&= f(x_k) - \frac{1}{2}\alpha_k r_k^\top r_k < f(x_k)
\end{aligned}$$

2. **Метод Флетчера-Ривза:** Здесь всё зависит от конкретной реализации линейного поиска. Если возможно каким-то образом быстро и точно найти $\arg\min\limits_{\alpha} f(x_k + \alpha p_k)$, то вроде как всё очевидно из определения. В противном случае имеем

$$\nabla f(x_k)^\top p_k = -\|\nabla f(x_k)\|_2^2 + \beta_k^{FR} \nabla f(x_k)^T p_{k-1}$$

Видим, что при $\alpha_{k-1} := \arg\min\limits_{\alpha} f(x_{k-1} + \alpha p_{k-1})$ по необходимому условию получаем, что

$$\nabla f(x_k)^T p_{k-1} = 0$$

Тогда $\nabla f(x_k)^\top p_k < 0$, что, в силу геометрического смысла градиента, указывает: p_k — направление убывания. Тем не менее, возможно себе представить такой выбор шага, при котором β_k^{FR} доминирует и $\nabla f(x_k)^\top p_k > 0$. Тогда p_k направлением убывания не будет по тем же соображениям о смысле градиента.

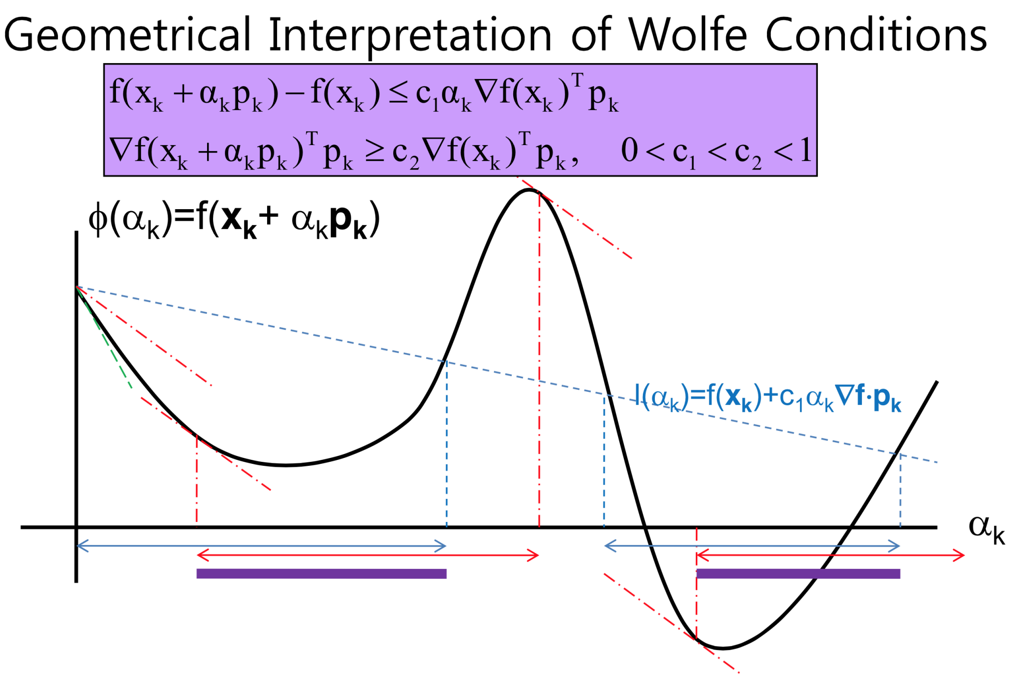
Для ответа на следующий вопрос полезно вспомнить формулировку условий Вольфа.
Для направлений p_k:

$$\begin{aligned}
&\exists c_1, c_2 \forall k : 0 < c_1 < c_2 < 1, \\
&i : f(x_k + \alpha_k p_k) \leqslant f(x_k) + c_1 \cdot \alpha_k p_k^\top \nabla f(x_k) \\
&ii : p_k^\top \nabla f(x_k + \alpha_k p_k) \geqslant c_2 \cdot p_k^\top \nabla f(x_k)
\end{aligned}$$

Первое условие отсекает случаи, когда точка изменилась сильно, а значение функции нет. Второе условие, напротив, не даёт делать слишком маленькие шаги. [1]
Сформулируем, что такое "метод дробления шага":

$$\alpha_k = \alpha \rho^{s_k}, \text{ где } s_k \in \mathbb{N}, \rho \in (0,1) \text{ — такие, что } i, ii \text{ выполняются}$$

а α — некоторое начальное приближение. От его выбора как раз и зависит, получится ли найти подходящее α_k или нет. Рассмотрим на примере картинки, бессовестно украденной с одного блога.



[1] Геометрический смысл условий хорошо расписан в ноутбуке с седьмого семинара, но ещё можно почитать, например, здесь: The Wolfe conditions

Зелёное направление соответствует p_k, красное — ограничению кривизны, синее — необходимому условию убывания. Видно, что если взять α слишком близким к нулю (слева от первой красной проекции на оси абсцисс), то его уменьшением нельзя будет добиться выполнения условий кривизны. Первая идея, которая приходит в голову — завести разные константы u и d для ограничений сверху и снизу, чтобы уметь двигаться и влево, и вправо. Тем не менее, конкретные соотношения между u и d неочевидны, и я пока не знаю, как их подбирать адаптивно, т.к. это может сильно зависеть от специфики задачи. □

2 Задача 2

Доказать, что сопряжённые направления линейно независимы.

Доказательство. Направления $\{p_i\}_{i=1}^n$, сопряжённые относительно $A \in S_{++}^n$ определяются соотношениями $p_i A p_j = 0 \iff i \neq j$. Геометрически это означает, что они ортогональны в метрическом пространстве, для которого A — матрица Грама. Говорят, что такие направления A-*ортогональны*. Покажем, что из A-*ортогональности* следует линейная независимость.
Предположим противное:

$$(1)\ \exists \lambda_1, \dots, \lambda_n : \sum_{i=1}^n \lambda_i^2 \neq 0, \sum_{i=1}^n \lambda_i p_i = 0$$

Будем последовательно домножать (1) слева на $p_j^\top A$.

$$\begin{aligned} &p_j^\top A \sum_{i=1}^n \lambda_i p_i = 0 \\ &\lambda_j p_j^\top A p_j = 0 \\ &\lambda_j = 0 \end{aligned}$$

Отсюда выводим противоречие с $\sum_{i=1}^n \lambda_i^2 \neq 0$. □