Машинное обучение, ММП ВМК МГУ
Заметки о решении задачи условной оптимизации

Липкина Анна

1 Немного определений

Для решения задачи условной оптимизации нам потребуется знать некоторые определения:

Определение 1. Выпуклое множество. Пусть V — векторное пространство. Тогда $C \subseteq V$ — *выпуклое множество* $\Leftrightarrow$

$$\forall\, x, y \in C : \alpha x + (1-\alpha)y \in C \quad \forall\, \alpha \in (0,1).$$

Примеры выпуклых множеств:

1. $\emptyset, V$;
2. $\{x\}$ — множество из одного элемента;
3. Полупространства $H_{+(-)} := \{x \in V : \langle a, x\rangle \geqslant 0 (\leqslant 0)\}, a \in V, a \neq 0$;
4. Шар $B(a, r) := \{x \in V \| x - a\| \leqslant r\}, a \in V, r > 0, \|\cdot\|$ — любая норма в V.

Определение 2. Выпуклая функция. Пусть V — вещественное векторное пространство, $U \subseteq V$ — выпуклое множество, $f : U \to \mathbb{R}$. Тогда *функция* f — *выпуклая* $\Leftrightarrow$

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y) \quad \forall x, y \in U,\ \forall\, \alpha \in [0,1].$$

Примеры выпуклых функций:

1. *Любая линейная функция* f является выпуклой (и вогнутой):
$$f(\alpha x + (1-\alpha)y) = f(\alpha x) + f((1-\alpha)y) = \alpha f(x) + (1-\alpha)f(y).$$
2. Если $f_1, \ldots, f_k$ — выпуклые функции, $\alpha_i \geqslant 0\ \forall\, i = \overline{1,k} \Rightarrow$
$$\sum_{i=1}^{k} \alpha_i f_i(x) \text{ — выпуклая функция.}$$

2 Задача условной оптимизации

Представим, что мы хотим минимизировать некоторую функцию, учитывая некоторые ограничения на значения носителя. Формально, эту **задачу условной оптимизации** можно записать так:

$$P := \begin{cases} f(x) \to \min\limits_{x \in V} \\ g_i(x) \leqslant 0, i = \overline{1, m} \\ h_j(x) = 0, j = \overline{1, p} \end{cases}$$

Замечание: Почему нельзя использовать какие-нибудь градиентные методы? Например, градиентый спуск? Потому что мы легко можем выйти за данные ограничения, если не делать дополнительных приемов проверки.
Замечание: Если в изначальной постановке была задача максимизации

$$f(x) \to max,$$

то можем ее заменить на задачу минимизации:

$$-f(x) \to min.$$

Аналогично поступаем с ограничением $g_i(x) \geqslant 0$. Лучше всего всегда приводить задачу условной оптимизации к виду P, чтобы не думать о различных случаях.

3 Ладно, и что дальше?

Перед тем как решать задачу, разберем все необходимые понятия и теоремы. Затем рассмотрим различные способы решения задачи.

3.1 Понятия, теоремы

Мы должны проверить, что задача P нам попалась "хорошая". Дальше поймем, что же такое "хорошая" задача. Также будем по умолчанию считать, что функция $f(x)$ является гладкой.

Определение 3. Выпуклая задача. Задача P называется *выпуклой*, если верны следующие утверждения:

1. V — выпуклое множество.
2. $f(x), g_i(x)\ \forall\, i = \overline{1, m}$ — выпуклые функции.
3. $h_j(x)\ \forall\, j = \overline{1, p}$ — линейные.

Замечание: Как правило, $V = \mathbb{R}^d$, и поэтому оно является выпуклым множеством. Остальные пункты достаточно легко проверяются.
Введем еще определения.

Определение 4. Допустимое множество.

$$Q := \{x \in V : g_i(x) \leqslant 0, h_j(x) = 0 \quad \forall\, i = \overline{1, m} \,\forall\, j = \overline{1, p}\}$$

Определение 5. Условие регулярности Слейтера (сильное). *Применяется для выпуклых задач оптимизации* P.

$$\exists\, x_0 \in Q : g_i(x_0) < 0 \quad \forall\, i = \overline{1, m}.$$

Заметим, что так как $x_0 \in Q$, то $h_j(x_0) = 0 \,\forall\, j = \overline{1, p}$.

Определение 6. Условие регулярности Слейтера (слабое). *Применяется для выпуклых задач оптимизации* P.

$$\exists\, x_0 \in Q : g_k(x_0) < 0 \quad \forall\, k : g_k - \textbf{НЕ}\text{линейная}.$$

Заметим, что так как $x_0 \in Q$, то $h_j(x_0) = 0 \,\forall\, j = \overline{1, p}$.

Определение 7. Итак, задача P является **хорошей**, если

1. P — выпуклая.
2. Выполнено любое из перечисленных условий регулярности Слейтера[1].

Определение 8. Лагранжиан $\mathcal{L}$ для задачи P выписывается следующим образом:

$$\begin{cases} \mathcal{L}(x, \vec{\lambda}, \vec{\mu}) = f(x) + \sum\limits_{i=1}^{m} \lambda_i g_i(x) + \sum\limits_{j=1}^{p} \mu_j h_j(x) \\ x \in V \subseteq \mathbb{R}^d \text{ (для простоты пока рассматриваем } \mathbb{R}^d) \\ \lambda_i \geqslant 0 \,\forall\, i = \overline{1, m}, \quad \vec{\lambda} = (\lambda_1, \ldots, \lambda_m) \\ \mu_j \in \mathbb{R} \,\forall\, j = \overline{1, p}, \quad \vec{\mu} = (\mu_1, \ldots, \mu_p) \end{cases}$$

Замечания (QA):

1. *Как понять, сколько у нас коэффициентов λ_i и μ_j?* Каждый коэффициент λ_i отвечает за **одно** ограничение g_i из задачи P. Соответственно их столько, сколько ограничений g_i, то есть m штук. И мы не можем считать их все одинаковыми. Априори мы полагаем их **разными.** Аналогично, каждый μ_j отвечается за **одно** ограничение h_j из задачи P. И их тоже столько же, сколько ограничений h_j, то есть p штук. И они так же априори **все разные!**

2. *А если у нас ограничения заданы не по отдельности, а в матричном виде* $Ax = b, A \in \mathbb{R}^{p \times d}, b \in \mathbb{R}^p$, d — *размерность* x? На самом деле это p **ограничений** вида $h_j(x) := \langle a_j, x \rangle - b_j = 0$.

3. *Так, а как запомнить, почему $\lambda_i \geqslant 0$, а $\mu_j \in \mathbb{R}$?* Совсем *нематематическое* правило такое. λ_i у нас соответствуют $g_i(x) \leqslant 0$, и чтобы "скомпенсировать неравенство" , λ_i должны быть $\lambda_i \geqslant 0$. А μ_j соответствуют $h_j(x) = 0$, тут как бы "компенсировать" нечего, и поэтому $\mu_j \in \mathbb{R}$ — любое вещественное число.

[1]На самом деле условий регулярности еще больше, почитать можно, например, тут.

4. *Хм, я вижу что оптимизируемый функционал у меня не зависит в явном виде от вектора, а зависит от нескольких переменных, что мне делать?* Например, пусть $f(v, y, z) = v + 3y + (z-2)^2 \to \min\limits_{v,y,z}$. Можем ввести вектор $\vec{x} := (v, y, z)$, тогда задача станет вида $f(\vec{x}) \to \min\limits_{\vec{x}}$. Причем ограничения не обязательно зависят полностью от x, например, они могут быть:

$$g_1(\vec{x}) = \|\vec{x}\|^2 \leqslant 0, g_2(\vec{x}) = v^2 - y \leqslant 0, h_1(\vec{x}) = z + 2 = 0.$$

Определение 9. Условия Каруша-Куна-Таккера (ККТ) (Необходимое условие!). Пусть в точке $x^* \in Q$ *выполнено условие регулярности (какое-то).* Тогда, если x^* — точка локального минимума $f(x)$, то $\exists\ \vec{\lambda}^* \in \mathbb{R}^m_+, \vec{\mu}^* \in \mathbb{R}^p$ такие, что выполнены условия ККТ:

$$\text{ККТ} := \begin{cases} \nabla_x \mathcal{L}(x, \vec{\lambda}^*, \vec{\mu}^*)|_{x=x^*} = 0 \text{ — \textit{условие стационарности}} \\ \lambda_i^* \geqslant 0 \ \forall\ i = \overline{1,m} \text{ — \textit{да, это просто формализация условий на лямбды}} \\ g_i(x^*) \leqslant 0 \ \forall\ i = \overline{1,m} \text{ — \textit{это формализация того, что} } x^* \in Q \\ h_j(x^*) = 0 \ \forall\ j = \overline{1,p} \text{ — \textit{и это формализация того, что} } x^* \in Q \\ \lambda_i^* g_i(x^*) = 0 \ \forall\ i = \overline{1,m} \text{ — \textit{а это условие дополняющей нежесткости}} \end{cases}$$

Замечание: Лагранжиан является функционалом, а x — вектором, поэтому производная Лагранжиана по x будет градиентом — тоже вектором. Что значит, что градиент равен 0? Это значит, что каждая компонента градиента равна 0! А теперь вспоминаем, что выше мы приводили пример про то, что делать, если функция зависит от нескольких переменных (мы объединяли переменные в один вектор). Также вспоминаем, что градиент — это вектор частных производных. Поэтому, условие стационарности становится совсем просто выписать: достаточно написать d уравнений, где производные Лагранжиана по x_i-й компоненте равны 0. В случае 4 выше, например, $\frac{\partial \mathcal{L}}{\partial v} = 0, \frac{\partial \mathcal{L}}{\partial y} = 0, \frac{\partial \mathcal{L}}{\partial z} = 0.$

А теперь, внимание, фокус!

Определение 10. Теорема ККТ (Достаточное условие!). Если P — выпуклая задача оптимизации и для набора $(x^*, \vec{\lambda}^*, \vec{\mu}^*)$ выполнены все условия ККТ, то x^* — точка глобального минимума.

Замечания (QA):

1. *Погодите, но у нас же Слейтер и так применяется к выпуклым задачам, зачем нам еще раз требовать выпуклость?* В условиях ККТ фигурирует фраза "выполнено какое-то условие регулярности" . Мы для простоты сейчас разобрали только два условия Слейтера, которые действительно применимы для выпуклых задач. А вообще, существуют и другие условия регулярности, которые не обязательно применяются к выпуклым задачам.

2. *А где здесь двойственная функция вообще?* Ее здесь, действительно, в явном виде нет. Есть только двойственные переменные.

Собственно, что нам делать-то с этой теоремой? Выпишем алгоритм действий:

1. Приводим задачу условной оптимизации к каноническому виду P.
2. Проверяем, является ли она выпуклой.
3. Проверяем, выполняется ли хотя бы одно условие регулярности (в нашем случае — одно из сильного или слабого Слейтера).
4. Выписываем условия ККТ.
5. Если задача является "хорошей" , то ККТ становятся достаточными условиями для поиска глобального минимума.
6. А вот тут самое веселое. Если нам надо просто найти минимум функции, то можно попробовать угадать решение $(x^*, \vec{\lambda}^*, \vec{\mu}^*)$, которое удовлетворяет ККТ. Причем на любом найденном решении (если их несколько), которые удовлетворяют ККТ, будет достигаться глобальный минимум функции (кстати, подумайте, какие бывают выпуклые гладкие функции с несколькими глобальными минимумами). Если решение не угадывается, то пытаемся решить полученную систему. Очень хорошо в данном случае помогают условия дополняющей нежесткости. Ведь если произведение множителей равно 0, то какое-то из множетелей само равно 0!

3.2 Пример

$$\begin{cases} (x-4)^2 + (y-4)^2 \to \min\limits_{x,y} \\ x + y \leqslant 4 \\ x + 3y \leqslant 9 \end{cases}$$

1. Задача уже приведена к каноническому виду.
2. Очевидно, она является выпуклой.
3. Сильное условие Слейтера выполнено. Например, возьмем точку $(1, 1)$, тогда неравества из ограничений станут строгими. А можно было сказать проще: выполнено слабое условие Слейтера, так как все ограничения — линейные функции.
4. Задача "хорошая" , а значит — можно выписать условия ККТ и найти все решения — они и будут точками глобального минимума!

Лагранжиан:

$$L(x, y, \lambda_1, \lambda_2) = (x-4)^2 + (y-4)^2 + \lambda_1(x + y - 4) + \lambda_2(x + 3y - 9)$$

Условия ККТ:

$$\begin{cases} 2(x-4)+\lambda_1+\lambda_2=0 \\ 2(y-4)+\lambda_1+3\lambda_2=0 \\ x+y\leqslant 4, \lambda_1 \geqslant 0, \lambda_1(x+y-4)=0 \\ x+3y\leqslant 9, \lambda_2 \geqslant 0, \lambda_2(x+3y-9)=0 \end{cases}$$

Обратим внимание на условие дополняющей нежесткости, они являются мощным инструментов для решения уравнений.

1. Так как $\lambda_1(x+y-4)=0$, то тут возможны два случая:
   (a) $\lambda_1=0, x+y<4$ (из ограничений).
   (b) $\lambda_1\geqslant 0, x+y=4$.
2. Аналогично с $\lambda_2(x+3y-9)=0$:
   (a) $\lambda_2=0, x+3y<9$.
   (b) $\lambda_2\geqslant 0, x+3y=9$.

То есть всего получается 4 случая, которые нужно разобрать. Разбор каждого из них и нахождение решений предоставляется читателю.