Математическое моделирование и методы оптимизации
Часть 4.

Тремба Андрей Александрович
Москва, к.ф.-м.н., с.н.с.
Институт проблем управления РАН

21 июня 2019 г., Иннополис

Четвёртая часть

Задачи с ограничениями

- Типы ограничений
- «Простые» ограничения
- Некоторые трюки
- Метод множителей Лагранжа
- Линейное и квадратичное программирование
- CVX

Ограничения: запись и определения.

Постановка задачи

$$
\begin{aligned}
\min_{\substack{x \in Q \\ g_i(x) \leq 0 \\ h_j(x) = 0}} \quad & f(x)
\end{aligned}
$$

Допустимое множество:

$$
\mathbb{X} = \left\{ x \in \mathbb{R}^n : \begin{array}{c} x \in Q \\ g_i(x) \leq 0 \\ h_j(x) = 0 \end{array} \right\}
$$

Сложности

Если $\mathbb{X} \neq \mathbb{R}^n$, меняется всё:

- Определения минимума
- Свойства задачи
- Условия оптимальности
- Методы

Определения

- x^* является локальным минимумом, если $\exists \varepsilon > 0$, такое что

$$f(x) \geq f(x^*), \quad \forall x \in \mathbb{X} : \|x - x^*\| \leq \varepsilon$$

- x^* называется глобальным минимумом, если

$$f(x) \geq f(x^*), \quad \forall x \in \mathbb{X}$$

Как множество влияет на минимум.

Два множества ограничений

$$\min_{x \in \mathbb{X}} f(x)$$

Lemma

Если $\mathbb{Y} \subseteq \mathbb{X}$, *то*

$$\min_{x \in \mathbb{Y}} f(x) \overset{?}{\lesseqgtr} \min_{x \in \mathbb{X}} f(x)$$

Два множества ограничений

$$\min_{x \in \mathbb{X}} f(x)$$

Lemma

Если $\mathbb{Y} \subseteq \mathbb{X}$, *то*

$$\min_{x \in \mathbb{Y}} f(x) \geq \min_{x \in \mathbb{X}} f(x)$$

Таблица

	$f(x)$	$f_1(x) \geq f_2(x)$
$\mathbb{X}$	$\min\limits_{x \in \mathbb{X}}$	$\min\limits_{x \in \mathbb{X}} f_1(x) \geq \min\limits_{x \in \mathbb{X}} f_2(x)$
$\mathbb{Y} \subseteq \mathbb{X}$	$\min\limits_{x \in \mathbb{Y}} f(x) \geq \min\limits_{x \in \mathbb{X}} f(x)$	$\min\limits_{x \in \mathbb{Y}} f_1(x) \geq \min\limits_{x \in \mathbb{Y}} f_2(x)$

Основная проблема условной оптимизации

Для задачи

$$\min_{x \in \mathbb{X}} f(x)$$

условие

$$\nabla f(x) = 0$$

или

$$\partial f(x) \ni 0$$

уже не является необходимым условием!

Даже для выпуклых задач*.

Влияние множества на выпуклость

$\mathbb{X} \setminus f(\cdot)$	выпуклая	невыпуклая
выпуклое	**выпуклая задача**	невыпуклая
невыпуклое	невыпуклая	невыпуклая

Влияние множества на выпуклость

$\mathbb{X} \setminus f(\cdot)$	выпуклая (на $\mathbb{X}$)	невыпуклая
выпуклое	**выпуклая задача**	невыпуклая
невыпуклое	невыпуклая	невыпуклая

Необходимые условия оптимальности

$$\min_{x \in \mathbb{X}} f(x)$$

Теорема

Пусть $f(x)$ дифференцируема и $\mathbb{X}$ выпуклое. Если x^* локальный минимум, то

$$(\nabla f(x^*), x - x^*) \geq 0, \quad \forall x \in \mathbb{X}$$

Необходимые условия оптимальности

$$\min_{x \in \mathbb{X}} f(x)$$

Теорема

Пусть $f(x)$ дифференцируема и $\mathbb{X}$ выпуклое. Если x^* локальный минимум, то

$$(\nabla f(x^*), x - x^*) \geq 0, \quad \forall x \in \mathbb{X}$$

Следствие

Если точка минимума x^* - внутренняя точка $\mathbb{X}$, то

$$\nabla f(x^*) = 0$$

Выпуклая задача оптимизации

Выпуклое $\mathbb{X}$, функция $f(x)$ выпуклая (хотя бы на $\mathbb{X}$):

$$\min_{x \in \mathbb{X}} f(x)$$

Следствие: достаточное условие

Если $\nabla f(x^*) = 0$, то x^* является локальным минимумом.

Выпуклая задача оптимизации

Выпуклое $\mathbb{X}$, функция $f(x)$ выпуклая (хотя бы на $\mathbb{X}$):

$$\min_{x \in \mathbb{X}} f(x)$$

Следствие: достаточное условие

Если $\nabla f(x^*) = 0$, то x^* является глобальным минимумом.

Теорема

Локальный минимум x^* выпуклой задачи

$$f(x) \geq f(x^*), \; \forall x \in \mathbb{X}, \|x - x^*\| \leq \varepsilon$$

является и глобальным минимумом

$$f(x) \geq f(x^*), \; \forall x \in \mathbb{X}$$

Типы ограничений

Ограничения в виде «простого» множества

Пусть $\mathbb{X} = Q$ – «простое»

$$
\begin{aligned}
\min_{x \in Q} \quad & f(x) \\
\cancel{g_i(x) \leq 0} & \\
\cancel{h_j(x) = 0} &
\end{aligned}
$$

- «удаление» множества ограничений
- использование «простоты» Q

«Удалиние» ограничения-множества

«Спрятать» в функцию с расширенным множеством значений:

$$\min_{x \in Q} f(x) \Leftrightarrow \min_x \left(f(x) + \mathbb{I}_Q(x) \right) = \min_x g(x)$$

«Удалиние» ограничения-множества

«Спрятать» в функцию с расширенным множеством значений:

$$\min_{x\in Q} f(x) \Leftrightarrow \min_x \left(f(x) + \mathbb{I}_Q(x)\right) = \min_x g(x)$$

Использовать барьеры (внутренние или внешние)

$$\min_{x\in Q} f(x) \approx \min_x \left(f(x) + mH(x)\right)$$

Эквивалентность при приделе: $m \to \ldots$

Ограничение в виде простого множества

Идея: делать шаги и оставаться в Q.

$$x^k \in Q \to \overline{x}^{k+1} \to x^{k+1} \in Q$$

Ограничение в виде простого множества

Идея: делать шаги и оставаться в Q.

$$x^k \in Q \to \overline{x}^{k+1} \to x^{k+1} \in Q$$

Взять ближайшую точку из Q!

$$Proj_Q(x) = \arg\min_{y \in Q} \|x - y\|$$

Ограничение в виде простого множества

Идея: делать шаги и оставаться в Q.

$$x^k \in Q \to \overline{x}^{k+1} \to x^{k+1} \in Q$$

Взять ближайшую точку из Q!

$$Proj_Q(x) = \arg\min_{y \in Q} \|x - y\|$$

$$x^{k+1} = \mathrm{Proj}_Q(\overline{x}^{k+1})$$

Метод проекции градиента

Шаг градиентного метод + проекция

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

$$x^{k+1} = \mathrm{Proj}_Q(\overline{x}^{k+1})$$

or

$$x^{k+1} = \mathrm{Proj}_Q(x^k - \gamma_k \nabla f(x^k))$$

Также: субградиентный метод с проектированием.

$$x^{k+1} = \mathrm{Proj}_Q(x^k - \gamma_k \partial f(x^k))$$

Метод условного градиента (метод Франк-Вульфа)

$$\min_{x \in Q} f(x)$$

Вспомогательная задача

$$\min_{x \in Q}(c, x) = -\max_{x \in Q}(-c, x)$$

Метод условного градиента (метод Франк-Вульфа)

$$\min_{x \in Q} f(x)$$

Вспомогательная задача

$$\min_{x \in Q} (c, x) = -\max_{x \in Q} (-c, x)$$

1. $$\overline{x}^k = \arg\min_{x \in Q} (\nabla f(x^k), x)$$
2. $$\gamma_k = \arg\min_{\gamma \in [0,1]} f(x^k + \gamma(\overline{x}^k - x^k))$$
3. $$x^{k+1} = (1 - \gamma_k) x^k + \gamma_k \overline{x}^k$$

Метод секущей плоскости

1. Метод центра тяжести
2. Метод эллипсоидов

$$\min_{\cancel{x \in Q}} \qquad f(x)$$

$$g_i(x) \leq 0, \; i = 1, \ldots, m$$
$$h_j(x) = 0, \; j = 1, \ldots, \ell$$

$$
\begin{aligned}
\min_{\cancel{x \in Q}} \quad & f(x) \\
g_i(x) \le 0,\ & i = 1, \ldots, m \\
h_j(x) = 0,\ & j = 1, \ldots, \ell
\end{aligned}
$$

Выпуклая задача с ограничениями $f(x)$, $g_i(x)$ выпуклые, $h_j(x)$ аффинные:

$$
\begin{aligned}
\min \quad & f(x) \\
g_i(x) \le 0 & \\
{\color{blue} A_j x - b_j = 0} &
\end{aligned}
$$

«Удаление» функциональных ограничений

Как «спрятать» несколько ограничений в одно:

$$g_i(x) \leq 0 \Leftrightarrow g(x) \leq 0$$

$$g(x) \doteq \max_{i=1,\ldots,m} g_i(x)$$

«Удаление» функциональных ограничений

Как «спрятать» несколько ограничений в одно:

$$g_i(x) \le 0 \Leftrightarrow g(x) \le 0$$

$$g(x) \doteq \max_{i=1,\dots,m} g_i(x)$$

Барьеры в явном виде (увеличение M):

$$\min_x \left(f(x) + M_g \sum_i [g_i(x)]_+ + M_h \sum_j h_j(x)^2 \right)$$

Как избавиться от линейных ограничений-равенств

$$
\begin{aligned}
\min_{\substack{x \in Q \\ g_i(x) \le 0 \\ Ax = b,\ A \in \mathbb{R}^{k \times n}}} \quad & f(x)
\end{aligned}
$$

$$
\{x \in \mathbb{R}^n : Ax = b\} =^* \{Fz + \widehat{x} : z \in \mathbb{R}^{n-k}\}
$$

Как избавиться от линейных ограничений-равенств

$$\begin{aligned} \min_{\substack{x \in Q \\ g_i(x) \le 0 \\ Ax = b,\ A \in \mathbb{R}^{k \times n}}} \quad & f(x) \end{aligned}$$

$$\{x \in \mathbb{R}^n : Ax = b\} =^* \{Fz + \widehat{x} : z \in \mathbb{R}^{n-k}\}$$

$$\min_{\substack{Fz + \widehat{x} \in Q \\ g_i(Fz + \widehat{x}) \le 0}} f(Fz + \widehat{x}) \Leftrightarrow \min_{\substack{Fz + \widehat{x} \in Q \\ \widehat{g}_i(z) \le 0}} \widetilde{f}(z)$$

Необходимые условия оптимальности

Для ограничений-равенств

$$\min f(x)$$
$$h_j(x) = 0$$
$$j = 1, \ldots, \ell$$

Каруш-Кун-Таккер, ККТ

Пусть $f(\cdot), h_j(\cdot)$ дифференцируемые Если x^* - регулярный[a] минимум, то существуют $\nu^* \in \mathbb{R}^\ell$:

$$\nabla f(x^*) + \langle \nu^*, \nabla h(x^*) \rangle = 0$$

$$\nabla f(x^*) + \sum_{j=1}^{\ell} \nu_j^* \nabla h_j(x^*) = 0$$

[a] $h_j(x^*) = 0$ и $\nabla h_j(x^*)$ линейно независимы

Промежуточные итоги

1. Выпуклые и невыпуклые задачи с ограничениями
2. Несколько методов: проекции градиента, условного градиента (Франк-Вульфа), метод секущих плоскостей и барьеры.
3. Ограничения в виде множеств
4. Ограничения в виде функций
5. Необходимые и достаточные условия

Функция Лагранжа, множители Лагранжа

Встречайте!

$$\begin{aligned} &\min \quad f(x) \\ &h_j(x) = 0 \\ &j = 1, \ldots, \ell \end{aligned}$$

Функция Лагранжа:

$$L(x, \nu) = f(x) + (\nu, h(x)) = f(x) + \sum_{j=1}^{\ell} \nu_j h_j(x)$$

Встречайте!

$$\min_{\substack{h_j(x)=0\\ j=1,\ldots,\ell}} f(x)$$

Функция Лагранжа:

$$L(x,\nu) = f(x) + (\nu, h(x)) = f(x) + \sum_{j=1}^{\ell} \nu_j h_j(x)$$

Необходимые условия для того, чтобы регулярная* точка x^* была минимумом - это существование ν^*:

$$\nabla_x L(x^*,\nu^*) = 0, \quad \nabla_\nu L(x^*,\nu^*) = 0$$

Функция Лагранжа в общем виде

$$\begin{aligned} \min \quad & f(x) \\ g_i(x) \leq 0 & \\ h_j(x) = 0 & \end{aligned}$$

Функция Лагранжа:

$$L(x,\lambda,\nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \nu_j h_j(x)$$

Функция Лагранжа в общем виде

$$\begin{aligned} \min \quad & f(x) \\ g_i(x) \leq 0 & \\ h_j(x) = 0 & \end{aligned}$$

Функция Лагранжа:

$$L(x,\lambda,\nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \nu_j h_j(x)$$

$$= f(x) + (\lambda, g(x)) + (\nu, h(x))$$

Условия ККТ (в т.ч. для невыпуклых)

Если x^* - локальный регулярный[1] минимум, то существуют множители Лагранжа $\lambda_i^* \geq 0$ и ν_j^*, такие, что $\lambda_i^* g_i^*(x^*) = 0$, и

$$L'_x(x^*, \lambda^*, \nu^*) =$$

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x) + \sum_{j=1}^{\ell} \nu_j^* \nabla h_j(x) = 0$$

[1] Все градиенты $\{\nabla h_j(x^*), \nabla g_i(x^*), \ \ j = 1, \ldots \ell, i : g_i(x^*) = 0\}$ линейно независимы, включая градиенты и ограничений-равенств, и ограничений-неравенств.

Системная работа с ограничениями-функциями. Двойственность.

Двойственность

Исходная («прямая») задача и её решение:

$$p^* \doteq \min_{\substack{g_i(x) \le 0 \\ h_j(x) = 0}} f(x)$$

Её функция Лагранжа

$$L(x, \lambda, \nu) = f(x) + (\lambda, g(x)) + (\nu, h(x)) =$$
$$= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \nu_j h_j(x)$$

Двойственная задача

Двойственная функция (определение):

$$d(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

Что можно сказать о выпуклости?

Двойственная задача

$$\sup_{\lambda_i \geq 0, \nu_i} d(\lambda, \nu)$$

Слабая двойственность

Решение двойственной задачи

$$q^* \doteq \sup_{\lambda_i \geq 0, \nu_i} d(\lambda, \nu)$$

не превосходит решения прямой задачи

$$p^* \doteq \min_{\substack{g_i(x) \leq 0 \\ h_j(x) = 0}} f(x)$$

Слабая двойственность

$$q^* \leq p^*$$

«Седловая» форма слабой двойственности

$$\begin{aligned} \min \quad & f(x) \\ g_i(x) \leq 0 & \\ h_j(x) = 0 & \end{aligned}$$

Для любых $f(\cdot), g_i(\cdot), h_j(\cdot)$:

$$\max_{\lambda_i \geq 0, \nu_i} \min_x L(x, \lambda, \nu) = q^* \leq p^* = \min_x \max_{\lambda_i \geq 0, \nu_i} L(x, \lambda, \nu)$$

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \nu_j h_j(x)$$

Зачем решать двойственные задачи

$$q^* = \sup_{\lambda_i \geq 0, \nu_i} d(\lambda, \nu) \leq p^*$$

+ Двойственная задача «простая», в том числе по числу переменных

Зачем решать двойственные задачи

$$q^* = \sup_{\lambda_i \geq 0, \nu_i} d(\lambda, \nu) \leq p^*$$

- Двойственная задача «простая», в том числе по числу переменных
- Она доставляет нижнюю границу решения прямой задачи.

Зачем решать двойственные задачи

$$q^* = \sup_{\lambda_i \geq 0, \nu_i} d(\lambda, \nu) \leq p^*$$

+ Двойственная задача «простая», в том числе по числу переменных

+ Она доставляет нижнюю границу решения прямой задачи.

– Необходимо вычислять двойственную функцию $d(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$,

Сильная двойственность

Если прямая задача выпукла и выполнены условия регулярности[2], то выполняется сильная двойственность

$$q^* = \sup_{\lambda_i \geq 0, \nu_i} d(\lambda, \nu) = \min_{\substack{g_i(x) \leq 0 \\ h_j(x) = 0}} f(x) = p^*$$

[2]В выпуклом случае они проще: существует $x^a : h_j(x^a) = 0$ и $g_i(x^a) < 0$ (условие Слейтера)

Сильная двойственность

Если прямая задача выпукла и выполнены условия регулярности[2], то выполняется сильная двойственность

$$q^* = \sup_{\lambda_i \geq 0, \nu_i} d(\lambda, \nu) = \min_{\substack{g_i(x) \leq 0 \\ h_j(x) = 0}} f(x) = p^*$$

+ Решение прямой задачи x^* может быть восстановлено* из решения двойственной задачи λ^*, ν^*.

[2] В выпуклом случае они проще: существует $x^a : h_j(x^a) = 0$ и $g_i(x^a) < 0$ (условие Слейтера)

Сильная двойственность: другая характеризация

Сильная двойственность:

$$\max_{\lambda_i \geq 0, \nu_i} \min_x L(x, \lambda, \nu) = q^* = p^* = \min_x \max_{\lambda_i \geq 0, \nu_i} L(x, \lambda, \nu)$$

означает существование седловой точки:

$$L(x^*, \lambda, \nu)\Big|_{\lambda \geq 0} \leq L(x^*, \lambda^*, \nu^*) \leq L(x, \lambda^*, \nu^*)$$

+ можно использовать алгоритмы нахождения седловой точки функции,
+ можно упростить задачу

Использование двойственности Лагранжа

1. Сформировать функцию Лагранжа
2. Найти двойственную функцию
3. Решить двойственную задачу
4. Восстановить решение прямой задачи (если возможно)

Выводы об ограничениях

- Ограничения усложняют жизнь
- Простые способы работы с ограничениями: преобразования и алгоритмы
- Множители Лагранжа и функция Лагранжа
- Двойственная функция,
  сильная двойственность

Разные функции Лагранжа. Выбор ограничений

Ограничения разных типов

$$\begin{aligned} \min_{\substack{x \in Q_k \qquad k = 1, \dots \\ g_i(x) \le 0 \qquad i = 1, \dots \\ h_j(x) = 0 \qquad j = 1, \dots}} \quad f(x) \end{aligned}$$

Двойственная функция

$$d(\lambda, \nu) = \inf_{x \in \cap_k Q_k} L(x, \lambda, \nu)$$

Ограничения разных типов

$$
\begin{aligned}
\min_{\substack{x \in Q_k \quad k = 1, \ldots \\ g_i(x) \le 0 \quad i = 1, \ldots \\ h_j(x) = 0 \quad j = 1, \ldots \\ x \in \operatorname{\mathbf{dom}} f \\ x \in \operatorname{\mathbf{dom}} g_i \quad i = 1, \ldots \\ x \in \operatorname{\mathbf{dom}} h_j \quad j = 1, \ldots}} \quad f(x)
\end{aligned}
$$

Двойственная функция

$$
d(\lambda, \nu) = \inf_{x \in \cap_k Q_k} L(x, \lambda, \nu) = \inf_{x \in \cap_k Q_k \cap \operatorname{\mathbf{dom}} L} L(x, \lambda,
$$

$$
= \inf_{x \in \cap_k Q_k \cap \operatorname{\mathbf{dom}} f \cap_i \operatorname{\mathbf{dom}} g_i \cap_j \operatorname{\mathbf{dom}} h_j} L(x, \lambda, \nu)
$$

Выбор ограничений

$$
\begin{aligned}
\min_{\substack{x \in Q \\ g_i(x) \le 0 \\ h_j(x) = 0}} \quad & f(x)
\end{aligned}
$$

Ограничения, заданные множеством(-ами)

$$\Updownarrow$$

Ограничения, заданные функциями

Разные двойственные задачи

$$\max_{\substack{\nu \\ \lambda \geq 0}} \quad d(\lambda, \nu)$$

Ключевое свойство: некоторые функции-ограничения могут быть «спрятаны» в Q (и наоборот): $Q_h = Q \cap \{x : h_j(x) = 0\}$

$$L_h(x, \lambda) = f(x) + \sum_{i=\overline{1,m}} \lambda_i g_i(x),$$

$$d_h(\lambda) = \min_{x \in Q_h} L_h(x, \lambda),$$

$$\max_{\lambda \geq 0} \quad d_h(\lambda)$$

Разные двойственные задачи

$$\max_{\substack{\nu \\ \lambda \geq 0}, (\nu, \lambda) \in \mathbf{dom}\, d} d(\lambda, \nu)$$

Ключевое свойство: некоторые функции-ограничения могут быть «спрятаны» в Q (и наоборот): $Q_h = Q \cap \{x : h_j(x) = 0\}$

$$L_h(x, \lambda) = f(x) + \sum_{i=\overline{1,m}} \lambda_i g_i(x),$$

$$d_h(\lambda) = \min_{x \in Q_h} L_h(x, \lambda),$$

$$\max_{\lambda \geq 0, \lambda \in \mathbf{dom}\, d_h} d_h(\lambda)$$

Использование двойственности Лагранжа

1. Привести целевую функцию и/или ограничения к удобной форме
2. Выбрать, какие функциональные граничения войдут в функцию Лагранжа
3. Сформировать функцию Лагранжа
4. Найти двойственную функцию (решить задачу без ограничений)
5. Поставить и решить двойственную задачу
6. Восстановить решение прямой задачи

Полезные свойства функции Лагранжа

- Поиск решения прямой задачи как $x^* = \arg\min_Q L(x, \lambda^*, \nu^*)$
- Убрать ограничения из условия дополняющей нежёсткости $\lambda_i^* g_i(x^*) = 0$
- Анализ чувствительности
- Связь с регуляризацией
- Связь с «овыпуклением»

Анализ чувствительности

$$p^*(u,v) = \min_{\substack{g_i(x) \le u_i \\ h_j(x) = v_j}} f(x)$$

$$p^*(u,v) \ge p^*(0,0) - (\lambda^*, u) - (\nu^*, v)$$

- большое λ^*, уменьшаем $u_i < 0 \to$ рост p^*
- большое $\nu^* \gg 0$ и $v_i < 0$, либо
  большое $\nu^* \ll 0$ и $v_i > 0 \to$ рост p^*
- малое λ^*, увеличиваем $u_i > 0$
- малое $\nu^* > 0$ и $v_i > 0$, либо
  малое $\nu^* < 0$ и $v_i < 0$

Анализ чувствительности, продолжение

Сильная выпуклость и дифференцируемость

$$p^*(u, v) = \min_{\substack{g_i(x) \le u_i \\ h_j(x) = v_j}} f(x)$$

$$\frac{\partial p^*(0,0)}{\partial u_i} = -\lambda_i^*$$

$$\frac{\partial p^*(0,0)}{\partial v_i} = -\nu_i^*$$

Связь с регуляризацией

Штрафной член для $\min_{x \in Q} f(x)$:

$$f_1^*(c) = \min_x \left(f(x) + c\|x\| \right)$$

(SVM, Lasso, TV)

Связь с регуляризацией

Штрафной член для $\min_{x \in Q} f(x)$:

$$f_1^*(c) = \min_x \left(f(x) + c\|x\| \right)$$

(SVM, Lasso, TV)

Условная оптимизация

$$f_2^*(r) = \min_{\|x\| \le r} f(x)$$

$$L(x, \lambda) = f(x) + \lambda(\|x\| - r)$$

Связь с регуляризацией

Штрафной член для $\min_{x \in Q} f(x)$:

$$f_1^*(c) = \min_x \left(f(x) + c\|x\| \right)$$

(SVM, Lasso, TV)

Условная оптимизация

$$f_2^*(r) = \min_{\|x\| \leq r} f(x)$$

$$L(x, \lambda) = f(x) + \lambda(\|x\| - r)$$

$$x_1^*(c) = \arg\min_x \left({\color{blue} f(x)} + {\color{red} c}{\color{blue}\|x\|} \right)$$

Поиск седловой точки

Пример задачи поиска седловой точки

$$\max_{\lambda \geq 0, \nu} L(x^*, \lambda, \nu) = L(x^*, \lambda^*, \nu^*) = \min_{x \in Q} L(x, \nu^*, \lambda^*)$$

$$f(x) \to \min_{g_i(x) \leq 0}$$

$$\max_{\lambda \geq 0} L(x^*, \lambda) = L(x^*, \lambda^*) = \min_{x \in Q} L(x, \lambda^*)$$

Идея: чередовать шаги по градиенту и антиградиенту.

Двойственные методы

Решение задачи оптимизации с помощью задачи на поиск седловой точки

$$f(x) \to \min_{g_i(x) \leq 0}$$

Метод Эрроу-Гурвица-Удзавы:

$$\begin{aligned} x^{k+1} &= x^k - \gamma \nabla_x L(x^k, \lambda^k) = \\ &= x^k - \gamma (\nabla f(x^k) + \sum_i \lambda_i^k \nabla g_i(x^k)) \\ \lambda^{k+1} &= [\lambda^k + \gamma \nabla_\lambda L(x^k, \lambda^k)]_+ = \\ &= [\lambda^k + \gamma g(x^k)]_+ \end{aligned}$$

Есть проблемы со сходимостью!

«Рабочая» модификация

Градиентный (двойственный) подъём

$$x^{k+1} = \arg\min_x L(x, \lambda^k)$$

$$\lambda^{k+1} = [\lambda^k + \gamma g(x^k)]_+$$

Модифицированная функция Лагранжа

$$f(x) \to \min_{Ax=c}$$

$$L_\rho(x, \nu) = f(x) + (\nu, Ax - c) + \frac{\rho}{2}\|Ax - c\|^2$$

Градиентный двойственный подъём

$$x^{k+1} = \arg\min_x L_\rho(x, \nu^k)$$

$$\nu^{k+1} = \nu^k + \rho(Ax^{k+1} - c)$$

Частный случай: сепарабельные переменные

$$f(x) + g(z) \to \min_{Ax+Bz=c}$$

Модифицированная функция Лагранжа

$$L_\rho(x, z, \nu) = f(x) + g(z) + (\nu, Ax + Bz - c) +$$
$$+ \frac{\rho}{2}\|Ax + Bz - c\|^2$$

Метод множителей Лагранжа с переменой направления
(Alternating Direction Method of Multipliers)

$$\begin{cases} x^{k+1} = \arg\min_x \left(f(x) + \frac{\rho}{2}\|Ax + Bz^k - c + u^k\|_2^2\right) \\ z^{k+1} = \arg\min_z \left(g(z) + \frac{\rho}{2}\|Ax^{k+1} + Bz - c + u^k\| \right. \\ u^{k+1} = u^k + Ax^{k+1} + Bz^{k+1} - c \end{cases}$$

Связан с двойственным подъёмом со специальным шагом.

Пример ADMM: «сепарация» функции

$$\min_x (f(x) + g(x)) \Leftrightarrow \min_{x=z} (f(x) + g(z))$$

$$\begin{cases} x^{k+1} = \arg\min_x \left(f(x) + \frac{\rho}{2}\|x - z^k + u^k\|_2^2 \right), \\ z^{k+1} = \arg\min_z \left(g(z) + \frac{\rho}{2}\|x^{k+1} - z + u^k\|_2^2 \right), \\ u^{k+1} = u^k + x^{k+1} - z^{k+1} \end{cases}$$

ADMM: введение ограничений

$$\min_{x \in Q} f(x) \Leftrightarrow \min_x f(x) + I_Q(x) \Leftrightarrow \min_{x=z} f(x) + I_Q(z)$$

$$\begin{cases} x^{k+1} &= \arg\min_x \left(f(x) + \frac{\rho}{2} \|x - z^k + u^k\|_2^2 \right), \\ z^{k+1} &= \arg\min_z \left(I_Q(z) + \frac{\rho}{2} \|x^{k+1} - z + u^k\|_2^2 \right) = \\ &= \mathrm{Proj}_Q(x^{k+1} + u^k), \\ u^{k+1} &= u^k + x^{k+1} - z^{k+1} \end{cases}$$

Как на практике решаются выпуклые задачи

Запись задачи в другой форме

$$\min_{x \in \overline{Q}} f(x) \Leftrightarrow \min_{\substack{x \in \overline{Q} \\ f(x) \le \tau}} \tau$$

Общая форма

$$\min_{x \in Q} (c, x)$$

Метод внутренней точки

$$\min_{x \in Q} (c, x)$$

Последовательность задач с барьером $F(x)$ при $t_k \to \infty$:

$$x^{k,*} = \arg\min_x \left(t_k(c, x) + F(x) \right) = \arg\min_x f_k(x)$$

$$\nabla f_k(x) = t_k c + \nabla F(x)$$

$$\nabla^2 f_k(x) = \nabla^2 F(x)$$

$\mathbf{dom}\, F$ открытое, при этом $\mathbf{cl}\ \mathbf{dom}\, F = Q$

Метод внутренней точки

$$\min_{x \in Q} (c, x)$$

Последовательность задач с барьером $F(x)$ при $t_k \to \infty$:

$$x^{k,*} = \arg\min_x \big(t_k(c, x) + F(x)\big) = \arg\min_x f_k(x)$$

$$\nabla f_k(x) = t_k c + \nabla F(x)$$

$$\nabla^2 f_k(x) = \nabla^2 F(x)$$

$\mathbf{dom}\, F$ открытое, при этом $\mathbf{cl}\ \mathbf{dom}\, F = Q$

Достаточно* делать по одному шагу метода Ньютона и правильно* увеличивать t_k.

Частные случаи выпуклых задач: ЛП (LP), QP

1. Линейное программирование

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^T x, \\ & Ax = b \\ & Cx \le d \end{aligned}$$

2. Квадратичное программирование ($Q \succcurlyeq 0$)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} x^T Q x + c^T x \\ & Ax = b \\ & Cx \le d \end{aligned}$$

Частные случаи выпуклых задач: SDP

3. Задачи полуопределённого программирования (semidefinite programming) с симметричными матрицами $\mathcal{A}_i = \mathcal{A}_i^T \in \mathbb{R}^{m \times m}, i = 0, ..., n$:

$$\begin{aligned} &\min && c^T x \\ &Ax = b \\ &\mathcal{A}_0 + \textstyle\sum_{i=1}^n x_i \mathcal{A}_i \preccurlyeq 0 \end{aligned}$$

$$\begin{aligned} &\min && (C, X) \\ &X \succcurlyeq 0 \\ &(\mathbb{A}_i, X) = d_i \end{aligned}$$

Частный случаи невыпуклых задач: MILP

4. Целочисленная задача линейного программирования

$$
\begin{aligned}
\min_{x \in \mathbb{Z}^n} \quad & c^T x, \\
Ax &= b \\
Cx &\leq d
\end{aligned}
$$

Инструменты

Пакеты

1. Встроенные: Matlab, Numpy, ...
2. Некоммерческие специализированные: для LP, QP, SDP и пр.
3. Коммерческие общего назначения: MOSEK, CPLEX, GUROBI, BARON.
4. Общего назначения для выпуклых задач оптимизации: CVX (Matlab, Python)

Дерево принятия решений при выборе программ оптимизации (Decision Tree for Optimization Software)

http://plato.asu.edu/guide

DCP и CVX

«Дисциплинированное выпуклое программирование»: https://dcp.stanford.edu

Одна из реализаций: CVX (Matlab)

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2 \\ \text{subject to} & Cx = d \\ & \|x\|_\infty \le e \end{array}$$

```
m = 20; n = 10; p = 4;
A = randn(m,n); b = randn(m,1);
C = randn(p,n); d = randn(p,1); e = rand;
cvx_begin
    variable x(n)
    minimize( norm( A * x - b, 2 ) )
    subject to
        C * x == d
        norm( x, Inf ) <= e
cvx_end
```

Общие выводы. Что делать

- Правильно ставить задачу в удобной форме (моделирование, подготовка)
- Понимать тип задачи и подобрать метод
- Стремиться построить выпуклую (под)задачу, если не получится, то маломерную
- Знать, когда остановиться и как проверять условия оптимальности
- Преобразовывать задачу, в том числе с ограничениями
- Уметь пользоваться пакетами программ

Самый последний слайд

Анонимный короткий опрос-отзыв:

https://forms.gle/EGYDfuiw1SxtgeM98

Спасибо!