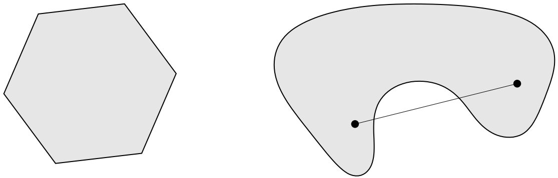
Сопряжённая функция и двойственность по Фенхелю

Михаил Фигурнов, аспирант

25 марта 2014

Выпуклое множество



Множество $C \subseteq \mathbb{R}^n$ называется **выпуклым**, если оно содержит линию между любыми двумя точками $x_1, x_2 \in C$:

$$\forall t : 0 \le t \le 1 \Rightarrow tx_1 + (1-t)x_2 \in C$$

Примеры: пустое множество, точка, плоскость, шар, конус, многогранник.
Более "абстрактные": полупространство, линейное подпространство.

Расширенная функция

Пусть функция задана на множестве $\operatorname{dom} f$. Доопределим её так, чтобы она вне этого множества равнялась ∞:

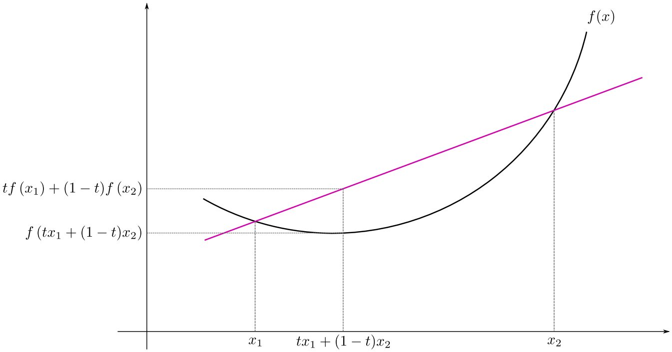
$$\tilde{f}(x) = \begin{cases} f(x), & x \in \operatorname{dom} f \\ \infty, & x \notin \operatorname{dom} f \end{cases}$$

Если мы минимизируем функцию, то значение минимума не изменится.

Можно восстановить область определения f:

$$\operatorname{dom} f = \{x \mid \tilde{f}(x) < \infty\}$$

Выпуклая функция



Функция f, заданная на выпуклом множестве $\mathrm{dom}\, f$, называется **выпуклой**, если для любых точек $x, y \in \mathrm{dom}\, f$:

$$\forall t : 0 \leq t \leq 1 \Rightarrow f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

Эквивалентно: надграфик f является выпуклым множеством.

Примеры выпуклых функций

- Линейная функция: $ax + b$
- Экспонента: e^x
- Степенная функция: $x^\alpha, \alpha \geq 1, x > 0$
- Отрицательная энтропия: $x \log x, x > 0$
- Индикатор выпуклого множества C:

$$I(x \in C) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases}$$

Операции, сохраняющие выпуклость функции

Пусть $f(x), f_i(x)$ выпуклые по x, а $f(x, y)$ выпуклая по x для любого y.

Тогда следующие функции являются выпуклыми:

- Композиция с линейной функцией: $f(Ax + b)$
- Неотрицательная сумма: $w_i \geq 0$, $w_1 f_1(x) + \dots + w_n f_n(x)$
- Обобщение: $w(y) > 0$, $\int w(y) f(x, y) dy$
- Поточечный максимум: $\max\{f_1(x), f_2(x)\}$
- Обобщение: $\sup_{y \in Y} f(x, y)$

Если $f(x, y)$ выпукла по (x, y), а C — непустое выпуклое множество, то $g(x) = \inf_{y \in C} f(x, y)$ — выпуклая функция.

$\min\{f_1(x), f_2(x)\}$ не обязательно выпуклая функция.

Композиция функций

Необходимое и достаточное условие выпуклости для дважды дифференцируемой функции:

$$f''(x) \geq 0$$

Когда дважды дифференцируемая функция $f(x) = h(g(x))$ является выпуклой?

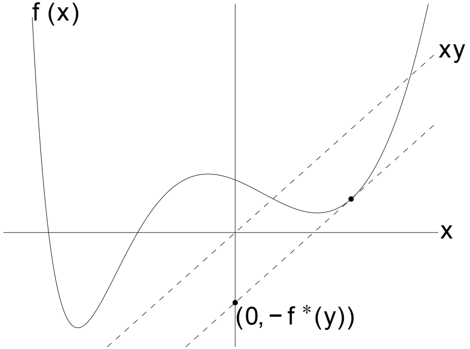
$$f''(x) = (h'(g(x))g'(x))' = h''(g(x))(g'(x))^2 + h'(g(x))g''(x) \geq 0$$

Достаточное условие. h выпуклая и неубывающая, g выпуклая:

$$h'' \geq 0, h' \geq 0, g'' \geq 0$$

Пример. Если $g(x)$ выпуклая, то $\exp g(x)$ выпуклая.

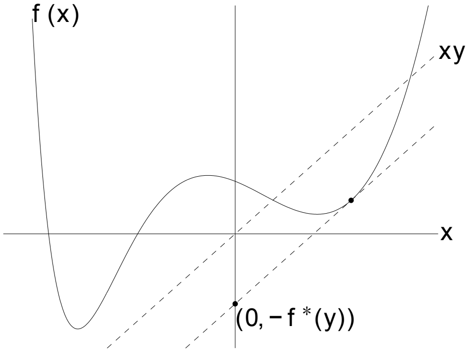
Определение сопряжённой функции



Рассмотрим функцию $f : \mathbb{R}^n \to \mathbb{R}$.
Сопряжённая функция (convex conjugate) $f^* : \mathbb{R}^n \to \mathbb{R}$:

$$f^*(y) = \sup_{x \in \mathrm{dom}\, f} (y^T x - f(x))$$

Сопряжённая функция выпуклая, даже если исходная невыпуклая (максимум семейства линейных функций).



Описываем функцию через точки касания опорных гиперплоскостей.

Если функция выпуклая, не потеряем информацию.

Вычисление сопряжённой функции I

Отрицательная энтропия.

$$f(x) = x\log x, \operatorname{dom} f = \{x \mid x > 0\}$$
$$f^*(y) = \sup_{x>0} xy - x\log x$$

В точке максимума производная по x равна нулю:

$$\frac{d}{dx}(xy - x\log x) = y - \log x - 1 = 0$$

$$\log x = y - 1$$
$$x = e^{y-1}$$

Подставляем в выражение для f^*:

$$f^*(y) = e^{y-1}y - e^{y-1}(y-1) = e^{y-1}$$

$$\operatorname{dom} f^* = \mathbb{R}$$

Вычисление сопряжённой функции II

Индикаторная функция точки:

$$f(x) = I(x = a)$$

$$f^*(y) = \sup_x xy - I(x = a) = ay$$

Линейная функция, плюс индикатор $x \succeq 0$.

$$\begin{aligned} f(x) &= \theta^T x + I(x \succeq 0) \\ f^*(y) &= \sup_x y^T x - \theta^T x - I(x \succeq 0) = \\ &= \sup_{x \succeq 0} (y - \theta)^T x = \begin{cases} 0, & y \preceq \theta \\ \infty, & y \npreceq \theta \end{cases} = I(y \preceq \theta) \end{aligned}$$

Свойства сопряжённой функции

Неравенство Фенхеля

$$\forall x, y \;\; f(x) + f^*(y) \geq x^T y$$

Масштабирование и перенос

$$g(x) = af(x) + b, a > 0$$

$$g^*(y) = af^*(y/a) - b$$

Сумма независимых функций $f_1(u)$, $f_2(v)$

$$(f_1 + f_2)^* = f_1^* + f_2^*$$

Если f выпуклая и надграфик f замкнут, то

$$f^{**} = f$$

Двойственность по Лагранжу

Задача условной оптимизации в стандартном виде:

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1, \ldots, p \end{aligned}$$

Лагранжиан:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

Двойственная функция:

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

Главное свойство: $g(\lambda, \nu)$ при $\lambda \succeq 0$ не превосходит оптимального решения исходной задачи.

Двойственная функция через сопряжённую функцию

Задача с линейными ограничениями:

$$
\begin{aligned}
\min_x \quad & f_0(x) \\
\text{s.t.} \quad & Ax \preceq b \\
& Cx = d
\end{aligned}
$$

Лагранжиан:

$$
L(x, \lambda, \nu) = f_0(x) + \lambda^T(Ax - b) + \nu^T(Cx - d)
$$

Двойственная функция:

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_x (f_0(x) + \lambda^T(Ax - b) + \nu^T(Cx - d)) = \\
&= -b^T\lambda - d^T\nu + \inf_x (f_0(x) + (A^T\lambda + C^T\nu)^T x) = \\
&= -b^T\lambda - d^T\nu - \sup_x (-(A^T\lambda + C^T\nu)^T x - f_0(x)) = \\
&= -b^T\lambda - d^T\nu - f_0^*(-A^T\lambda - C^T\nu)
\end{aligned}
$$

Пример. Максимизация энтропии

$$
\begin{aligned}
\min_x \quad & f_0(x) = \sum_{i=1}^n x_i \log x_i \\
\text{s.t.} \quad & Ax \preceq b \\
& 1^T x = 1
\end{aligned}
$$

Знаем сопряжённую функцию для отрицательной энтропии:

$$
(x \log x)^* = e^{y-1} \Rightarrow f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}
$$

Двойственная функция:

$$
g(\lambda, \nu) = -b^T\lambda - d^T\nu - f_0^*(-A^T\lambda - C^T\nu) =
$$

$$
= -b^T\lambda - \nu - \sum_{i=1}^n e^{-a_i^T\lambda - \nu - 1},
$$

где a_i — i-й столбец матрицы A.

Теорема двойственности Фенхеля

Прямая задача:

$$\inf_x f(x) + g(Ax)$$

Двойственная по Фенхелю задача:

$$\sup_y -f^*(A^T y) - g^*(-y)$$

Теорема

Пусть даны функции f, g (не обязательно выпуклые) и линейный оператор A. Тогда выполняется слабая двойственность:

$$\inf_x f(x) + g(Ax) \geq \sup_y -f^*(A^T y) - g^*(-y)$$

Достаточное условие сильной двойственности (равенства): f, g выпуклые, существует точка $x_0 \in \operatorname{dom} f$ такая, что $g(x)$ непрерывна в точке Ax_0.

Геометрический смысл

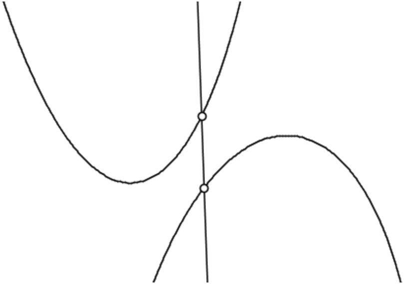
f, g выпуклые ($-g$ вогнутая).

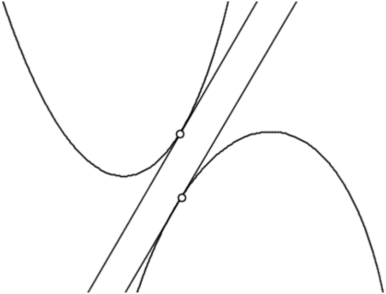
$$\inf_x f(x) + g(x)$$

Минимизируем зазор между функциями.

$$\sup_y -f^*(y) - g^*(-y)$$

Максимизируем расстояние между параллельными касательными.





Теорема двойственности Фенхеля утверждает, что оптимальные значения задач совпадают.

Роль в машинном обучении

$$
\begin{aligned}
w &— \text{вектор весов.}\\
f(w) &— \text{регуляризатор.}\\
A &— \text{матрица объект-признак.}\\
g(Aw) &— \text{эмпирический риск для вектора весов } w.
\end{aligned}
$$

Задача настройки линейного классификатора (размерность - число весов):

$$\inf_w f(w) + g(Aw)$$

Двойственная форма (размерность - число объектов):

$$\sup_y -f^*(A^T y) - g^*(-y)$$

Подробнее — см. [Rifkin]

Доказательство слабой двойственности

Вспомним **неравенство Фенхеля**:

$$\forall x, y \;\; f(x) + f^*(y) \geq x^T y$$

Запишем его два раза:

$$f(x) + f^*(A^T y) \geq x^T A^T y = (Ax)^T y$$
$$g(Ax) + g^*(-y) \geq (Ax)^T(-y) = -(Ax)^T y$$

Сложим:

$$f(x) + g(Ax) + f^*(A^T y) + g^*(-y) \geq 0$$
$$f(x) + g(Ax) \geq -f^*(A^T y) - g^*(-y)$$
$$\inf_x f(x) + g(Ax) \geq \sup_y -f^*(A^T y) - g^*(-y)$$

Сильную двойственность доказывать не будем, см. [Borwein, Lewis].

Энергия MRF

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ — вершины и рёбра ненаправленного графа.
x_i — переменные.
$x_i \in \mathcal{K}$ — множество меток переменных.
$\theta_i(x_i), \theta_{ij}(x_i, x_j)$ — потенциалы.

Минимизация энергии MRF ($x_i \in \mathcal{K}$):

$$\min_{x_i} \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{(ij) \in \mathcal{E}} \theta_{ij}(x_i, x_j)$$

LP-релаксация энергии MRF ($x_{ip}, x_{ij,pq} \in [0,1]$):

$$\begin{aligned}
\min_{x_{ip}, x_{ij,pq}} \quad & \sum_{i\in\mathcal{V}} \theta_{ip} x_{ip} + \sum_{(ij)\in\mathcal{E}} \sum_{p,q\in\mathcal{K}} \theta_{ij,pq} x_{ij,pq} = \theta^T x \\
\text{s.t.} \quad & \sum_{p\in\mathcal{K}} x_{ip} = 1, \ \forall i \in \mathcal{V} \\
& \sum_{p\in\mathcal{K}} x_{ij,pq} = x_{jq}, \ \forall (ij) \in \mathcal{E}, \forall q \in \mathcal{K} \\
& \sum_{q\in\mathcal{K}} x_{ij,pq} = x_{ip}, \ \forall (ij) \in \mathcal{E}, \forall p \in \mathcal{K} \\
& x_{ip} \geq 0, \ \forall i \in \mathcal{V}, \forall p \in \mathcal{K} \\
& x_{ij,pq} \geq 0, \ \forall (ij) \in \mathcal{E}, \forall p, q \in \mathcal{K}
\end{aligned}$$

Представим эту задачу в виде

$$\min_x f(x) + g(Ax)$$

Двойственная задача I

$$f(x) = \theta^T x + I(x \succeq 0) \text{ — целевая функция, ограничения вида } x_\bullet \geq 0$$

$$g(u) = \sum_{i=1}^{|\mathcal{V}|} I(u_i = 1) + \sum_{i=|\mathcal{V}|+1}^{N} I(u_i = 0) \text{ — остальные ограничения}$$

Сопряжённые функции уже знаем.
Двойственная по Фенхелю задача:

$$\max_y -f^*(A^T y) - g^*(-y) = \max_y -I(A^T y \preceq \theta) - \sum_{y=1}^{|\mathcal{V}|} (-y_i)$$

Более привычная форма:

$$\begin{aligned} &\max_y \sum_{i \in \mathcal{V}} y_i \\ \text{s.t.} \quad &A^T y \preceq \theta \end{aligned}$$

Сопоставим элементам вектора Ax переменные y.

$$
\begin{aligned}
y_i &: \quad \sum_{p \in \mathcal{K}} x_{ip} \\
y^1_{ij,q} &: \quad \sum_{p \in \mathcal{K}} x_{ij,pq} - x_{jq} \\
y^2_{ij,p} &: \quad \sum_{q \in \mathcal{K}} x_{ij,pq} - x_{ip}
\end{aligned}
$$

Компоненты вектора $A^T y$:

$$
(A^T y)_{ip} = y_i - \sum_{j:(ji) \in \mathcal{E}} y^1_{ji,p} - \sum_{j:(ij) \in \mathcal{E}} y^2_{ij,p}
$$

$$
(A^T y)_{ij,pq} = y^1_{ij,q} + y^2_{ij,p}
$$

Двойственная задача II

$$\begin{aligned} & \max_y \sum_{i \in \mathcal{V}} y_i \\ \text{s.t.} \quad & A^T y \preceq \theta \end{aligned}$$

Подставим $A^T y$:

$$\begin{aligned} \max_{y_i, y_{ij,p}} \quad & \sum_{i \in \mathcal{V}} y_i \\ \text{s.t.} \quad & y_i - \sum_{j:(ji) \in \mathcal{E}} y^1_{ji,p} - \sum_{j:(ij) \in \mathcal{E}} y^2_{ij,p} \leq \theta_{ip}, \ \forall i \in \mathcal{V}, \forall p \in \mathcal{K} \\ & y_{ij,p} + y_{ij,q} \leq \theta_{ij,pq}, \ \forall (ij) \in \mathcal{E}, \forall p, q \in \mathcal{K} \end{aligned}$$

Другая двойственная задача

Можно взять другую функцию $f(x)$ ($g(u)$ и A не меняем):

$$f(x) = \theta^T x + I(x \in [0,1]) = \sum_i \theta_i x_i + I(x_i \in [0,1])$$

$$f^*(y) = \sum_i (y_i - \theta_i)[y_i - \theta_i \geq 0] = \sum_i (y_i - \theta_i)_+$$

Это не накладывает дополнительных ограничений.

Получим другую двойственную задачу:

$$\max_y -(A^T y - \theta)_+ - \sum_{y=1}^{|\mathcal{V}|} (-y_i) =$$

$$= \max_{y_i, y_{ij,p}} \quad \sum_{i \in \mathcal{V}} y_i -$$

$$- \sum_{i \in \mathcal{V}} \sum_{p \in \mathcal{K}} (y_i - \sum_{j:(ji) \in \mathcal{E}} y^1_{ji,p} - \sum_{j:(ij) \in \mathcal{E}} y^2_{ij,p} - \theta_{ip})_+$$

$$- \sum_{(ij) \in \mathcal{E}} \sum_{p,q \in \mathcal{K}} (y_{ij,p} + y_{ij,q} - \theta_{ij,pq})_+$$

Две задачи с равным значением

(Но оптимальные значения переменных не совпадут!)

$$
\begin{aligned}
\max_{y_i, y_{ij,p}} \quad & \sum_{i \in \mathcal{V}} y_i \\
\text{s.t.} \quad & y_i - \sum_{j:(ji)\in\mathcal{E}} y^1_{ji,p} - \sum_{j:(ij)\in\mathcal{E}} y^2_{ij,p} \le \theta_{ip}, \ \forall i \in \mathcal{V}, \forall p \in \mathcal{K} \\
& y_{ij,p} + y_{ij,q} \le \theta_{ij,pq}, \ \forall (ij) \in \mathcal{E}, \forall p, q \in \mathcal{K}
\end{aligned}
$$

$$
\begin{aligned}
\max_{y_i, y_{ij,p}} \quad & \sum_{i \in \mathcal{V}} y_i - \\
& - \sum_{i \in \mathcal{V}} \sum_{p \in \mathcal{K}} (y_i - \sum_{j:(ji)\in\mathcal{E}} y^1_{ji,p} - \sum_{j:(ij)\in\mathcal{E}} y^2_{ij,p} - \theta_{ip})_+ \\
& - \sum_{(ij)\in\mathcal{E}} \sum_{p,q \in \mathcal{K}} (y_{ij,p} + y_{ij,q} - \theta_{ij,pq})_+
\end{aligned}
$$

Заключение

1. Сопряжённую функцию иногда легко подсчитать.
2. Сопряжённая функция позволяет явно находить двойственную функцию Лагранжа при линейных ограничениях.
3. Теорема двойственности по Фенхелю позволяет строить эквивалентные задачи, которые могут удобней решаться.
4. Например, так можно построить двойственные задачи минимизации энергии MRF.

Список литературы I

- Boyd, Stephen P., and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.
- Borwein, Jonathan M., and Adrian S. Lewis. Convex analysis and nonlinear optimization: theory and examples. Springer, 2010.
- http://en.wikipedia.org/wiki/Convex_function
- http://en.wikipedia.org/wiki/Convex_conjugate
- http://en.wikipedia.org/wiki/Fenchel's_duality_theorem

Список литературы II

- Bauschke, Heinz H., and Yves Lucet. WHAT IS a Fenchel conjugate? Notices of the AMS 59.1 (2012).
- Rifkin, Ryan M., and Ross A. Lippert. Value regularization and Fenchel duality. The Journal of Machine Learning Research 8 (2007): 441-479.
- Osokin, Anton. Fenchel dual for MRF energy minimization. Unpublished, 2013.