

Методы оптимизации в машинном обучении

Сопряженные нормы и функции. Двойственность Фенхеля.

1 Замкнутые функции

Напомним, что *надграфиком* функции $f : E \rightarrow \mathbb{R}$, заданной на множестве E , называется множество $\text{Epi } f := \{(x, t) \in E \times \mathbb{R} : f(x) \leq t\}$.

Определение 1 (Замкнутые функции). Пусть $f : E \rightarrow \mathbb{R}$ — функция, заданная на множестве E в нормированном пространстве V . Функция f называется *замкнутой*, если $\text{Epi } f$ является замкнутым множеством в пространстве $V \oplus \mathbb{R}$.

Утверждение 1 (Эквивалентное определение замкнутости через множества подуровней). Пусть $f : E \rightarrow \mathbb{R}$ — функция, заданная на множестве E в нормированном пространстве. Тогда f является замкнутой, если и только если множество подуровней $f^{-1}((-\infty, a]) := \{x \in E : f(x) \leq a\}$ является замкнутым для любого $a \in \mathbb{R}$.

Определение 2 (Полунепрерывность). Пусть $f : E \rightarrow \mathbb{R}$ — функция, заданная на множестве E в нормированном пространстве. Функция f называется *полунепрерывной снизу* (соответственно *сверху*) в точке $x_0 \in E$, если для любой последовательности $(x_k)_{k=1}^{\infty}$ точек из E , сходящейся к x_0 , выполняется $f(x_0) \leq \liminf_{k \rightarrow \infty} f(x_k)$ (соответственно $f(x_0) \geq \limsup_{k \rightarrow \infty} f(x_k)$). Если f является полунепрерывной снизу (соответственно сверху) в каждой точке $x \in E$, то f просто называется *полунепрерывной снизу* (соответственно *сверху*).

Пример 1. Пусть $E := [-1, 1]$, и пусть $f, g : E \rightarrow \mathbb{R}$ — функции

$$f(x) := \begin{cases} 0, & \text{если } x \neq 0, \\ -1, & \text{если } x = 0, \end{cases} \quad g(x) := \begin{cases} 0, & \text{если } x \neq 0, \\ 1, & \text{если } x = 0. \end{cases}$$

Обе функции f и g не являются непрерывными, однако f является полунепрерывной снизу, а g является полунепрерывной сверху.

Нетрудно видеть, что f является непрерывной, если и только если f одновременно является полунепрерывной снизу и сверху.

Утверждение 2 (Критерий замкнутости). Пусть $f : E \rightarrow \mathbb{R}$ — функция, заданная на множестве E в нормированном пространстве. Тогда f является замкнутой, если и только если

1. f является полунепрерывной снизу;
2. f обладает барьерным свойством: для всех $x_0 \in \partial E \setminus E$ выполнено $f(x) \rightarrow +\infty$ при $x \rightarrow x_0; x \in E$.

Замечание 1. Если множество $\partial E \setminus E$ является пустым, т. е. если множество E замкнутое, то барьерное свойство выполняется бессодержательно. Таким образом, в случае замкнутой области определения, замкнутость функции эквивалентна полунепрерывности снизу.

Приведенный критерий является одним из главных поставщиков базовых примеров замкнутых функций.

Пример 2. Любая полунепрерывная (в том числе и просто непрерывная) функция $f : E \rightarrow \mathbb{R}$, заданная на *замкнутом* множестве E в нормированном пространстве, замкнутая.

Пример 3. Пусть $E := (0, +\infty)$ — открытое множество в \mathbb{R} . Тогда функции $x \mapsto -\ln x$ и $x \mapsto \frac{1}{x}$, заданные на E , замкнутые, поскольку они обе непрерывны, и обладают барьерным свойством: в данном случае $\partial E \setminus E = \{0\}$ и обе функции стремятся к $+\infty$ при $x \rightarrow 0; x > 0$. В то же время функции $x \mapsto \ln x$ и $x \mapsto 0$, заданные на E , являются непрерывными, но не замкнутыми, поскольку не обладают барьерным свойством: первая из них стремится к $-\infty$ при $x \rightarrow 0; x > 0$, а вторая стремится к 0.

Утверждение 3 (Операции, сохраняющие замкнутость). Пусть E и G — множества в нормированных пространствах V и W соответственно, и пусть $f : E \rightarrow \mathbb{R}$ и $g : G \rightarrow \mathbb{R}$ — замкнутые функции.

1. (Умножение на положительный скаляр) Если $c > 0$, то функция $cf : E \rightarrow \mathbb{R}$ замкнутая.
2. (Сумма) Если $V = W$, то функция $f + g : E \cap G \rightarrow \mathbb{R}$ замкнутая.
3. (Композиция с непрерывным преобразованием) Если $T : V \rightarrow W$ — непрерывное преобразование, то композиция $g \circ T : T^{-1}(G) \rightarrow \mathbb{R}$ замкнутая.
4. (Поточечный супремум) Пусть I — произвольное (не обязательно конечное и не обязательно счетное) множество, и пусть для каждого $i \in I$ заданы множество E_i в пространстве V и замкнутая функция $f_i : E_i \rightarrow \mathbb{R}$. Тогда функция $\sup_{i \in I} f_i : E \rightarrow \mathbb{R}$, определенная на множестве $E := \{x \in \bigcap_{i \in I} E_i : \sup_{i \in I} f_i(x) < +\infty\}$, замкнутая.
5. (Поточечный минимум) Пусть $\phi : E \times G \rightarrow \mathbb{R}$ — замкнутая функция (относительно пространства $V \oplus W$). Если множество G компактное, то функция $h : E \rightarrow \mathbb{R}$, определенная как $h(x) := \min_{y \in G} \phi(x, y)$, замкнутая. (Минимум здесь всегда достигается согласно теореме Вейерштрасса.)

Пример 4 (Максимум конечного числа функций). Пусть $f_1 : E_1 \rightarrow \mathbb{R}, \dots, f_m : E_m \rightarrow \mathbb{R}$ — замкнутые функции, заданные на множествах E_1, \dots, E_m в нормированном пространстве. Тогда поточечный максимум $\max_{1 \leq i \leq m} f_i : \bigcap_{1 \leq i \leq m} E_i \rightarrow \mathbb{R}$ этих функций является замкнутой функцией как частный случай поточечного супремума (с индексным множеством $I := \{1, \dots, m\}$).

Пример 5 (Супремум семейства аффинных функций). Пусть V — евклидово пространство, I — произвольное множество, и пусть для каждого $i \in I$ заданы вектор $a_i \in V$ и скаляр $b_i \in \mathbb{R}$. Пусть $E := \{x \in V : \sup_{i \in I} \{\langle a_i, x \rangle + b_i\} < +\infty\}$. Тогда функция $f : E \rightarrow \mathbb{R}$, заданная как $f(x) := \sup_{i \in I} \{\langle a_i, x \rangle + b_i\}$, замкнутая. Это следует из правила поточечного супремума и того, что аффинная функция является замкнутой (как непрерывная функция).

2 Сопряженные функции

Определение 3 (Сопряженная функция). Пусть $f : E \rightarrow \mathbb{R}$ — функция, заданная на множестве E в евклидовом пространстве V . Сопряженной функцией для функции f называется функция $f^* : E_* \rightarrow \mathbb{R}$, определенная как

$$f^*(s) := \sup_{x \in E} \{\langle s, x \rangle - f(x)\},$$

где $E_* := \{s \in V : \sup_{x \in E} \{\langle s, x \rangle - f(x)\} < +\infty\}$. Сопряженную функцию также часто называют сопряженной функцией Фенхеля. Преобразование $f \mapsto f^*$ называют преобразованием Фенхеля, преобразованием Лежандра или преобразованием Фенхеля–Лежандра.

Из определения сразу же следует, что сопряженная функция f^* является выпуклой и замкнутой функцией как поточечный супремум семейства аффинных функций, независимо от того, является ли при этом исходная функция f выпуклой или замкнутой.

Замечание 2. Вообще говоря, сопряженная функция f^* зависит от скалярного произведения, введенного в пространстве V . Таким образом, при работе с сопряженными функциями необходимо четко указывать, какое конкретно используется скалярное произведение. Всюду в дальнейшем, если не оговорено иное, будем считать, что во всех пространствах, с которыми мы работаем, введено стандартное скалярное произведение, и сопряженная функция определяется относительно него. Отметим также, что существует более общий взгляд на сопряженные функции, в котором сопряженная функция определяется на *двойственном пространстве* E^* всевозможных непрерывных линейных функционалов на E ; в этом случае $s \in E^*$ является линейным функционалом, а операция $\langle s, x \rangle$ интерпретируется как вычисление линейного функционала s на аргументе x . При таком определении никаких проблем вышеуказанного характера не возникает (и при этом даже можно работать с пространствами, на которых скалярное произведение ввести невозможно). Тем не менее, для простоты и понятности мы не будем рассматривать это обобщение.

Следующее утверждение является тривиальным следствием из определения сопряженной функции:

Утверждение 4 (Неравенство Фенхеля–Юнга). Пусть $f : E \rightarrow \mathbb{R}$ — функция, заданная на множестве E в евклидовом пространстве, и пусть $f^* : E_* \rightarrow \mathbb{R}$ — сопряженная функция. Тогда для всех $x \in E$ и всех $s \in E_*$ выполнено

$$\langle s, x \rangle \leq f^*(s) + f(x).$$

Несмотря на свою простоту, это утверждение имеет многочисленные приложения. Неравенство Фенхеля–Юнга дает систематичный способ построения оценок вида

$$\langle s, x \rangle \leq g(s) + f(x), \tag{1}$$

для всех s и x , где f и g — некоторые функции. Сопряженная функция может быть мотивирована следующим вопросом: какие функции f и g нужно выбрать в неравенстве (1), чтобы оно было «максимально точным»? Однозначного ответа на такой вопрос нет, поскольку, например, если пара (f, g) дает «максимально точную» оценку, то пара $(f + c, g - c)$ также будет давать «максимально точную» оценку для любого $c \in \mathbb{R}$. Тем не менее, если зафиксировать хотя бы одну из этих функций, скажем, f , то другая функция g , дающая «максимально точную» оценку в неравенстве (1) уже определяется однозначно — это в точности сопряженная функция f^* из определения 3.

Перейдем к некоторым важным примерам сопряженных функций.

Для первого примера (а также в дальнейшем) нам понадобится понятие индикаторной функции. *Индикаторной функцией* множества E называется функция $\delta_E : E \rightarrow \mathbb{R}$, определенная как $\delta_E(x) := 0$. (Таким образом, функция δ_E тождественно равна нулю на множестве E ; за пределами E функция δ_E не определена.)

Пример 6 (Линейная форма). Пусть V — евклидово пространство, и пусть $f : V \rightarrow \mathbb{R}$ — функция $f(x) := \langle a, x \rangle$, где $a \in V$. Поскольку

$$\sup_{x \in V} \{\langle s, x \rangle - \langle a, x \rangle\} = \sup_{x \in V} \langle s - a, x \rangle = \begin{cases} 0, & \text{если } s = a, \\ +\infty, & \text{иначе} \end{cases}$$

для всех $s \in V$, то сопряженная функция f^* — это индикаторная функция $\delta_{\{a\}}$ одноэлементного множества $\{a\}$.

Пример 7 (Индикатор одноэлементного множества). Пусть V — евклидово пространство, $a \in V$, и пусть $\delta_{\{a\}} : \{a\} \rightarrow \mathbb{R}$ — индикаторная функция одноэлементного множества $\{a\}$. Поскольку для $s \in V$ выполнено

$$\sup_{x \in \{a\}} \{\langle s, x \rangle - \delta_{\{a\}}(x)\} = \langle s, a \rangle,$$

то сопряженная функция $\delta_{\{a\}}^*$ равна линейной форме $x \mapsto \langle a, x \rangle$, рассмотренной в примере 6.

Для следующего примера нам понадобится понятие *сопряженной нормы*. Пусть V — конечномерное евклидово пространство, и пусть $\|\cdot\|$ — произвольная норма в V (не обязательно индуцированная скалярным произведением). Тогда *сопряженной нормой* для $\|\cdot\|$ называется норма $\|\cdot\|_*$, определенная как

$$\|s\|_* := \max_{\|v\|=1} |\langle s, v \rangle|.$$

Заметим, что это определение корректное, поскольку максимум достигается по теореме Вейерштрасса в силу компактности единичной сферы.

Пример 8 (l^p -норма). Пусть $p \in [1, +\infty]$, и пусть $\|\cdot\|_p : \mathbb{R}^n \rightarrow \mathbb{R}$ — l^p -норма. Тогда из неравенства Гельдера следует, что сопряженная норма $(\|\cdot\|_p)_*$ — это l^q -норма $\|\cdot\|_q$, где $q \in [1, +\infty]$ определяется из равенства $1/p + 1/q = 1$. В частности, l^1 -норма является сопряженной к l^∞ -норме; в то же время, l^∞ -норма является сопряженной к l^1 -норме, а l^2 норма является самосопряженной.

Упражнение 1 (Основные свойства сопряженной нормы). Пусть V — конечномерное евклидово пространство, и пусть $\|\cdot\| : V \rightarrow \mathbb{R}$ — норма в V (не обязательно индуцированная скалярным произведением). Покажите, что:

1. Сопряженная норма $\|\cdot\|_*$, действительно, является нормой в V .

2. (Неравенство Гельдера) Для всех $s, v \in V$ выполнено

$$|\langle s, v \rangle| \leq \|s\|_* \|v\|.$$

3. (Эквивалентные формы записи сопряженной нормы) Для всех $s \in V$ выполнено

$$\|s\|_* = \max_{\|v\|=1} \langle s, v \rangle = \max_{\|v\| \leq 1} |\langle s, v \rangle| = \max_{\|v\| \leq 1} \langle s, v \rangle.$$

4. (Сопряжение является инволюцией) $(\|\cdot\|_*)_* = \|\cdot\|$.

Пример 9 (Норма). Пусть V — евклидово пространство, и пусть $\|\cdot\|$ — произвольная норма в V (не обязательно индуцированная скалярным произведением). Тогда сопряженная функция $\|\cdot\|_*$ равна индикаторной функции замкнутого единичного шара $\overline{B}_{\|\cdot\|_*}(0, 1)$ с центром в нуле относительно сопряженной нормы $\|\cdot\|_*$:

$$\|\cdot\|_* = \delta_{\overline{B}_{\|\cdot\|_*}(0, 1)}.$$

Доказательство. Пусть $s \in V$. Согласно определению сопряженной нормы, $\langle s, x \rangle \leq \|s\|_* \|x\|$ для всех $x \in V$, причем неравенство достигается для некоторого $x \in V$. Отсюда

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} = \sup_{t \geq 0} (\|s\|_* - 1)t = \begin{cases} 0, & \text{если } \|s\|_* \leq 1, \\ +\infty, & \text{иначе.} \end{cases}$$

Таким образом, областью определения сопряженной функции $\|\cdot\|_*$ является $\overline{B}_{\|\cdot\|_*}(0, 1)$, и всюду на этом множестве функция $\|\cdot\|_*$ равна нулю. Это означает, что $\|\cdot\|_* = \delta_{\overline{B}_{\|\cdot\|_*}(0, 1)}$. \square

Замечание 3. Обратите внимание, что сопряженная функция $\|\cdot\|_*$ для нормы $\|\cdot\|$ не равна сопряженной норме $\|\cdot\|_*$. Вместо этого, она равна индикатору $\delta_{\overline{B}_{\|\cdot\|_*}(0, 1)}$ единичного шара относительно сопряженной нормы. Несмотря на схожие обозначения, функции $\|\cdot\|_*$ и $\|\cdot\|$ являются разными: например, сопряженная норма $\|\cdot\|_*$ всегда определена на всем пространстве V (как и любая норма), в то время как сопряженная функция $\|\cdot\|_*$ оказывается определенной на собственном подмножестве V .

Пример 10 (Индикатор единичного шара). Пусть V — евклидово пространство, и пусть $\|\cdot\|$ — произвольная норма в V (не обязательно индуцированная скалярным произведением). Тогда сопряженной функцией для индикатора $\delta_{\overline{B}_{\|\cdot\|}(0, 1)}$ замкнутого единичного шара с центром в нуле относительно нормы $\|\cdot\|$ является сопряженная норма $\|\cdot\|_*$:

$$(\delta_{\overline{B}_{\|\cdot\|}(0, 1)})_* = \|\cdot\|_*.$$

Доказательство. Для $s \in V$ имеем

$$\sup_{x \in \overline{B}_{\|\cdot\|}(0,1)} \langle s, x \rangle = \sup_{\|x\| \leq 1} \langle s, x \rangle = \|s\|_*,$$

где последнее равенство следует из упражнения 1. \square

Пример 11 (Логистическая функция). Пусть $f : \mathbb{R} \rightarrow \mathbb{R}$ — функция $f(x) := \ln(1+e^x)$. Тогда $f^* = \text{Ent}_b$, где $\text{Ent}_b : [0, 1] \rightarrow \mathbb{R}$ — бинарная энтропийная функция

$$\text{Ent}_b(x) := \begin{cases} x \ln x + (1-x) \ln(1-x), & \text{если } 0 < x < 1, \\ 0, & \text{если } x = 0 \text{ или } x = 1. \end{cases}$$

Доказательство. Пусть $s \in \mathbb{R}$. Вычислим

$$\sup_{x \in \mathbb{R}} \{sx - \ln(1 + e^x)\}. \quad (2)$$

Для этого рассмотрим различные случаи.

Пусть $s < 0$. Из монотонности логарифма и того, что $e^x < 1$ при $x < 0$, следует $\ln(1 + e^x) < \ln 2$ для всех $x < 0$. Значит, $sx - \ln(1 + e^x) > sx - \ln 2$ для всех $x < 0$. Поскольку $sx \rightarrow +\infty$ при $x \rightarrow -\infty$, то отсюда следует, что $sx - \ln(1 + e^x) \rightarrow +\infty$ при $x \rightarrow -\infty$; $x < 0$. Таким образом, супремум (2) равен $+\infty$.

В случае $s > 1$, то аналогичные рассуждения дают неравенство $\ln(1 + e^x) < \ln(e^x + e^x) = \ln 2 + x$ при $x > 0$. Отсюда $sx - \ln(1 + e^x) > (s-1)x - \ln 2$ для всех $x > 0$. Устремляя $x \rightarrow +\infty$; $x > 0$, получаем, что супремум (2) равен $+\infty$.

Пусть $s = 0$. Поскольку $\ln(1 + e^x) \geq 0$ для всех $x \in \mathbb{R}$ и $\ln(1 + e^x) \rightarrow 0$ при $x \rightarrow -\infty$, то в этом случае супремум (2) равен нулю.

Если $s = 1$, то из неравенства $\ln(1 + e^x) \geq x$ для всех $x \in \mathbb{R}$, следует, что супремум (2) не больше нуля. Но, поскольку $\ln(1 + e^x) = x + \ln(1 + e^{-x})$ для всех $x \in \mathbb{R}$ и $\ln(1 + e^{-x}) \rightarrow 0$ при $x \rightarrow +\infty$, супремум (2) в точности равен нулю.

Наконец, пусть $0 < s < 1$. Несложные вычисления показывают, что в этом случае функция $x \mapsto sx - \ln(1 + e^x)$ имеет точку стационарности $\bar{x} := \ln s - \ln(1-s)$, а соответствующее значение функции в \bar{x} равно $s \ln s + (1-s) \ln(1-s)$. Поскольку функция $x \mapsto sx - \ln(1 + e^x)$ вогнутая, то найденное значение — это и есть супремум (2). \square

Пример 12 (Бинарная энтропийная функция). Пусть $\text{Ent}_b : [0, 1] \rightarrow \mathbb{R}$ — бинарная энтропийная функция, определенная в примере 11. Тогда $\text{Ent}_b^* : \mathbb{R} \rightarrow \mathbb{R}$ — это логистическая функция

$$\text{Ent}_b^*(s) = \ln(1 + e^s)$$

для всех $s \in \mathbb{R}$.

Доказательство. Пусть $s \in \mathbb{R}$. Вычислим

$$\sup_{0 \leq x \leq 1} \{sx - \text{Ent}_b(x)\}. \quad (3)$$

Для этого найдем

$$\sup_{0 < x < 1} \{sx - \text{Ent}_b(x)\} = \sup_{0 < x < 1} \{sx - x \ln x - (1-x) \ln(1-x)\}.$$

Поскольку множество $(0, 1)$ открытое, а функция $x \mapsto sx - x \ln x - (1-x) \ln(1-x)$, определенная на этом множестве, является вогнутой, то достаточно найти точку стационарности и соответствующее

значение этой функции. Дифференцируя, находим, что $\bar{x} := \frac{1}{1+e^{-s}}$ является точкой стационарности с соответствующим значением функции $\ln(1+e^s)$. Таким образом,

$$\sup_{0 < x < 1} \{sx - \text{Ent}_b(x)\} = \ln(1+e^s).$$

Поскольку функция $sx - \text{Ent}_b(x) = 0$ для $x = 0$ и $sx - \text{Ent}_b(x) = s$ для $x = 1$, то супремум (3) равен максимуму из $0, 1$ и $\ln(1+e^s)$. Остается воспользоваться тем, что $\ln(1+e^s) \geq \max\{0, s\}$. \square

Как показывают примеры 6, 7, 9, 10, 11, 12, преобразование Фенхеля $f \mapsto f^*$ зачастую является инволюцией: если f^* — сопряженная функция для f , то сопряженная функция f^{**} к функции f^* — это исходная функция f . Следующая крайне важная теорема говорит о том, что это не случайно, и что такой результат является отличительной особенностью класса выпуклых замкнутых функций:

Теорема 1 (Теорема Фенхеля–Моро). *Пусть f — функция, заданная на непустом множестве в евклидовом пространстве. Тогда $f = f^{**}$, если и только если f является выпуклой и замкнутой.*

Замечание 4. В одну сторону теорема Фенхеля–Моро очевидна: если $f = f^{**}$, то f обязана быть выпуклой и замкнутой, поскольку сопряженная функция к любой функции (в том числе к f^*) является выпуклой и замкнутой вне зависимости от того, обладала ли исходная функция этими свойствами или нет. Нетривиальной частью теоремы Фенхеля–Моро является обратное утверждение: что выпуклость и замкнутость f является не только необходимым условием, но и достаточным.

Утверждение 5 (Исчисление сопряженных функций). *Пусть $f : E \rightarrow \mathbb{R}$ и $g : G \rightarrow \mathbb{R}$ — функции, заданные на множествах E и G в евклидовых пространствах V и W соответственно, и пусть $f^* : E_* \rightarrow \mathbb{R}$ и $g^* : G_* \rightarrow \mathbb{R}$ — соответствующие сопряженные функции.*

1. (Умножение на положительный скаляр) Если $c > 0$, то $(cf)^* : cE_* \rightarrow \mathbb{R}$ — это функция $(cf)^*(s) = cf^*(s/c)$ для всех $s \in cE_*$.
2. (Сдвиг аргумента) Пусть $a \in V$, и пусть $h : E - a \rightarrow \mathbb{R}$ — функция $h(x) := f(x + a)$. Тогда $h^* : E_* \rightarrow \mathbb{R}$ — это функция $h^*(s) = f^*(s) - \langle s, a \rangle$ для всех $s \in E_*$.
3. (Прибавление аффинной функции) Пусть $a \in V$, $b \in \mathbb{R}$, и пусть $h : E \rightarrow \mathbb{R}$ — функция $h(x) := f(x) + \langle a, x \rangle + b$. Тогда $h^* : E_* + a \rightarrow \mathbb{R}$ — это функция $h^*(s) = f^*(s - a) - b$ для всех $s \in E_* + a$.
4. (Композиция с обратимым линейным преобразованием) Пусть $A : V \rightarrow V$ — обратимое линейное преобразование. Тогда $(f \circ A)^* : A^*(E_*) \rightarrow \mathbb{R}$ — это функция $(f \circ A)^*(s) = f^*(A^{-*}s)$ для всех $s \in A^*(E_*)$.
5. (Сепарабельная сумма) Пусть $\phi : E \times G \rightarrow \mathbb{R}$ — функция $\phi(x, y) := f(x) + g(y)$. Тогда $\phi^* : E_* \times G_* \rightarrow \mathbb{R}$ — это функция $\phi^*(s, u) = f^*(s) + g^*(u)$ для всех $s \in E_*$ и всех $u \in G_*$.

Пример 13 (Норма с коэффициентом). Этот пример является обобщением примера 9. Пусть V — евклидово пространство, $\|\cdot\|$ — произвольная норма в V (не обязательно индуцированная скалярным произведением), и пусть $\lambda > 0$. Тогда $(\lambda\|\cdot\|)^* = \delta_{\overline{B}_{\|\cdot\|_*}(0, \lambda)}$, т. е. сопряженная функция для функции $x \mapsto \lambda\|x\|$ — это индикатор замкнутого шара радиуса λ с центром в нуле относительно сопряженной нормы $\|\cdot\|_*$.

Пример 14 (Многомерная логистическая функция). Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ — функция $f(x) := \sum_{i=1}^n \ln(1 + e^{x_i})$. Тогда $f^* : [0, 1]^n \rightarrow \mathbb{R}$ — функция

$$f^*(s) = \sum_{i=1}^n \text{Ent}_b(x_i)$$

для всех $s \in \mathbb{R}^n$. Это следует из свойства сепарабельной суммы и примера 11.

3 Двойственность Фенхеля

Пусть $f : E \rightarrow \mathbb{R}$ и $g : G \rightarrow \mathbb{R}$ — функции, заданные на множествах E и G в евклидовых пространствах V и W соответственно, и пусть $A : V \rightarrow W$ — линейное преобразование. *Задачей Рокафеллара* называется задача минимизации следующего вида:

$$\min_{x \in E \cap A^{-1}(G)} f(x) + g(Ax), \quad (4)$$

где $A^{-1}(G) := \{x \in V : Ax \in G\}$ — прообраз G .

Для этой задачи можно построить двойственную задачу Лагранжа с помощью следующего приема, который называется *разделение переменных*. Введем дополнительную переменную $y = Ax$. Тогда задача Рокафеллара переписывается в виде

$$\min_{x \in E, y \in G} f(x) + g(y) \quad \text{s. t. } y = Ax.$$

Пусть $L : E \times G \times W$ — функция Лагранжа для этой задачи:

$$L(x, y, u) := f(x) + g(y) + \langle u, Ax - y \rangle.$$

Найдем двойственную функцию Лагранжа. Для этого зафиксируем $u \in W$ и вычислим

$$\begin{aligned} \inf_{x \in E, y \in G} L(x, y, u) &= \inf_{y \in G} \{g(y) - \langle u, y \rangle\} + \inf_{x \in E} \{f(x) + \langle A^*u, x \rangle\} \\ &= -\sup_{y \in G} \{\langle u, y \rangle - g(y)\} - \sup_{x \in E} \{\langle -A^*u, x \rangle - f(x)\}. \end{aligned}$$

Здесь A^* — сопряженный оператор для A . Первое равенство следует из свойства сепарабельности инфимума, а второе из того, что $\inf X = -\sup(-X)$ для любого множества $X \subseteq \mathbb{R}$.

Пусть $f^* : E_* \rightarrow \mathbb{R}$ и $g^* : G_* \rightarrow \mathbb{R}$ — сопряженные функции для f и g соответственно. Из определения следует, что

$$\sup_{y \in G} \{\langle u, y \rangle - g(y)\} = \begin{cases} g^*(u), & \text{если } u \in G_*, \\ +\infty, & \text{иначе.} \end{cases}$$

Аналогично,

$$\sup_{x \in E} \{\langle -A^*u, x \rangle - f(x)\} = \begin{cases} f^*(-A^*u), & \text{если } u \in (-A^*)^{-1}(E_*), \\ +\infty, & \text{иначе.} \end{cases}$$

Таким образом,

$$-\sup_{y \in G} \{\langle u, y \rangle - g(y)\} - \sup_{x \in E} \{\langle -A^*u, x \rangle - f(x)\} = \begin{cases} -g^*(u) - f^*(-A^*u), & \text{если } u \in G_* \cap (-A^*)^{-1}(E_*), \\ -\infty, & \text{иначе.} \end{cases}$$

В результате, двойственная функция Лагранжа ϕ определена на множестве $G_* \cap (-A^*)^{-1}(E_*)$ и равна

$$\phi(u) = -g^*(u) - f^*(-A^*u).$$

Значит, двойственная задача Лагранжа следующая:

$$\max_{u \in G_* \cap (-A^*)^{-1}(E_*)} -g^*(u) - f^*(-A^*u). \quad (5)$$

Задача (5) называется *двойственной задачей (Фенхеля)* к задаче (4).

Следующая фундаментальная теорема резюмирует вышеописанные рассуждения и предоставляет достаточные условия для сильной двойственности:

Теорема 2 (Двойственность Фенхеля). Пусть $f : E \rightarrow \mathbb{R}$, $g : G \rightarrow \mathbb{R}$ — функции, заданные на множествах E и G в евклидовых пространствах V и W соответственно, $f^* : E_* \rightarrow \mathbb{R}$, $g^* : G_* \rightarrow \mathbb{R}$ — соответствующие сопряженные функции, $A : V \rightarrow W$ — линейное преобразование, и пусть $p, d \in [-\infty, +\infty]$ — оптимальные значения в прямой и двойственной задаче

$$p := \inf_{x \in E \cap A^{-1}(G)} \{f(x) + g(Ax)\},$$

$$d := \sup_{u \in G_* \cap (-A^*)^{-1}(E_*)} \{-g^*(u) - f^*(-A^*u)\}.$$

Тогда выполняется слабая двойственность: $p \geq d$. Если при этом функции f и g выпуклые и выполняется условие $0 \in \text{int}(G - A(E))$ (или более сильное условие $A(E) \cap \text{int}(G) \neq \emptyset$) тогда выполняется сильная двойственность: $p = d$; если при этом d конечно, то супремум в двойственной задаче достигается. При этом точки $\bar{x} \in E \cap A^{-1}(G)$ и $\bar{u} \in G_* \cap (-A^*)^{-1}(E_*)$ являются оптимальными решениями для прямой и двойственной задач, если и только если $\langle -A^*\bar{u}, \bar{x} \rangle = f^*(-A^*\bar{u}) + f(\bar{x})$ и $\langle \bar{u}, A\bar{x} \rangle = g^*(\bar{u}) + g(A\bar{x})$, т.е. если соответствующие неравенства Фенхеля–Юнга переходят в равенства.

Если перейти от задачи максимизации к эквивалентной ей задаче минимизации минус функции, то двойственную задачу к задаче

$$\min_{x \in E \cap A^{-1}(G)} f(x) + g(Ax)$$

можно записать в форме

$$\min_{u \in G_* \cap (-A^*)^{-1}(E_*)} g^*(u) + f^*(-A^*u).$$

В таком виде двойственность Фенхеля является полностью симметричной: двойственной задачей к задаче Рокафеллара с параметрами (f, g, A) является задача Рокафеллара с параметрами $(g^*, f^*, -A^*)$; если при этом функции f и g являются выпуклыми и замкнутыми, то двойственной задачей к двойственной задаче является исходная задача Рокафеллара, поскольку $f^{**} = f$, $g^{**} = g$ (см. теорему 1) и $-((-A^*)^*)^* = A$ (это простой результат из линейной алгебры, выпуклость и замкнутость f и g здесь ни при чем).

Пример 15. Построим двойственную задачу для задачи логистической регрессии

$$\inf_{x \in \mathbb{R}^n} \sum_{i=1}^m \ln(1 + e^{(a_i, x)}), \quad (6)$$

где $a_1, \dots, a_m \in \mathbb{R}^n$.

Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ и $g : \mathbb{R}^m \rightarrow \mathbb{R}$ — функции

$$f(x) := 0, \quad g(y) := \sum_{i=1}^m \ln(1 + e^{y_i}),$$

и пусть $A := [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$ — матрицами со строками a_1, \dots, a_m . Тогда задачу (6) можно переписать в виде задачи Рокафеллара

$$p := \inf_{x \in \mathbb{R}^n} f(x) + g(Ax).$$

Согласно примерам 6 и 14, соответствующие сопряженные функции следующие:

$$f^* = \delta_{\{0\}}, \quad g^* : [0, 1]^m \rightarrow \mathbb{R} \text{ — функция } g^*(u) = \sum_{i=1}^m \text{Ent}_b(u_i).$$

Таким образом, двойственная задача имеет вид

$$d := \sup_{u \in [0,1]^m} - \sum_{i=1}^m \text{Ent}_b(u_i) \quad \text{s. t. } A^T u = 0,$$

т.е. минимизация сепарабельной суммы бинарных энтропийных функций на линейном подпространстве $\{u \in \mathbb{R}^n : A^T u = 0\}$. Этот пример показывает, что двойственная задача для безусловной задачи может быть условной. Согласно теореме 2, в данном случае имеет место сильная двойственность: $p = d$ и супремум в двойственной задаче достигается, если он конечный (почему условие $A(E) \cap \text{int}(G) \neq \emptyset$ выполняется?).

Пример 16. Теперь построим двойственную задачу для логистической регрессии с l^1 -регуляризатором:

$$p := \inf_{x \in \mathbb{R}^n} \sum_{i=1}^m \ln(1 + e^{a_i x}) + \lambda \|x\|_1,$$

где $a_1, \dots, a_m \in \mathbb{R}^n$. Как и раньше, пусть $A := [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$ — матрица со строками a_1, \dots, a_m . В этом случае $f : \mathbb{R}^n \rightarrow \mathbb{R}$ и $g : \mathbb{R}^m \rightarrow \mathbb{R}$ — это функции

$$f(x) := \lambda \|x\|_1, \quad g(y) := \sum_{i=1}^m \ln(1 + e^{y_i}).$$

Согласно примерам 13, 8 и 14, соответствующие сопряженные функции

$$f^* = \delta_{\overline{B}_{\|\cdot\|_\infty}(0, \lambda)}, \quad g^* : [0, 1]^m \rightarrow \mathbb{R} \text{ — функция } g^*(u) = \sum_{i=1}^m \text{Ent}_b(u_i).$$

Значит, двойственная задача имеет вид

$$d := \sup_{u \in [0,1]^m} - \sum_{i=1}^m \text{Ent}_b(u_i) \quad \text{s. t. } \|A^T u\|_\infty \leq \lambda.$$

Точно так же, как и в предыдущем примере, в данном случае имеет место сильная двойственность: $p = d$ и супремум в двойственной задаче достигается, если он конечный.