

1 Метод зеркального спуска

Рассмотрим стандартную задачу выпуклой негладкой оптимизации:

$$\min_{x \in Q} f(x), \quad (1.1)$$

где $f : \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая негладкая функция, а Q — выпуклое замкнутое множество в \mathbb{R}^n .

Метод зеркального спуска является обобщением стандартного субградиентного метода. Напомним, что субградиентный метод для решения задачи (1.1) начинается с некоторой точки $x_0 \in \mathbb{R}^n$ и далее итеративно строит последовательность $(x_k)_{k=0}^\infty$ по правилу

$$x_{k+1} := \Pi_Q(x_k - h_k g_k), \quad (1.2)$$

где $g_k \in \partial f(x_k)$ — субградиент функции f в точке x_k , $h_k > 0$ — длина шага, $\Pi_Q(y) := \operatorname{argmin}_{x \in Q} \|x - y\|_2$ — евклидова проекция точки $y \in \mathbb{R}^n$ на множество Q . Заменяя оператор проектирования Π_Q на его определение, итерацию (1.2) можно переписать следующим образом:

$$x_{k+1} = \operatorname{argmin}_{x \in Q} \left\{ \langle g_k, x \rangle + \frac{1}{2h_k} \|x - x_k\|_2^2 \right\}.$$

Отсюда видно, что итерация субградиентного метода имеет определенный геометрический смысл: новая точка x_{k+1} минимизирует на множестве Q линейную модель $f(x_k) + \langle g_k, x - x_k \rangle$ целевой функции плюс регуляризатор $(1/2)\|x - x_k\|_2^2$, запрещающий новой точке x_{k+1} отдаляться слишком далеко от текущей точки x_k . При такой интерпретации возникает естественный вопрос: нельзя ли в качестве регуляризатора вместо квадрата евклидова расстояния использовать какую-либо другую функцию $V(x; x_k)$, т. е. рассматривать итерации вида

$$x_{k+1} = \operatorname{argmin}_{x \in Q} \left\{ \langle g_k, x \rangle + \frac{1}{h_k} V(x; x_k) \right\}?$$

В этом случае траектория $(x_k)_{k=0}^\infty$ метода уже будет другой и можно ожидать, что при «правильном» выборе регуляризатора V , а также длин шагов h_k , новый метод будет обладать более быстрой сходимостью. При этом естественно ожидать, что выбор регуляризатора V должен определяться геометрией рассматриваемого множества Q : например, если $Q = \Delta_n$ — стандартный симплекс, то точки множества Q соответствуют дискретным вероятностным распределениям, и в этом случае более естественным расстоянием между точками множества Q является дивергенция Кульбака–Лейблера

$$V(x; x_k) = \sum_{i=1}^n x_i \ln \frac{x_i}{x_{k,i}},$$

а не обычная евклидова метрика $V(x; x_k) = (1/2)\|x - x_k\|_2^2$, никак не учитывающая особенности множества Q .

Итак, ключевой ингредиент метода зеркального спуска — это функция V . Ясно, что эта функция не может быть совсем произвольной (например, для некоторых функций соответствующая задача оптимизации, определяющая следующую точку x_{k+1} , вообще может не иметь решений). Из приведенных выше рассуждений понятно, что для того, чтобы метод зеркального спуска «правильным образом» обобщал субградиентный метод, функция V должна иметь те же свойства, что и квадрат евклидова расстояния. Оказывается, что основное свойство квадрата евклидова расстояния, отвечающее за сходимость стандартного субградиентного метода, заключается в том, что для некоторой¹ функции $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ справедливо следующее алгебраическое представление:

$$\frac{1}{2} \|y - x\|_2^2 = \omega(y) - \omega(x) - \langle \nabla \omega(x), y - x \rangle,$$

¹В данном случае $\omega(x) := (1/2)\|x\|_2^2$.

причем функция ω сильно выпукла (относительно некоторой нормы). Таким образом, в методе зеркального спуска регуляризирующая функция V задается с помощью другой функции ω согласно выписанному выше алгебраическому выражению. Функция ω называется *прокс-функцией* или *функцией, порождающей расстояние*. Приведем соответствующее формальное определение.

Определение 1.1 (Прокс-функция). Пусть $\|\cdot\|$ — норма в пространстве \mathbb{R}^n (не обязательно евклидова). Пусть $Q \subseteq \mathbb{R}^n$ — выпуклое замкнутое непустое множество. Пусть $\omega : Q \rightarrow \mathbb{R}$ — выпуклая и непрерывная функция. Обозначим через $Q^\circ := \{x \in Q : \partial\omega(x) \neq \emptyset\}$ подмножество Q , на котором функция ω субдифференцируема². Функция ω называется *прокс-функцией* для множества Q , связанной с нормой $\|\cdot\|$, если

(а) Функция ω допускает *непрерывный выбор субградиентов*, т. е. существует функция $\omega' : Q^\circ \rightarrow \mathbb{R}^n$, такая, что для всех $x \in Q^\circ$ выполнено $\omega'(x) \in \partial\omega(x)$, и функция ω' непрерывна на множестве Q° .

(б) Функция ω сильно выпукла с параметром 1 относительно нормы $\|\cdot\|$, т. е.

$$\langle \omega'(x) - \omega'(y), x - y \rangle \geq \|x - y\|^2. \quad (1.3)$$

для всех $x, y \in Q^\circ$.

Пример 1.2 (Евклидова прокс-функция). Пусть $\|\cdot\|$ — евклидова норма $\|x\| := \|x\|_2 := \langle x, x \rangle^{1/2}$. Рассмотрим функцию $\omega : Q \rightarrow \mathbb{R}$, заданную формулой

$$\omega(x) := \frac{1}{2} \|x\|_2^2.$$

Эта функция непрерывная и выпуклая, для нее $Q^\circ = Q$, $\omega'(x) = x$, и неравенство (1.3) переходит в тождественное равенство. Таким образом, ω является прокс-функцией для множества Q , связанной с евклидовой нормой. Эта прокс-функция называется *евклидовой*.

Пример 1.3 (Энтропийная прокс-функция). Пусть $Q := \Delta_n := \{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1\}$ — стандартный симплекс в пространстве \mathbb{R}^n . Пусть также $\|\cdot\|$ — ℓ_1 -норма $\|x\| := \|x\|_1 := \sum_{i=1}^n |x_i|$. Рассмотрим функцию $\omega : \Delta_n \rightarrow \mathbb{R}$, заданную формулой³

$$\omega(x) := \sum_{i=1}^n x_i \ln x_i.$$

Эта функция непрерывная и выпуклая, для нее $Q^\circ = \{x \in \mathbb{R}_{++}^n : \sum_{i=1}^n x_i = 1\}$ — внутренность Δ_n , и $\omega'(x) = (\ln x_i + 1)_{i=1}^n$.

Покажем, что функция ω сильно выпукла с параметром 1 относительно ℓ_1 -нормы, т. е. что для всех $x, y \in \Delta_n^\circ$ справедливо неравенство

$$\sum_{i=1}^n (\ln x_i - \ln y_i)(x_i - y_i) \geq \|x - y\|_1^2. \quad (1.4)$$

Действительно, пусть $x, y \in \Delta_n^\circ$ — произвольные точки. Обозначим $h := y - x$. Заметим, что для доказательства неравенства (1.4) достаточно показать, что⁴

$$\sum_{i=1}^n \frac{h_i^2}{x_i} \geq \|h\|_1^2. \quad (1.5)$$

²Из выпуклого анализа известно, что это множество заведомо содержит относительную внутренность множества Q . Также из выпуклого анализа известно, что выпуклое непустое множество всегда имеет непустую относительную внутренность. Таким образом, при сделанных предположениях множество Q° заведомо непусто.

³Здесь и в дальнейшем выражение вида $x \ln x$ при $x = 0$ будем понимать как ноль.

⁴В самом деле, для разности логарифмов справедливо представление $\ln y_i - \ln x_i = h_i/x_i + o(h_i)$ при $h_i \rightarrow 0$. Чтобы получить неравенство (1.4) из неравенства (1.5), нужно воспользоваться этим представлением и перейти к пределу по $h \rightarrow 0$.

Остается понять, что неравенство (1.5) вытекает из условия $\sum_{i=1}^n x_i = 1$ и неравенства Коши–Буняковского:

$$\sum_{i=1}^n \frac{h_i^2}{x_i} = \left(\sum_{i=1}^n \frac{h_i^2}{x_i} \right) \left(\sum_{i=1}^n x_i \right) \geq \left(\sum_{i=1}^n \frac{|h_i|}{\sqrt{x_i}} \sqrt{x_i} \right)^2 = \left(\sum_{i=1}^n |h_i| \right)^2 = \|h\|_1^2.$$

Таким образом, функция ω является прокс-функцией для множества Δ_n , связанной с ℓ_1 -нормой. Эта прокс-функция называется *энтропийной*.

Имея в распоряжении прокс-функцию ω , можно задать регуляризующую функцию V с помощью рассмотренного выше алгебраического выражения. Это алгебраическое выражение задает функцию, которая называется *дивергенция Брегмана* и встречается не только в контексте метода зеркального спуска.

Определение 1.4 (Дивергенция Брегмана). Пусть $Q \subseteq \mathbb{R}^n$ — непустое замкнутое выпуклое множество, и $\omega : Q \rightarrow \mathbb{R}$ — прокс-функция для множества Q . Обозначим через $Q^\circ := \{x \in Q : \partial\omega(x) \neq \emptyset\}$ подмножество Q , на котором функция ω субдифференцируема. Пусть также $\omega' : Q^\circ \rightarrow \mathbb{R}^n$ — непрерывный выбор субградиентов для функции ω . *Дивергенцией Брегмана* для функции ω (и непрерывного выбора субградиентов ω') называется функция $V_\omega : Q \times Q^\circ \rightarrow \mathbb{R}$, определенная по формуле

$$V_\omega(y; x) := \omega(y) - \omega(x) - \langle \omega'(x), y - x \rangle.$$

Пример 1.5 (Дивергенция Брегмана для евклидовой прокс-функции). Пусть $\omega : Q \rightarrow \mathbb{R}$ — евклидова прокс-функция $\omega(x) = (1/2)\|x\|_2^2$. Тогда

$$\begin{aligned} V_\omega(y; x) &= \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|x\|_2^2 - \langle x, y - x \rangle \\ &= \frac{1}{2}\|y\|_2^2 - \langle x, y \rangle + \frac{1}{2}\|x\|_2^2 \\ &= \frac{1}{2}\|y - x\|_2^2. \end{aligned}$$

Таким образом, дивергенция Брегмана $V_\omega(y; x)$ для евклидовой прокс-функции равна половине квадрата евклидова расстояния между точками x и y .

Пример 1.6 (Дивергенция Брегмана для энтропийной прокс-функции). Пусть $Q := \Delta_n$ — стандартный симплекс, и $\omega : \Delta_n \rightarrow \mathbb{R}$ — энтропийная прокс-функция $\omega(x) := \sum_{i=1}^n x_i \ln x_i$. Тогда

$$\begin{aligned} V_\omega(y; x) &= \sum_{i=1}^n y_i \ln y_i - \sum_{i=1}^n x_i \ln x_i - \sum_{i=1}^n (\ln x_i + 1)(y_i - x_i) \\ &= \sum_{i=1}^n y_i \ln y_i - \sum_{i=1}^n y_i \ln x_i - \sum_{i=1}^n y_i + \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n y_i \ln \frac{y_i}{x_i}. \end{aligned}$$

Заметим, что полученное выражение равно дивергенции Кульбака–Лейблера между дискретными распределениями, соответствующими точкам y и x . Таким образом, дивергенция Брегмана для энтропийной прокс-функции — это дивергенция Кульбака–Лейблера.

Замечание 1.7. Дивергенцию Брегмана часто также называют *расстоянием Брегмана*. Однако, как показывает последний пример, дивергенция Брегмана, вообще говоря, не является симметричной относительно своих аргументов и поэтому не может быть расстоянием. Тем не менее, некоторым свойствам расстояния дивергенция Брегмана все же удовлетворяет.

Например, из сильной выпуклости функции ω следует, что

$$V_\omega(y; x) \geq \frac{1}{2} \|y - x\|^2$$

для всех $y \in Q$ и всех $x \in Q^\circ$ (здесь $\|\cdot\|$ — норма, с которой согласована прокс-функция ω). Поэтому дивергенция Брегмана всюду неотрицательная и обращается в ноль тогда и только тогда, когда оба ее аргумента совпадают.

Имея в распоряжении нужный аналог расстояния между двумя точками — дивергенцию Брегмана — можно наконец выписать формальное определение одного шага метода зеркального спуска.

Определение 1.8 (Шаг зеркального спуска). Пусть $Q \subseteq \mathbb{R}^n$ — выпуклое замкнутое непустое множество, и $\omega : Q \rightarrow \mathbb{R}$ — прокс-функция для множества Q . Обозначим через $Q^\circ := \{x \in Q : \partial\omega(x) \neq \emptyset\}$ подмножество Q , на котором функция ω субдифференцируема. Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция. Результатом *шага зеркального спуска* на множестве Q из точки $\bar{x} \in Q^\circ$ для прокс-функции ω , субградиента $g \in \partial f(\bar{x})$ и длины шага $h \geq 0$, называется точка

$$\text{Mirr}_{Q,\omega}(\bar{x}; g, h) := \operatorname{argmin}_{x \in Q} \{h\langle g, x \rangle + V_\omega(x; \bar{x})\}. \quad (1.6)$$

Замечание 1.9. Поскольку допустимое множество Q в задаче (1.6) выпуклое замкнутое и непустое, а целевая функция $x \mapsto h\langle g, x \rangle + V_\omega(x; \bar{x})$ непрерывная и сильно-выпуклая на этом множестве (в силу сильной выпуклости прокс-функции ω), то задача (1.6) всегда имеет, и при том единственное, решение. Таким образом, шаг зеркального спуска определен корректно.

Пример 1.10 (Шаг зеркального спуска для евклидовой прокс-функции). Пусть $\omega : Q \rightarrow \mathbb{R}$ — евклидова прокс-функция $\omega(x) := (1/2)\|x\|_2^2$. В этом случае $V_\omega(y; x) = (1/2)\|y - x\|_2^2$, и

$$\text{Mirr}_{Q,\omega}(\bar{x}; g, h) = \operatorname{argmin}_{x \in Q} \left\{ h\langle g, x \rangle + \frac{1}{2} \|x - \bar{x}\|_2^2 \right\} = \Pi_Q(\bar{x} - hg).$$

Таким образом, для евклидовой прокс-функции шаг зеркального спуска — то же самое, что и шаг обычного субградиентного метода с проектированием.

Пример 1.11 (Шаг зеркального спуска для энтропийной прокс-функции). Пусть $Q := \Delta_n$ — стандартный симплекс, и $\omega : \Delta_n \rightarrow \mathbb{R}$ — энтропийная прокс-функция $\omega(x) := \sum_{i=1}^n x_i \ln x_i$. Тогда

$$\text{Mirr}_{Q,\omega}(\bar{x}; g, h) = \operatorname{argmin}_{x \in \Delta_n} \left\{ h\langle g, x \rangle + \sum_{i=1}^n x_i \ln \frac{x_i}{\bar{x}_i} \right\}.$$

Эта задача имеет простое аналитическое решение⁵:

$$\text{Mirr}_{Q,\omega}(\bar{x}; g, h) = \frac{\bar{x} \exp(-hg)}{\sum_{i=1}^n \bar{x}_i \exp(-hg_i)},$$

Замечание 1.12. Поясним смысл названия «зеркальный спуск». Пусть $x_+ := \text{Mirr}_{Q,\omega}(\bar{x}; g, h)$ — результат шага зеркального спуска. Для простоты объяснения⁶ будем считать, что внутренность множества Q непустая, точки x_+ и \bar{x} принадлежат внутренности множества Q , и что прокс-функция ω непрерывно дифференцируема всюду на внутренности множества Q . Записывая для задачи (1.6) условие оптимальности первого порядка (равенство нулю градиента), получаем, что точка x_+ определяется следующим нелинейным уравнением:

$$\nabla\omega(x_+) = \nabla\omega(\bar{x}) - hg.$$

Это уравнение имеет определенную геометрическую интерпретацию:

⁵Здесь операции взятия экспоненты от вектора и произведения двух векторов выполняются покомпонентно.

⁶Аналогичную интерпретацию можно дать и без указанных предположений с помощью понятия *сопряженной функции*. Однако не будем здесь этого делать и ограничимся более простым объяснением.

- (a) Градиент $\nabla\omega : \text{int}(Q) \rightarrow \mathbb{R}^n$ прокс-функции ω задает отображение внутренности множества Q (будем называть это «прямым пространством») в пространство \mathbb{R}^n (будем называть его «двойственным пространством»).
- (b) С помощью этого отображения точка \bar{x} , лежащая в «прямом пространстве», преобразуется в точку $\nabla\omega(\bar{x})$, лежащую в «двойственном пространстве».
- (c) Далее в «двойственном пространстве» из точки $\nabla\omega(\bar{x})$ выполняется стандартный градиентный шаг (для вектора g и длины шага h); получается точка $\nabla\omega(x_+)$.
- (d) Наконец, точка $\nabla\omega(x_+)$ из «двойственного пространства» отображается обратно в «прямое пространство» в точку x_+ (с помощью обратного преобразования для $\nabla\omega$; требования, накладываемые на прокс-функцию ω , гарантируют, что отображение $\nabla\omega$ взаимнооднозначно).

Таким образом, сам спуск выполняется в «двойственном пространстве» (на градиентах функции ω), а наблюдаемый процесс, протекаемый в «прямом пространстве» (на точках x), является лишь отражением этого спуска. Отсюда и название «зеркальный спуск».

Заметим, что шаг зеркального спуска определен не из всех точек множества Q , а только из тех точек $\bar{x} \in Q^\circ$, в которых прокс-функция ω субдифференцируема. Например, если множество Q есть стандартный симплекс Δ_n , то шаг зеркального спуска можно выполнять только из внутренних точек симплекса, т. е. таких точек $\bar{x} \in \Delta_n$, в которых все координаты строго положительные. Следующая лемма показывает, что результатом шага зеркального спуска, выполненного из допустимой точки $\bar{x} \in Q^\circ$, может быть только допустимая точка.

Лемма 1.13 (Техническая лемма). Пусть $Q \subseteq \mathbb{R}^n$ — выпуклое замкнутое непустое множество, и $\omega : Q \rightarrow \mathbb{R}$ — прокс-функция для множества Q . Пусть также $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая непрерывно дифференцируемая функция. Обозначим $x_+ := \arg\min_{x \in Q} \{\phi(x) + \omega(x)\}$, и через $Q^\circ := \{x \in Q : \partial\omega(x) \neq \emptyset\}$ — подмножество Q , на котором функция ω субдифференцируема. Тогда x_+ определена корректно (существует и единственная) и, кроме того, $x_+ \in Q^\circ$.

Доказательство. См. Ben-Tal and Nemirovski, книга Lectures on Modern Convex Optimization, 2016, Lemma 5.6.1. \square

Итак, начав однажды с некоторой допустимой точки $x_0 \in Q^\circ$, можно итеративно повторять шаги зеркального спуска, каждый раз выполняя очередной шаг из только что полученной допустимой точки $x_k \in Q^\circ$. Остается лишь понять, как выбрать начальную точку x_0 , чтобы она была допустимой. Для этого достаточно в предыдущей лемме взять в качестве функции ϕ тождественный ноль. В этом случае точка x_0 будет соответствовать минимуму прокс-функции ω на множестве Q . Введем соответствующее определение.

Определение 1.14 (Прокс-центр множества). Пусть $Q \subseteq \mathbb{R}^n$ — выпуклое замкнутое непустое множество, и $\omega : Q \rightarrow \mathbb{R}$ — прокс-функция для множества Q . Прокс-центром множества Q (соответствующим прокс-функции ω) называется точка

$$x_{Q,\omega} := \arg\min_{x \in Q} \omega(x).$$

Пример 1.15 (Прокс-центр для евклидовой прокс-функции). Пусть $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ — евклидова прокс-функция $\omega(x) := (1/2)\|x\|_2^2$. В этом случае прокс-центром множества Q является евклидова проекция точки 0 на множество Q :

$$x_{Q,\omega} = \arg\min_{x \in Q} \|x\|_2 = \Pi_Q(0).$$

Пример 1.16 (Прокс-центр симплекса для энтропийной прокс-функции). Пусть $Q := \Delta_n$ — стандартный симплекс в пространстве \mathbb{R}^n , и пусть $\omega : \Delta_n \rightarrow \mathbb{R}$ — энтропийная прокс-функция $\omega(x) :=$

$\sum_{i=1}^n x_i \ln x_i$. В этом случае прокс-центром множества Q является точка

$$x_{Q,\omega} = \operatorname{argmin}_{x \in \Delta_n} \sum_{i=1}^n x_i \ln x_i.$$

Решая эту оптимизационную задачу аналитически, получаем, что прокс-центр — это точка, соответствующая равномерному распределению:

$$x_{Q,\omega} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right).$$

Теперь у нас имеются все компоненты метода зеркального спуска. Соберем их все вместе и сформулируем схему метода.

Метод зеркального спуска
<p>Вход: выпуклое замкнутое непустое множество $Q \subseteq \mathbb{R}^n$; прокс-функция $\omega : Q \rightarrow \mathbb{R}$ для множества Q; выпуклая функция $f : \mathbb{R}^n \rightarrow \mathbb{R}$; последовательность $(h_k)_{k=0}^\infty$ длин шагов $h_k \geq 0$.</p> <p>Метод:</p> <p>(а) Начать с точки $x_0 := x_{Q,\omega}$ (прокс-центр множества Q).</p> <p>(б) На каждой итерации $k \geq 0$ выбрать произвольный субградиент $g_k \in \partial f(x_k)$ и выполнить из точки x_k шаг зеркального спуска:</p> $x_{k+1} := \operatorname{Mirr}_{Q,\omega}(x_k; g_k, h_k).$

Перейдем к анализу скорости сходимости метода зеркального спуска. Как обычно, ключевым элементом в анализе скорости сходимости различных градиентных методов является лемма о том, как изменяется «расстояние» между некоторой фиксированной точкой u и траекторией x_k метода за одну итерацию. В нашем случае, исходя из конструкции метода, «расстояние» естественно измерять с помощью введенной дивергенции Брегмана.

Лемма 1.17. Пусть $Q \subseteq \mathbb{R}^n$ — выпуклое замкнутое непустое множество, и $\omega : Q \rightarrow \mathbb{R}$ — прокс-функция для множества Q (связанная с некоторой нормой $\|\cdot\|$). Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция, и $(h_k)_{k=0}^\infty$ — последовательность неотрицательных чисел. Пусть также $(x_k)_{k=0}^\infty$ — последовательность точек $x_k \in Q$, построенная методом зеркального спуска для минимизации функции f на множестве Q для последовательности длин шагов $(h_k)_{k=0}^\infty$, и $(g_k)_{k=0}^\infty$ — соответствующая последовательность выбранных методом субградиентов $g_k \in \partial f(x_k)$. Тогда для всех $u \in Q$ и всех $k \geq 0$ справедливо следующее неравенство⁷:

$$V_\omega(u; x_{k+1}) \leq V_\omega(u; x_k) - h_k \langle g_k, x_k - u \rangle + \frac{h_k^2}{2} \|g_k\|_*^2.$$

Доказательство. Пусть $u \in Q$ — произвольная точка, и $k \geq 0$ — произвольный номер.

Воспользуемся определением дивергенции Брегмана, чтобы связать следующую невязку $V_\omega(u; x_{k+1})$ с текущей невязкой $V_\omega(u; x_k)$:

$$\begin{aligned} V_\omega(u; x_{k+1}) &= \omega(u) - \omega(x_{k+1}) - \langle \omega'(x_{k+1}), u - x_{k+1} \rangle \\ &= [\omega(u) - \omega(x_k) - \langle \omega'(x_k), u - x_k \rangle] \\ &\quad + \omega(x_k) - \omega(x_{k+1}) - \langle \omega'(x_{k+1}), u - x_{k+1} \rangle + \langle \omega'(x_k), u - x_k \rangle \\ &= V_\omega(u; x_k) + \omega(x_k) - \omega(x_{k+1}) - \langle \omega'(x_{k+1}), u - x_{k+1} \rangle + \langle \omega'(x_k), u - x_k \rangle. \end{aligned}$$

⁷Здесь и в дальнейшем $\|\cdot\|_*$ — двойственная норма $\|s\|_* := \max_x \{\langle s, x \rangle : \|x\| \leq 1\}$.

Из условия оптимальности первого порядка для шага зеркального спуска получаем, что

$$\langle h_k g_k + \omega'(x_{k+1}) - \omega'(x_k), u - x_{k+1} \rangle \geq 0,$$

или, эквивалентно,

$$\langle \omega'(x_{k+1}) - \omega'(x_k), u - x_{k+1} \rangle \geq h_k \langle g_k, x_{k+1} - u \rangle.$$

Воспользуемся этим неравенством в полученном выше тождестве для дивергенции Брегмана:

$$\begin{aligned} V_\omega(u; x_{k+1}) &= V_\omega(u; x_k) - \langle \omega'(x_{k+1}) - \omega'(x_k), u - x_{k+1} \rangle \\ &\quad + \omega(x_k) - \omega(x_{k+1}) - \langle \omega'(x_{k+1}), u - x_{k+1} \rangle + \langle \omega'(x_k), u - x_k \rangle \\ &\quad + \langle \omega'(x_{k+1}) - \omega'(x_k), u - x_{k+1} \rangle \\ &= V_\omega(u; x_k) - \langle \omega'(x_{k+1}) - \omega'(x_k), u - x_{k+1} \rangle - V_\omega(x_{k+1}; x_k) \\ &\leq V_\omega(x_k; u) - h_k \langle g_k, x_{k+1} - u \rangle - V_\omega(x_{k+1}; x_k). \end{aligned} \tag{1.7}$$

Оценим сумму последних двух слагаемых. Прежде всего, перепишем ее эквивалентным образом:

$$h_k \langle g_k, x_{k+1} - u \rangle + V_\omega(x_{k+1}; x_k) = h_k \langle g_k, x_k - u \rangle + h_k \langle g_k, x_{k+1} - x_k \rangle + V_\omega(x_{k+1}; x_k).$$

Учитывая сильную выпуклость прокс-функции ω , можно записать следующее неравенство:

$$V_\omega(x_{k+1}; x_k) \geq \frac{1}{2} \|x_{k+1} - x_k\|^2.$$

Значит,

$$h_k \langle g_k, x_{k+1} - x_k \rangle + V_\omega(x_{k+1}; x_k) \geq h_k \langle g_k, x_{k+1} - x_k \rangle + \frac{1}{2} \|x_{k+1} - x_k\|^2.$$

Согласно неравенству Фенхеля–Юнга⁸,

$$h_k \langle g_k, x_{k+1} - x_k \rangle + \frac{1}{2} \|x_{k+1} - x_k\|^2 \geq -\frac{h_k^2}{2} \|g_k\|_*^2.$$

В итоге,

$$h_k \langle g_k, x_{k+1} - u \rangle + V_\omega(x_{k+1}; x_k) \geq h_k \langle g_k, x_k - u \rangle - \frac{h_k^2}{2} \|g_k\|_*^2. \tag{1.8}$$

Для завершения доказательства леммы осталось применить полученное неравенство (1.8) в неравенстве (1.7). \square

Из полученной леммы легко получить следующий результат о скорости сходимости метода зеркального спуска.

Теорема 1.18 (Скорость сходимости зеркального спуска). Пусть $Q \subseteq \mathbb{R}^n$ — выпуклое замкнутое непустое множество, и $\omega : Q \rightarrow \mathbb{R}$ — прокс-функция для множества Q (связанная с некоторой нормой $\|\cdot\|$). Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция, и $(h_k)_{k=0}^\infty$ — последовательность неотрицательных чисел. Пусть множество точек минимума X^* функции f на множестве Q непусто, и $f^* := \min_{x \in Q} f(x)$ — соответствующее минимальное значение функции f . Пусть также $(x_k)_{k=0}^\infty$ — последовательность точек $x_k \in Q$, построенная методом зеркального спуска для минимизации функции f на множестве Q для последовательности длин шагов $(h_k)_{k=0}^\infty$, и $(g_k)_{k=0}^\infty$ — соответствующая последовательность выбранных методом субградиентов $g_k \in \partial f(x_k)$. Обозначим через $\bar{x}_k := \operatorname{argmin}_x \{f(x) : x \in \{x_0, \dots, x_{k-1}\}\}$ точку, соответствующую рекордному наблюдаемому значению целевой функции. Пусть также $R > 0$ — число, такое, что $R^2/2 \geq \inf_{x \in X^*} V_\omega(x; x_0)$. Тогда для всех $k \geq 1$ справедливо следующее неравенство:

$$f(\bar{x}_k) - f^* \leq \frac{R^2 + \sum_{i=0}^{k-1} h_i^2 \|g_i\|_*^2}{2 \sum_{i=0}^{k-1} h_i}. \tag{1.9}$$

В частности, если фиксирована желаемая точность $\varepsilon > 0$, все субградиенты функции f равномерно ограничены⁹ по двойственной норме константой $M > 0$ на множестве Q , и длины шагов выбраны

⁸Неравенство Фенхеля–Юнга говорит, что $\langle s, x \rangle \geq \frac{1}{2} \|s\|_*^2 + \frac{1}{2} \|x\|^2$ для всех $s, x \in \mathbb{R}^n$.

⁹Это означает, что для всех $x \in Q$ и всех $g \in \partial f(x)$ выполнено $\|g\|_* \leq M$.

по формуле

$$h_k := \frac{\varepsilon}{M\|g_k\|_*} \quad (1.10)$$

для всех $k \geq 0$, то после

$$K := \frac{M^2 R^2}{\varepsilon^2}$$

шагов, гарантированно будет найдена точка, имеющая ошибку не более ε :

$$f(\bar{x}_k) - f^* \leq \varepsilon$$

для всех $k \geq K$.

Доказательство. Пусть $k \geq 1$ — произвольный номер.

Согласно лемме 1.17, для всех $u \in Q$ и всех $i \geq 0$ справедливо неравенство

$$h_i \langle g_i, x_i - u \rangle \leq V_\omega(u; x_i) - V_\omega(u; x_{i+1}) + \frac{h_i^2}{2} \|g_i\|_*^2.$$

Просуммировав эти неравенства для $0 \leq i \leq k-1$ и отбросив отрицательное слагаемое $-V_\omega(u; x_k)$, получим

$$\sum_{i=0}^{k-1} h_i \langle g_i, x_i - u \rangle \leq V_\omega(u; x_0) + \frac{1}{2} \sum_{i=0}^{k-1} h_i^2 \|g_i\|_*^2. \quad (1.11)$$

для всех $u \in Q$.

Поскольку функция f выпукла, то $\langle g_i, x_i - u \rangle \geq f(x_i) - f(u)$ для всех $u \in Q$ и всех $i \geq 0$. Как следствие,

$$\sum_{i=0}^{k-1} h_i \langle g_i, x_i - u \rangle \geq \sum_{i=0}^{k-1} h_i [f(x_i) - f(u)]$$

для всех $u \in Q$. Кроме того, из определения точки \bar{x}_k , следует неравенство $f(x_i) - f(u) \geq f(\bar{x}_k) - f(u)$ для всех $0 \leq i \leq k-1$. Объединяя это с предыдущим неравенством, получаем, что

$$\sum_{i=0}^{k-1} h_i \langle g_i, x_i - u \rangle \geq [f(\bar{x}_k) - f(u)] \sum_{i=0}^{k-1} h_i \quad (1.12)$$

для всех $u \in Q$.

Комбинируя неравенства (1.11) и (1.12), получаем, что

$$f(\bar{x}_k) - f(u) \leq \frac{2V_\omega(u; x_0) + \sum_{i=0}^{k-1} h_i^2 \|g_i\|_*^2}{2 \sum_{i=0}^{k-1} h_i}$$

для всех $u \in Q$. Для завершения доказательства первой части теоремы осталось перейти в полученном неравенстве к инфимуму по $u \in X^*$.

Перейдем теперь ко второй части теоремы. Выбирая шаги по формуле (1.10) и используя доказанное неравенство (1.9), получаем, что

$$f(\bar{x}_k) - f^* \leq \frac{R^2 + \frac{\varepsilon^2}{M^2} k}{2 \frac{\varepsilon}{M} \sum_{i=0}^{k-1} \frac{1}{\|g_i\|_*}} = \frac{M^2 R^2 + \varepsilon^2 k}{2 \varepsilon M \sum_{i=0}^{k-1} \frac{1}{\|g_i\|_*}}.$$

для всех $k \geq 1$. Используя ограниченность субградиентов, получаем, что

$$f(\bar{x}_k) - f^* \leq \frac{M^2 R^2 + \varepsilon^2 k}{2 \varepsilon k}.$$

для всех $k \geq 1$. Отсюда для всех $k \geq K$ имеем

$$f(\bar{x}_k) - f^* \leq \frac{M^2 R^2 + \varepsilon^2 K}{2\varepsilon K}.$$

Для завершения доказательства осталось воспользоваться определением K . \square

В заключение сравним эффективность зеркального спуска и стандартного субградиентного метода для минимизации на симплексе.

Пример 1.19 (Зеркальный спуск для минимизации на симплексе). Пусть $n \geq 2$, и $Q := \Delta_n$ — стандартный симплекс в пространстве \mathbb{R}^n , и пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция. Рассмотрим две разных прокс-функции для множества Δ_n .

- (а) (Евклидова прокс-функция) Пусть $\omega : \Delta_n \rightarrow \mathbb{R}$ — евклидова прокс-функция $\omega(x) := (1/2)\|x\|_2^2$. В этом случае метод зеркального спуска соответствует стандартному субградиентному методу с проектированием. Воспользуемся теоремой 1.18, чтобы оценить число итераций K_1 , через которое метод гарантировано найдет минимум функции f с точностью ε . Для этого нужно оценить параметр R , участвующий в формулировке теоремы.

Прокс-центр симплекса для евклидовой прокс-функции равен $x_0 = \Pi_{\Delta_n}(0) = (1/n, \dots, 1/n)$. Заметим, что в худшем случае

$$R^2 = 2 \max_{x \in \Delta_n} V_\omega(x; x_0) = \max_{x \in \Delta_n} \|x - x_0\|_2^2.$$

Поскольку максимум выпуклой функции на симплексе достигается в одной из вершин, и все координаты точки x_0 одинаковые, то

$$\begin{aligned} R^2 &= \left\| (1, 0, \dots, 0) - \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right) \right\|_2^2 \\ &= \left(1 - \frac{1}{n} \right)^2 + \frac{n-1}{n^2} = 1 - \frac{2}{n} + \frac{1}{n} \\ &= 1 - \frac{1}{n}. \end{aligned}$$

Таким образом, в данной ситуации $1/2 \leq R \leq 1$, и итоговая оценка числа шагов для достижения точности ε составляет

$$K_1 := O\left(\frac{M_2^2}{\varepsilon^2}\right),$$

где $M_2 > 0$ — константа, равномерно ограничивающая субградиенты по ℓ_2 -норме.

- (б) (Энтропийная прокс-функция) Пусть теперь $\omega : \Delta_n \rightarrow \mathbb{R}$ — энтропийная прокс-функция $\omega(x) := \sum_{i=1}^n x_i \ln x_i$.

В этом случае прокс-центр тот же самый $x_0 = (1/n, \dots, 1/n)$, но

$$R^2 := 2 \max_{x \in \Delta_n} V_\omega(x; x_0) = 2 \max_{x \in \Delta_n} \sum_{i=1}^n x_i \ln(nx_i) = 2 \ln n.$$

Итоговая оценка числа шагов для достижения точности ε составляет

$$K_2 := O\left(\frac{M_\infty^2 \ln n}{\varepsilon^2}\right),$$

где $M_\infty > 0$ — константа, равномерно ограничивающая субградиенты по ℓ_∞ -норме.

Осталось понять, как связаны константы M_2 и M_∞ . Напомним, что для любого вектора $g \in \mathbb{R}^n$ всегда справедливы следующие неравенства:

$$\|g\|_\infty \leq \|g\|_2 \leq \sqrt{n}\|g\|_\infty.$$

Отсюда следует, что константы M_2 и M_∞ связаны неравенствами

$$M_\infty \leq M_2 \leq \sqrt{n}M_\infty.$$

В итоге эффективности K_1 и K_2 рассматриваемых методов связаны между собой неравенствами

$$K_2 \leq O(K_1 \ln n) \leq O(nK_2).$$

Это означает, что зеркальный спуск с энтропийной прокс-функцией всегда будет делать не более, чем в $O(\ln n)$ раз больше итераций, чем предписано стандартному градиентному спуску. Кроме того, возможны ситуации, когда последнее неравенство переходит в равенство (есть точки, в которых субградиент g имеет одинаковые компоненты). В этом случае зеркальный спуск будет работать в $O(n/\ln n)$ раз быстрее, чем стандартный субградиентный спуск, что при больших n может быть очень существенно.