



Московский государственный университет имени М.В. Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра системного анализа

Сафонова Елизавета Сергеевна

Сглаживание функций посредством преобразований Моро-Бертсекаса

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:

к.ф.-м.н., доцент

С. Н. Смирнов

Москва, 2023

Содержание

1	Инструментарий	4
1.1	Понятие выпуклой функции	4
1.2	Инфимальная конволюция	4
1.3	Преобразование Моро	5
1.4	Связь преобразования Моро и проксимального оператора	8
1.5	Преобразование Бертсекаса	9
2	Иллюстративные примеры	10
2.1	Пример 1	10
2.2	Пример 2	12
2.3	Пример 3	13
3	Случай в \mathbb{R}^2	16
3.1	Проведение экспериментов	17
3.2	Точность аппроксимации	19
4	Заключение	20
	Литература	21

Введение

С возникновением интегрального и дифференциального исчисления в трудах И. Ньютона и Г. В. Лейбница появилась возможность решать широкий класс задач, описываемых гладкими функциями. Данный аппарат удовлетворял потребностям науки и техники более двух столетий. Однако постепенно вопросы как теоретического, так и прикладного характера привели к исследованию недифференцируемых функций.

Первым, кто «прикоснулся» к негладкому анализу, был П. Л. Чебышёв. В 1853 году он решил задачу нахождения алгебраического многочлена n -ой степени со старшим коэффициентом единица, наименее отклоняющегося от нуля на заданном отрезке [1]. С тех пор интерес к проблематике негладкого анализа не угасает: появляются всё новые, более сложные реальные задачи, которые требуют усовершенствования математических моделей, алгоритмов, поиска новых подходов и инструментов. Сегодня же необходимость в сглаживании возникает в практических задачах финансов, экономики, медицины, инженерии и др.

В работе рассматриваются два преобразования: Моро¹ и Бертсекаса. Их реализация обладает рядом существенных преимуществ при поиске решения той или иной поставленной задачи.

Если говорить про преобразование Моро, то в [10] было показано, что оно обладает многими важными дифференциальными свойствами. В своем первоначальном виде огибающая Моро была представлена в середине 1960-х годов Ж. Моро [11]. Применение преобразования можно встретить во многих приложениях: оптимизация, нелинейный анализ, машинное обучение и т.д. [4–8]. Также, в качестве примера, можно привести задачу оптимального управления системами, описываемыми уравнениями в частных производных при наличии ограничений на состояние системы. В этом случае для дифференциальной задачи строится функция Лагранжа и затем находятся ее стационарные точки. Основная трудность связана с возможным отсутствием какой-либо гладкости множителей Лагранжа. Именно поэтому используются методы регуляризации дифференциальных задач. Такой подход, в частности, можно найти в статье [9], в которой авторы отмечают, что подзадачи, получившиеся в результате регуляризации Моро–Иосиды, решаются эффективно полугладкими методами Ньютона.

В качестве более общего варианта выступает преобразование, предложенное Д.Бертсекасом. Он исследовал подходы к сглаживанию для решения одной из важных задач оптимизации, которая называется минимаксной. Найденный метод решения использовал преобразование и состоял в замене негладких функций их гладкими приближениями. Он впервые был введен в [12] и обобщен в [13]. И если в аналитической форме преобразования, предложенного Моро, под инфимумом² стоит сумма самой

¹Преобразование Моро называют также огибающей Моро или же регуляризацией Моро— Иосиды.

²Имея в виду, что преобразования производятся с выпуклыми функциями.

функции и $\frac{1}{2}$ квадрата нормы (в обобщенном случае — сумма функции и квадратичной формы с положительно определенной матрицей), то Бертсекас добавил сдвиг — скалярное произведение и ввел параметризацию. Таким образом, в результате применения преобразования получается дифференцируемая задача, аппроксимирующая исходную.

Целью исследования является рассмотрение процедуры сглаживания собственных выпуклых функций, проведение серии экспериментов по сглаживанию выбранных функций в одномерном и двумерном случаях.

Структура работы

В разделе 1 вводятся обозначения и определения, описывается необходимый инструментарий.

В разделе 2 приводятся иллюстративные примеры, осуществляется аналитический вывод формул для каждого случая.

В разделе 3 рассматривается двумерный случай и проводятся эксперименты.

В разделе 4 описаны результаты и обозначены перспективы дальнейших возможных исследований.

1 Инструментарий

1.1 Понятие выпуклой функции

Первые «негладкие» методы исследования недифференцируемых функций появились в рамках выпуклого анализа [2, 3], который, наряду с теорией минимакса, послужил основой для формирования негладкого анализа.

Обозначим через $\overline{\mathbb{R}} = \mathbb{R}^n \cup \{-\infty, +\infty\}$ расширенную вещественную прямую. Будем рассматривать функции $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$. С каждой такой функцией f можно связать множества

$$\begin{aligned} \text{epi } f &= \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \alpha\}, \\ \text{dom } f &= \{x \in \mathbb{R}^n : f(x) < +\infty\}, \end{aligned}$$

называемые соответственно надграфиком функции f и ее эффективным множеством.

Определение 1. Функция f называется собственной, если $\text{dom } f \neq \emptyset$ и $f(x) > -\infty \forall x$. Функция, не являющаяся собственной, называется несобственной.

Определение 2. Функция f называется выпуклой, если ее надграфик есть выпуклое множество в \mathbb{R}^{n+1} . Функция f называется вогнутой, если функция $(-f)$ является выпуклой.

1.2 Инфимальная конволюция

Будем следовать обозначениям, введенным в [3].

Определение 3. Пусть $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$. Инфимальной конволюцией собственных выпуклых функций f_1, f_2 называется функция

$$f(x) = (f_1 \square f_2)(x) = \inf_{\substack{x_1+x_2=x \\ x_1, x_2 \in \mathbb{R}^n}} \{f_1(x_1) + f_2(x_2)\} = \inf_{x_1 \in \mathbb{R}^n} \{f_1(x_1) + f_2(x - x_1)\}.$$

Функция f является выпуклой на \mathbb{R}^n . Операция взятия инфимальной конволюции двух выпуклых функции коммутативна и ассоциативна. Также можно переформулировать определение (3) в терминах сложения надграфиков функций:

$$(f_1 \square f_2)(x) = \inf \{\mu \in \mathbb{R} \mid (x, \mu) \in \text{epi } f_1 + \text{epi } f_2\}.$$

Определение 4. Супремальная конволюция собственных вогнутых функций f_1 и f_2 определяется посредством формулы

$$f_1 \sqcup f_2(x) = \sup \{f_1(x_1) + f_2(x_2) : x_1 + x_2 = x\},$$

или в терминах сложения подграфиков: $f_1 \square f_2(x) = \sup \{y : (x, y) \in (\text{hypo } f_1 + \text{hypo } f_2)\}$.

Супремальная конволюция связана с инфимальной конволюцией собственных выпуклых функций соотношением $f_1 \square f_2 = -(-f_1) \square (-f_2)$.

1.3 Преобразование Моро

Недифференцируемые выпуклые функции могут быть аппроксимированы гладкими функциями различными методами. Одним из самых популярных способов сглаживания является преобразование Моро, которое называют также огибающей Моро или же регуляризацией Моро-Иосиды. Дадим характеристику этому инструменту.

Здесь и далее будем считать, что $\|x\| = \sqrt{\langle x, x \rangle}$, $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$.

Определение 5. Огибающая Моро для $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ с параметром $\lambda > 0$ определяется как:

$$\mathfrak{M}_\lambda[f](y) = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2\lambda} \|x - y\|^2 \right\}.$$

Можно заметить, что аппроксимация Моро выпуклой функции представляет собой инфимальную конволюцию f с квадратичной функцией $g(x) = \frac{1}{2\lambda} \|x\|^2$:

$$\mathfrak{M}_\lambda[f](x) = \inf_{x_1, x_2} \{f(x_1) + g(x_2) : x_1 + x_2 = x\} = \inf_{x_1 \in \mathbb{R}^n} \{f(x_1) + g(x - x_1)\} = (f \square g)(x).$$

При этом иногда вместо $\frac{1}{2\lambda} \|x\|^2$ используют положительно определенную квадратичную форму $\langle Bx, x \rangle$.

Обоснуем сглаживание f посредством применения $\mathfrak{M}_\lambda[f]$. Обратимся к понятию сопряжения функции.

Определение 6. (Преобразование Фенхеля) Пусть задана функция $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$. Функцией, сопряженной к f , называется функция, определенная формулой

$$f^*(x^*) = \sup_x (\langle x^*, x \rangle - f(x)). \quad (1)$$

Выражение (1) известно как преобразование Лежандра-Юнга-Фенхеля или просто преобразование Фенхеля.

Утверждение 1. Для функций f_1, f_2 имеет место

$$(f_1 \square f_2)^* = f_1^* + f_2^*.$$

Доказательство. Для произвольного x^* имеем:

$$\begin{aligned}
(f_1 \square f_2)^*(x^*) &= \sup_x \{ \langle x^*, x \rangle - (f_1 \square f_2)(x) \} = \\
&= \sup_x \left\{ \langle x^*, x \rangle + \sup_{x_1+x_2=x} \{ -f_1(x_1) - f_2(x_2) \} \right\} = \\
&= \sup_x \sup_{x_1+x_2=x} \{ \langle x^*, x_1 \rangle - f_1(x_1) + \langle x^*, x_2 \rangle - f_2(x_2) \} = \\
&= \sup_{x_1, x_2} \{ \langle x^*, x_1 \rangle - f_1(x_1) + \langle x^*, x_2 \rangle - f_2(x_2) \} = f_1^*(x^*) + f_2^*(x^*).
\end{aligned}$$

□

Утверждение 2. Для $f(x) = \|x\|^2$ верно $f^{**} = f$:

Доказательство.

$$\begin{aligned}
f^*(x^*) &= \sup_x \{ \langle x^*, x \rangle - f(x) \} = \sup_x \{ \langle x^*, x \rangle - \|x\|^2 \} = \left[x = \frac{x^*}{2} \right] = \frac{1}{4} \|x^*\|^2, \\
f^{**}(x) &= \sup_x \{ \langle x, x^* \rangle - f^*(x^*) \} = \sup_x \left\{ \langle x, x^* \rangle - \frac{1}{4} \|x^*\|^2 \right\} = \left[x^* = 2x \right] = \|x\|^2.
\end{aligned}$$

□

Утверждение 3. Если B — симметричная, положительно определенная матрица размера $n \times n$, $f(x) = \langle Bx, x \rangle$, то верно $f^{**} = f$:

Доказательство.

$$\begin{aligned}
f^*(x^*) &= \sup_x \{ \langle x^*, x \rangle - f(x) \} = \sup_x \{ \langle x^*, x \rangle - \langle Bx, x \rangle \} = \left[x = \frac{1}{2} B^{-1} x^* \right] = \frac{1}{4} \langle x^*, B^{-1} x^* \rangle, \\
f^{**}(x) &= \sup_x \{ \langle x, x^* \rangle - f^*(x^*) \} = \sup_x \left\{ \langle x, x^* \rangle - \frac{1}{4} \langle x^*, B^{-1} x^* \rangle \right\} = \left[x^* = 2Bx \right] = \langle Bx, x \rangle.
\end{aligned}$$

□

Для произвольной выпуклой функции f умножение на положительный скаляр $\lambda > 0$ определяется соотношением:

$$(\lambda f)(x) = \lambda f(x).$$

Непосредственно вычисляется, что

$$(\lambda f)^*(x^*) = \lambda f^*\left(\frac{x^*}{\lambda}\right).$$

$$\mathfrak{M}_\lambda[f]^* = (f \square \frac{1}{2\lambda} \|\cdot\|^2)^* = f^* + \frac{\lambda}{2} \|\cdot\|^2$$

$$\mathfrak{M}_\lambda[f]^{**} = (f^* + \frac{\lambda}{2} \|\cdot\|^2)^*$$

Теорема 1. (Фенхеля-Моро) Пусть функция f — выпуклая, замкнутая, собственная. Тогда $f^{**} = f$.

Доказательство. Можно найти в [2] [стр. 56-57]. □

Значит:

$$\begin{aligned}\mathfrak{M}_\lambda[f]^{**} &= \mathfrak{M}_\lambda[f] \\ \mathfrak{M}_\lambda[f] &= (f^* + \frac{\lambda}{2} \|\cdot\|^2)^*\end{aligned}\tag{2}$$

В общем случае функция φ^* , сопряженная к замкнутой собственной выпуклой функции φ , является гладкой, когда φ сильно выпукла. Получается, преобразование Моро можно интерпретировать как получение гладкого приближения функции путем взятия ее сопряжения, добавления регуляризации, а затем повторного сопряжения. Без регуляризации это преобразование просто дало бы исходную функцию, но с квадратичной регуляризацией — «плавное» приближение. То есть в действительности огибающая Моро является сглаженной или регуляризованной формой функции.

$\mathfrak{M}_\lambda[f]$ дает аппроксимированное сглаженное «описание» f , поэтому формула (2) может играть важную роль в получении апостериорной оценки численной погрешности расчета.

Рассмотрим обратимость преобразования. Следующее утверждение дает «рецепт» поиска обратной функции.

Утверждение 4. [15] Для $\lambda > 0$ и выпуклой полунепрерывной снизу функции f , такой, что $\mathfrak{M}_\lambda[f] = g$, выполнено $f = -\mathfrak{M}_\lambda[-g]$.

Доказательство. Сначала рассмотрим частный случай $\lambda = 1$, из которого будет следовать равенство для остальных $\lambda > 0$. Будем обозначать $\mathfrak{M}_1[f] = \mathfrak{M}[f]$. Пусть $\mathfrak{M}[f](y) = g(y)$, тогда

$$\begin{aligned}g(y) = \mathfrak{M}[f](y) &= \inf_x \left\{ f(x) + \frac{1}{2} \|x - y\|^2 \right\} = \frac{1}{2} \|y\|^2 + \inf_x \left\{ f(x) - \langle x, y \rangle + \frac{1}{2} \|x\|^2 \right\} = \\ &= \frac{1}{2} \|y\|^2 - \sup_x \left\{ \langle x, y \rangle - f(x) - \frac{1}{2} \|x\|^2 \right\}.\end{aligned}$$

Для простоты обозначений определим функцию $p_2(x) = \frac{1}{2} \|x\|^2$, тогда предыдущее выражение примет вид:

$$g = p_2 - (f + p_2)^*.$$

Теорема Фенхеля-Моро справедлива для f и для $f + p_2$, поэтому можно записать:

$$f = -p_2 + (p_2 - g)^*.$$

С учетом предыдущих результатов получаем:

$$\begin{aligned} f(x) &= -\frac{1}{2}\|x\|^2 + \sup_y \left\{ \langle x, y \rangle - \frac{1}{2}\|y\|^2 + g(y) \right\} = \\ &= \sup_y \left\{ -\frac{1}{2}\|x - y\|^2 + g(y) \right\} = -\inf_y \left\{ -g(y) + \frac{1}{2}\|x - y\|^2 \right\} = -\mathfrak{M}[-g](x). \end{aligned}$$

Поэтому, если $\mathfrak{M}[f] = g$, то $f = -\mathfrak{M}[-g]$. Чтобы применить эти рассуждения к $\lambda \neq 1$, используем свойство:

$$\lambda \mathfrak{M}_\lambda \left[\frac{1}{\lambda} f \right] = \mathfrak{M}[f]$$

Объединяя выражения выше, находим, что

$$\mathfrak{M}_\lambda \left[\frac{1}{\lambda} f \right] = \frac{1}{\lambda} g,$$

$$\frac{1}{\lambda} f = -\mathfrak{M}_\lambda \left[-\frac{1}{\lambda} g \right],$$

что завершает вывод, поскольку коэффициент $\frac{1}{\lambda}$ может быть включен в f и g . \square

Огибающая Моро также имеет важные приложения в математической оптимизации: минимизация по $\mathfrak{M}_\lambda[f]$ и сведение к минимуму по f являются эквивалентными задачами в том смысле, что набор минимизаторов $\mathfrak{M}_\lambda[f]$ и f одинаков:

$$\inf_y \mathfrak{M}_\lambda[f](y) = \inf_y \inf_x \left\{ f(x) + \frac{1}{2\lambda}\|x - y\|^2 \right\} = \inf_x \inf_y \left\{ f(x) + \frac{1}{2\lambda}\|x - y\|^2 \right\} = \inf_x f(x).$$

1.4 Связь преобразования Моро и проксимального оператора

Определение 7. Проксимальный оператор $\text{prox}_f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ функции f определяется как:

$$\text{prox}_f(y) = \underset{x}{\operatorname{argmin}} \left(f(x) + \frac{1}{2}\|x - y\|^2 \right).$$

Функция, минимизированная в правой части, сильно выпукла и не везде бесконечна, поэтому у нее есть единственный минимизатор для каждого $y \in \mathbb{R}^n$ (даже когда $\operatorname{dom} f \subsetneq \mathbf{R}^n$) [14].

На практике чаще всего используется проксимальный оператор масштабируемой функции λf , где $\lambda > 0$:

$$\text{prox}_{\lambda f}(y) = \underset{x}{\operatorname{argmin}} \left(f(x) + \frac{1}{2\lambda}\|x - y\|^2 \right),$$

который также называется проксимальным оператором f с параметром λ .

Установим связь между нашими объектами исследования, для этого потребуется вычислить градиент.

$$\nabla \mathfrak{M}_\lambda[f](y) = \frac{\partial}{\partial y} \inf_x \left[f(x) + \frac{1}{2\lambda} \|x - y\|^2 \right] = \frac{\partial}{\partial y} \left[f(\hat{x}) + \frac{1}{2\lambda} \|\hat{x} - y\|^2 \right] = \frac{y - \hat{x}}{\lambda}.$$

Тогда градиент огибающей Моро задается формулой:

$$\nabla \mathfrak{M}_\lambda[f](y) = \frac{1}{\lambda} (y - \text{prox}_{\lambda f}(y)).$$

Или, если преобразовать:

$$\text{prox}_{\lambda f}(y) = y - \lambda \nabla \mathfrak{M}_\lambda[f](y),$$

что показывает, что проксимальный оператор можно рассматривать как шаг градиента с размером шага λ для минимизации $\mathfrak{M}_\lambda[f]$. Также можно заметить, что т.к. $y = \text{prox}_f(y) + \text{prox}_{f^*}(y)$, то выполнено соотношение:

$$\text{prox}_f(y) = \nabla \mathfrak{M}[f^*](y).$$

1.5 Преобразование Бертсекаса

Преобразование Бертсекаса $\mathfrak{B}_{\lambda, \mu}[f]$ функции f определяется как инфимальная конволюция функций f и \tilde{g} , где $\tilde{g}(x) = \frac{1}{2\lambda} \|x\|^2 + \langle \mu, x \rangle$, $\lambda > 0$, $\mu \in \mathbb{R}^n$. Поскольку $g(0) = \tilde{g}(0) = 0$, то $\mathfrak{M}_\lambda[f] \leq f$ и $\mathfrak{B}_{\lambda, \mu}[f] \leq f$. Легко видеть, что эти два преобразования связаны: с одной стороны, $\mathfrak{M}_\lambda[f] = \mathfrak{B}_{\lambda, 0}[f]$, а с другой стороны:

$$\mathfrak{B}_{\lambda, \mu}[f](y) = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \langle \mu, x - y \rangle + \frac{1}{2\lambda} \|x - y\|^2 \right\}.$$

Покажем, что из $\lambda \rightarrow 0$ следует $\mathfrak{B}_{\lambda, \mu}[f](y) \rightarrow f(y)$. Обозначим:

$$\varphi_1(x) = f(x),$$

$$\varphi_2(x) = \langle \mu, x - y \rangle + \frac{1}{2\lambda} \|x - y\|^2,$$

Тогда выражение $\mathfrak{B}_{\lambda, \mu}[f](y)$ перепишется как:

$$\inf_{x \in \mathbb{R}^n} \{ \varphi_1(x) + \varphi_2(x) \} = \sup_{x^* \in \mathbb{R}^n} \{ -\varphi_1^*(x^*) - \varphi_2^*(-x^*) \},$$

$$\varphi_1^*(x^*) = f^*(x^*),$$

$$\begin{aligned}
\varphi_2^*(-x^*) &= \sup_x \{ \langle -x^*, x \rangle - \varphi_2(x) \} = \sup_x \left\{ \langle -x^*, x \rangle - \langle \mu, x - y \rangle - \frac{1}{2\lambda} \|x - y\|^2 \right\} = \\
&= \langle \mu, y \rangle - \frac{1}{2\lambda} \|y\|^2 + \sup_x \left\{ \left\langle \frac{y}{\lambda} - x^* - \mu, x \right\rangle - \frac{1}{2\lambda} \|x\|^2 \right\} = \left[x = y - \lambda(x^* + \mu) \right] = \\
&= \langle \mu, y \rangle - \frac{1}{2\lambda} \|y\|^2 + \frac{1}{2\lambda} \|y\|^2 - \langle y, x^* \rangle - \langle \mu, y \rangle + \frac{\lambda}{2} \|x^* + \mu\|^2 = -\langle y, x^* \rangle + \frac{\lambda}{2} \|x^* + \mu\|^2.
\end{aligned}$$

$$\begin{aligned}
\mathfrak{B}_{\lambda, \mu}[f](y) &= \sup_{x^* \in \mathbb{R}^n} \left\{ -f^*(x^*) + \langle y, x^* \rangle - \frac{\lambda}{2} \|x^* + \mu\|^2 \right\}, \\
\lim_{\lambda \rightarrow 0} \mathfrak{B}_{\lambda, \mu}[f](y) &= f^{**}(y) = f(y), \quad \forall y, \mu \in \mathbb{R}^n.
\end{aligned}$$

2 Иллюстративные примеры

В этом разделе рассматривается одномерный случай. Приводятся конкретные примеры, дающие возможность лучше «почувствовать» объект исследования. Производится аналитический вывод сглаживающего преобразования для каждой выбранной функции. Устанавливается связь между изменениями параметров и поведением функции. Экспериментальным путем подтверждается вывод для λ , полученный в разделе 1.5.

2.1 Пример 1

Рассмотрим $f(x) = \max\{x, 0\}$:

$$\begin{aligned}
\mathfrak{B}_{\lambda, \mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ \max\{x, 0\} + \frac{1}{2\lambda} \|x - y\|^2 + \langle \mu, x - y \rangle \right\} = \\
&= \inf_{x \in \mathbb{R}^n} \left\{ \max\{x, 0\} + \frac{1}{2\lambda} |x - y|^2 + \mu(x - y) \right\}.
\end{aligned}$$

1) При $y \geq \lambda(\mu + 1)$ получаем:

$$\begin{aligned}
\mathfrak{B}_{\lambda, \mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ x + \frac{1}{2\lambda} (x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\
&= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} x^2 + x \left(1 + \mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = y - \lambda(\mu + 1) \right] = y - \frac{\lambda(\mu + 1)^2}{2}.
\end{aligned}$$

2) При $y \leq \lambda\mu$ имеем:

$$\begin{aligned}
\mathfrak{B}_{\lambda, \mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} (x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\
&= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} x^2 + x \left(\mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = y - \lambda\mu \right] = -\frac{\lambda\mu^2}{2}.
\end{aligned}$$

3) При $\lambda\mu \leq y \leq \lambda(\mu + 1)$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda}(x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda}x^2 + x \left(\mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = 0 \right] = \frac{y^2}{2\lambda} - \mu y.\end{aligned}$$

$$\mathfrak{B}_{\lambda,\mu}[f](y) = \begin{cases} y - \frac{\lambda(\mu+1)^2}{2}, & y \geq \lambda(\mu + 1), \\ \frac{y^2}{2\lambda} - \mu y, & \lambda\mu \leq y \leq \lambda(\mu + 1), \\ -\frac{\lambda\mu^2}{2}, & y \leq \lambda\mu. \end{cases}$$

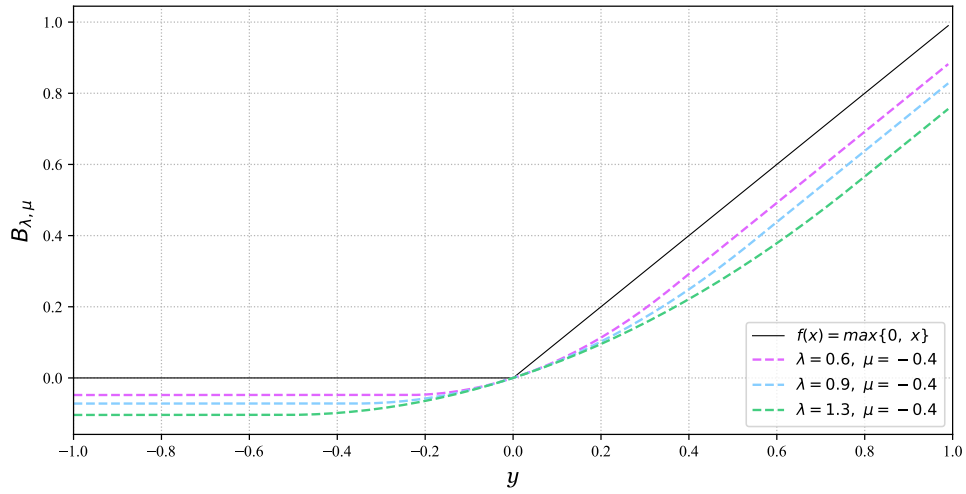


Рис. 1: Сглаживание функции $\max\{0, x\}$ преобразованием Бертсекаса: фиксирование μ , варьирование λ .

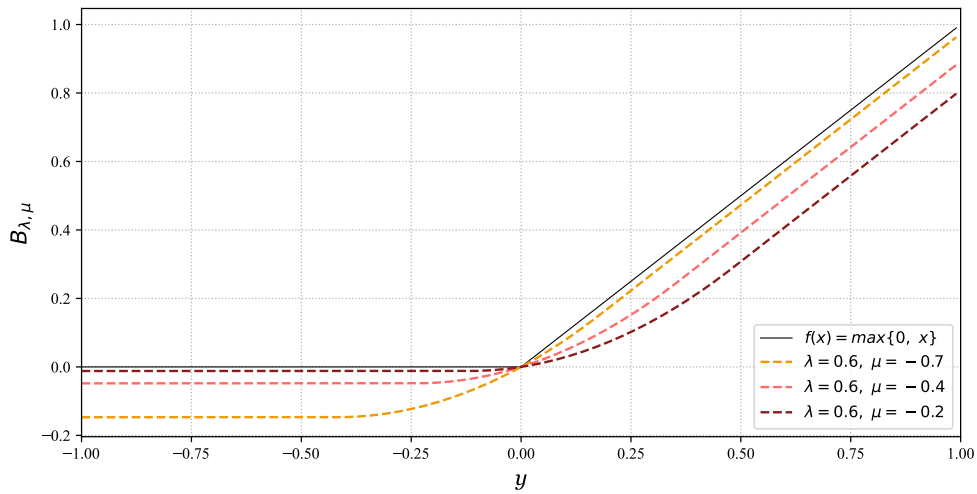


Рис. 2: Сглаживание функции $\max\{0, x\}$ преобразованием Бертсекаса: фиксирование λ , варьирование μ .

2.2 Пример 2

Рассмотрим $f(x) = |x|$:

$$\mathfrak{B}_{\lambda,\mu}[f](y) = \begin{cases} y - \frac{\lambda(\mu+1)^2}{2}, & y \geq \lambda(\mu+1), \\ \frac{y^2}{2\lambda} - \mu y, & \lambda(\mu-1) \leq y \leq \lambda(\mu+1), \\ -y - \frac{\lambda(\mu-1)^2}{2}, & y \leq \lambda(\mu-1). \end{cases}$$

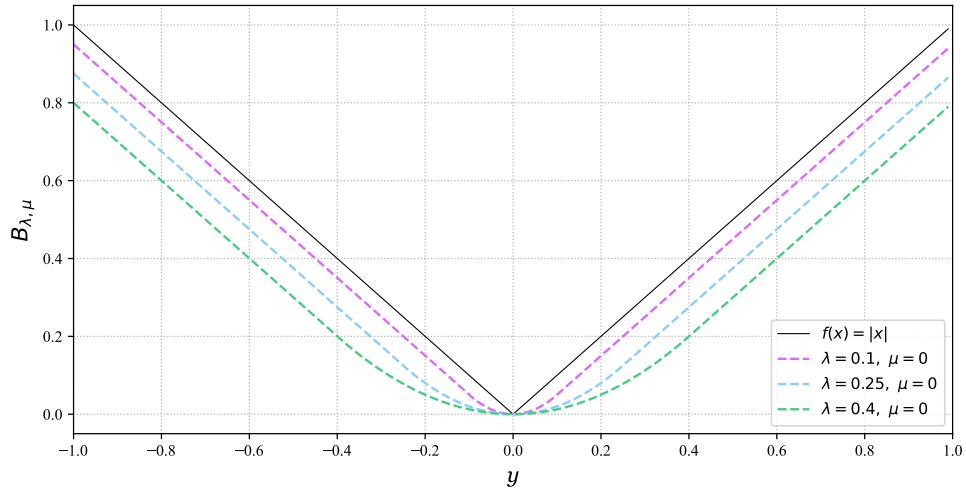


Рис. 3: Сглаживание функции $|x|$ преобразованием Бертсекаса: фиксирование μ , варьирование λ .

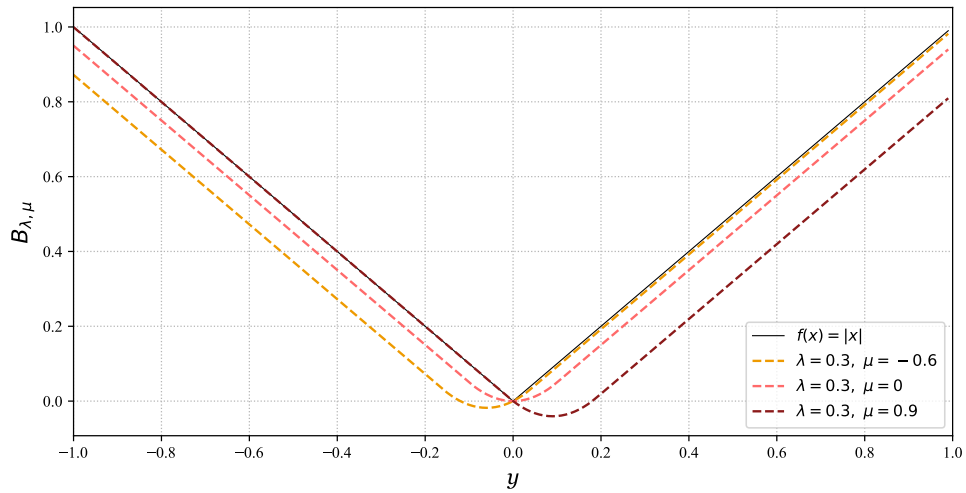


Рис. 4: Сглаживание функции $|x|$ преобразованием Бертсекаса: фиксирование λ , варьирование μ .

Замечание 1. Поскольку $\min\{f(x), 0\} = -\max\{-f(x), 0\}$ и $|f(x)| = \max\{f(x), 0\} + \max\{-f(x), 0\}$, значит, «выводить» функцию сглаживания для операторов \min и $|\cdot|$ не требуется.

Замечание 2. При данном подходе функция сглаживания принимает то же значение в точке перегиба, но в оставшейся части — отклоняется от исходной функции.

2.3 Пример 3

Рассмотрим $f(x) = \max\{|x|, x^2\}$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ \max\{|x|, x^2\} + \frac{1}{2\lambda} \|x - y\|^2 + \langle \mu, x - y \rangle \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \max\{|x|, x^2\} + \frac{1}{2\lambda} |x - y|^2 + \mu(x - y) \right\}.\end{aligned}$$

1) При $y \geq 1 + \lambda(\mu + 2)$ или $y \leq -1 + \lambda(\mu - 2)$ получаем:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ x^2 + \frac{1}{2\lambda} (x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \left(\frac{1}{2\lambda} + 1 \right) x^2 + x \left(\mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = \frac{y - \mu\lambda}{1 + 2\lambda} \right] = \frac{2y^2 - 4\lambda\mu y - \mu^2\lambda}{2(1 + 2\lambda)}.\end{aligned}$$

2) При $\lambda(\mu + 1) \leq y \leq 1 + \lambda(\mu + 1)$ имеем:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ x + \frac{1}{2\lambda} (x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} x^2 + x \left(1 + \mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = y - \lambda(\mu + 1) \right] = y - \frac{\lambda(\mu + 1)^2}{2}.\end{aligned}$$

3) При $\lambda(\mu - 1) - 1 \leq y \leq \lambda(\mu - 1)$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ -x + \frac{1}{2\lambda} (x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} x^2 + x \left(\mu - 1 - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = y - \lambda(\mu - 1) \right] = -y - \frac{\lambda(\mu - 1)^2}{2}.\end{aligned}$$

4) При $y \geq 1 + \lambda(\mu + 1)$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ x + \frac{1}{2\lambda} (x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} x^2 + x \left(1 + \mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = 1 \right] = \frac{(y - 1)^2}{2\lambda} - \mu(y - 1) + 1.\end{aligned}$$

5) При $y \leq \lambda(\mu + 1)$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ x + \frac{1}{2\lambda} (x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda} x^2 + x \left(1 + \mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = 0 \right] = \frac{y^2}{2\lambda} - \mu y.\end{aligned}$$

6) При $y \geq \lambda(\mu - 1)$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ -x + \frac{1}{2\lambda}(x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda}x^2 + x \left(\mu - 1 - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = 0 \right] = \frac{y^2}{2\lambda} - \mu y.\end{aligned}$$

7) При $y \leq \lambda(\mu - 1) - 1$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ -x + \frac{1}{2\lambda}(x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda}x^2 + x \left(\mu - 1 - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = -1 \right] = \frac{(y+1)^2}{2\lambda} - \mu(y+1) + 1.\end{aligned}$$

8) При $y \leq \lambda(\mu + 1) + 1$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ x^2 + \frac{1}{2\lambda}(x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \left(\frac{1}{2\lambda} + 1 \right) x^2 + x \left(\mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = 1 \right] = \frac{(y-1)^2}{2\lambda} - \mu(y-1) + 1.\end{aligned}$$

9) При $y \geq \lambda(\mu - 2) - 1$:

$$\begin{aligned}\mathfrak{B}_{\lambda,\mu}[f](y) &= \inf_{x \in \mathbb{R}^n} \left\{ x^2 + \frac{1}{2\lambda}(x^2 - 2xy + y^2) + \mu x - \mu y \right\} = \\ &= \inf_{x \in \mathbb{R}^n} \left\{ \left(\frac{1}{2\lambda} + 1 \right) x^2 + x \left(\mu - \frac{y}{\lambda} \right) + \frac{y^2}{2\lambda} - \mu y \right\} = \left[x = -1 \right] = \frac{(y+1)^2}{2\lambda} - \mu(y+1) + 1.\end{aligned}$$

$$\mathfrak{B}_{\lambda,\mu}[f](y) = \begin{cases} \frac{2y^2 - 4\lambda\mu y - \mu^2\lambda}{2(1+2\lambda)}, & y \geq 1 + \lambda(\mu + 2), y \leq -1 + \lambda(\mu - 2), \\ y - \frac{\lambda(\mu+1)^2}{2}, & \lambda(\mu + 1) \leq y \leq 1 + \lambda(\mu + 1), \\ -y - \frac{\lambda(\mu-1)^2}{2}, & \lambda(\mu - 1) - 1 \leq y \leq \lambda(\mu - 1), \\ \frac{(y-1)^2}{2\lambda} - \mu(y-1) + 1, & 1 + \lambda(\mu + 1) \leq y \leq \lambda(\mu + 2) + 1, \\ \frac{y^2}{2\lambda} - \mu y, & \lambda(\mu - 1) \leq y \leq \lambda(\mu + 1), \\ \frac{(y+1)^2}{2\lambda} - \mu(y+1) + 1, & \lambda(\mu - 2) - 1 \leq y \leq \lambda(\mu - 1) - 1. \end{cases}$$

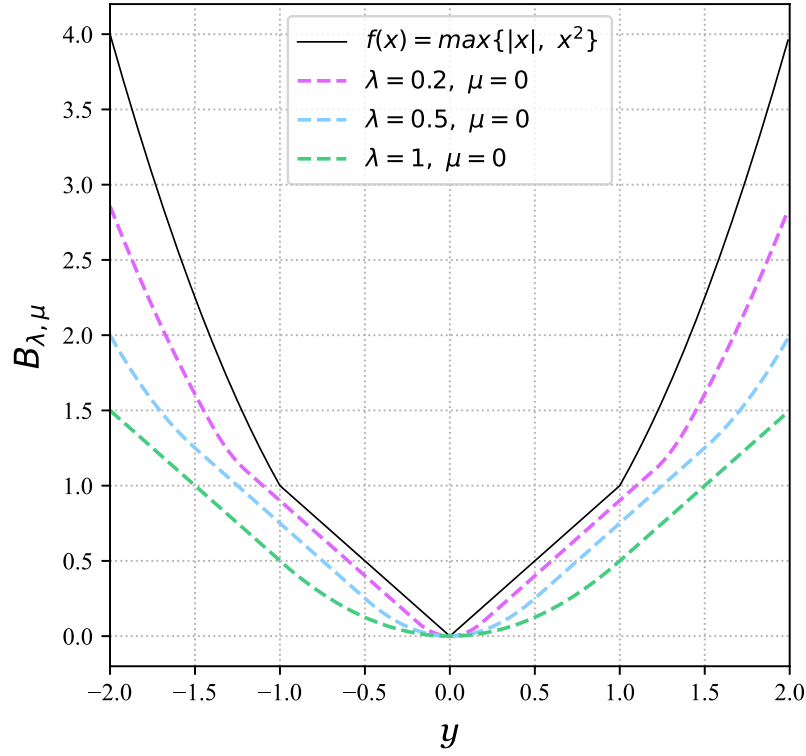


Рис. 5: Сглаживание функции $\max\{|x|, x^2\}$ преобразованием Бертсекаса: фиксирование μ , варьирование λ .

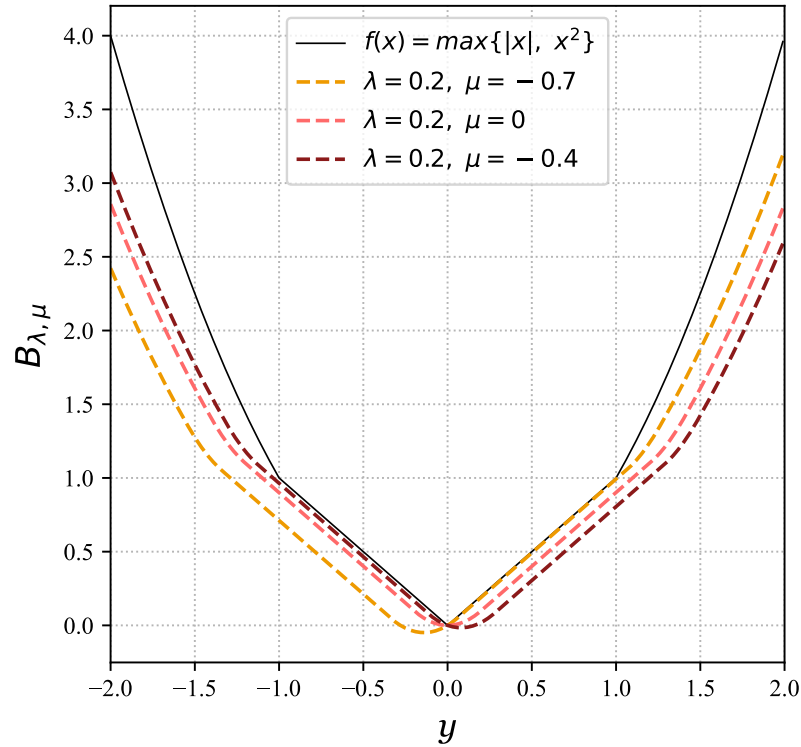


Рис. 6: Сглаживание функции $\max\{|x|, x^2\}$ преобразованием Бертсекаса: фиксирование λ , варьирование μ .

Замечание 3. Из всех трех примеров рассмотрения процедуры сглаживания функций видно, что параметр μ действительно выполняет роль сдвига графика и что при стремлении параметра λ к нулю аппроксимация получается более точной и приближенной к $f(x)$.

3 Случай в \mathbb{R}^2

В качестве двумерной модели будем рассматривать «payoff» экзотического «call» опциона³ — опциона «rainbow» (европейского типа). Он, без всяких сомнений, является одним из наиболее важных инструментов для хеджирования рисков, связанных с неопределенностью на финансовом рынке. Радужный опцион часто ассоциируют с корреляционной торговлей, поскольку стоимость опциона чувствительна к корреляции базовых активов.

Название для данного дериватива выбрано не случайным образом. Оно взято из публикации М. Рубинштейна [21], где опцион сравнивается с радугой: его выплата зависит, по крайней мере, от двух базовых рискованных активов, каждому из которых можно поставить в соответствие свой цвет.

Теория радужных опционов берет свое начало в работе [22], в которой была сконструирована новая модель и был изучен вопрос ее ценообразования. В ней держатель опциона на обмен активами имел право в момент исполнения получить некоторый актив А в обмен на другой актив В. Платежная функция в этом случае составляла $f = (S_A(T) - S_B(T))^+$, где $S_A(T), S_B(T)$ — спот-цены активов А и В, соответственно, Т — время экспирации.

Результаты, связанные с наличием только двух активов в радужном опционе, были получены в [23]. Позже в [24] был разобран случай для произвольного количества активов. Подробное описание самого дериватива и его цен можно найти в [25]. Стоит отметить, что во всех случаях динамика цен опиралась на модель Блэка-Шоулса.

Радужный опцион позволяет владельцу получить выплату, соответствующую лучшему финансовому результату из некоторого списка активов. В таблице приведены различные подвиды дериватива:

Таблица 1. Примеры выпуклых функций выплаты rainbow опциона. [16]	
Тип опциона	Функция выплат
«Best of assets or cash»	$\max(S_1, S_2, \dots, S_n, K)$
«Call on max»	$\max(\max(S_1, S_2, \dots, S_n) - K, 0)$
«Put on min»	$\max(K - \min(S_1, S_2, \dots, S_n), 0)$
«Put 2 and Call 1»	$\max(S_1 - S_2, 0)$
«Multi-strike»	$\max(S_1 - K_1, S_2 - K_2, \dots, S_n - K_n, 0)$
«Pyramid»	$\max(S_1 - K_1 + S_2 - K_2 + \dots + S_n - K_n - K, 0)$
«Madonna»	$\max\left(\sqrt{(S_1 - K_1)^2 + \dots + (S_n - K_n)^2} - K, 0\right)$

³Контракта, дающего право его покупателю купить базовую ценную бумагу в будущем в момент времени Т (момент экспирации) по заранее установленной цене исполнения опциона К (страйку).

Таблица 2. Примеры опционов rainbow с невыпуклыми функциями выплаты.

Тип опциона	Функция выплат
«Call on min»	$\max(\min(S_1, S_2, \dots, S_n) - K, 0)$
«Put on max»	$\max(K - \max(S_1, S_2, \dots, S_n), 0)$

Наиболее распространенным является «лучший из активов» («best of assets rainbow») или «денежный» («cash rainbow»). В момент экспирации этого опциона определяется, какой из рискованных активов дал наибольшую доходность за период существования опциона. Именно она и будет выплачена в соответствии со сроком исполнения.

Другими, менее популярными, видами радужных опционов являются:

- «call on max», дающий выплату в виде разницы между наибольшей ценой из списка активов и фиксированным страйком; по сути, такой опцион позволяет приобрести по фиксированному страйку актив, показавший наибольшую стоимость в момент экспирации;
- «call on min», позволяющий приобрести по страйку актив, показавший наименьшую стоимость;
- «put on max», позволяющий продать по страйку актив, показавший наибольшую стоимость, или реализовать разницу между страйком и ценой такого актива;
- «put on min», дающий возможность получить разницу между страйком и активом, показавшим наименьшую стоимость;
- «put 2 and call 1» или «exchange option», который позволяет получить выплату в виде разницы цен двух активов или некоторый процент от этой разницы.

3.1 Проведение экспериментов

Мы остановимся на рассмотрении опциона «call on max», зависящего от двух рискованных активов.

Алгоритм сглаживания будет заключаться в следующем:

1. В качестве функционала J используем выражение, стоящее под инфимумом преобразования Бертсекаса;
2. Фиксируем $y_i = [y_{i1}, y_{i2}]$;
3. Для y_i и $x_j = [x_{j1}, x_{j2}]$, находим $J(x_j, y_i)$ (т.е. вычисляем значение в точке);
4. Повторяем процедуру для других x_k , $k \neq j$;
5. В итоге получаем вектор значений $J(x, y_i)$;
6. Минимизируем функционал $J(x, y_i)$ по x , таким образом, находим $\mathfrak{B}_{\lambda, \mu}[f](y_i)$;
7. Выполняем предыдущие пункты для y_m , $m \neq i$.

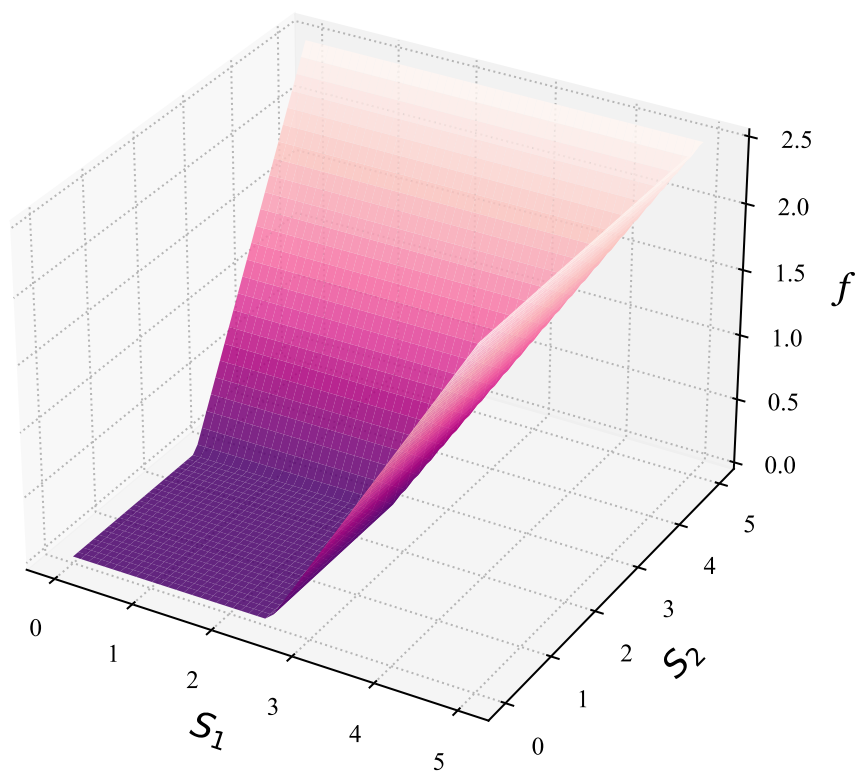


Рис. 7: Исходная функция выплат по опциону «call on max» $(\max(S_1, S_2) - K)^+$, $K = 2.5$.

Результаты сглаживающего преобразования Бертсекаса:

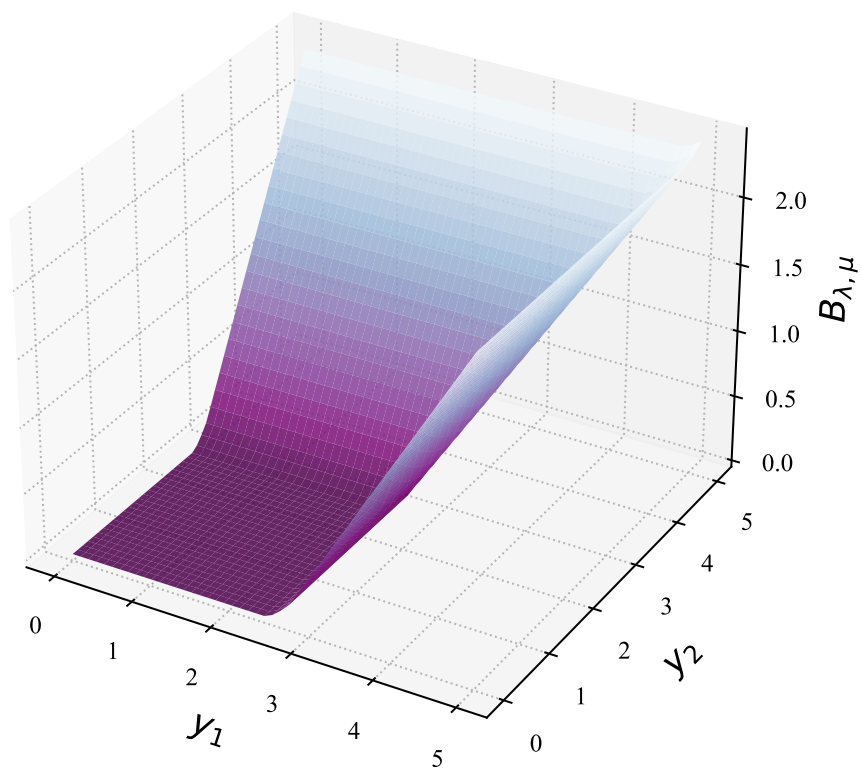


Рис. 8: Сглаживание функции выплаты опциона «call on max» $(\max(S_1, S_2) - K)^+$ преобразованием Бертсекаса: $\mu = (-0.1, -0.1)$, $\lambda = 0.3$, $K = 2.5$.

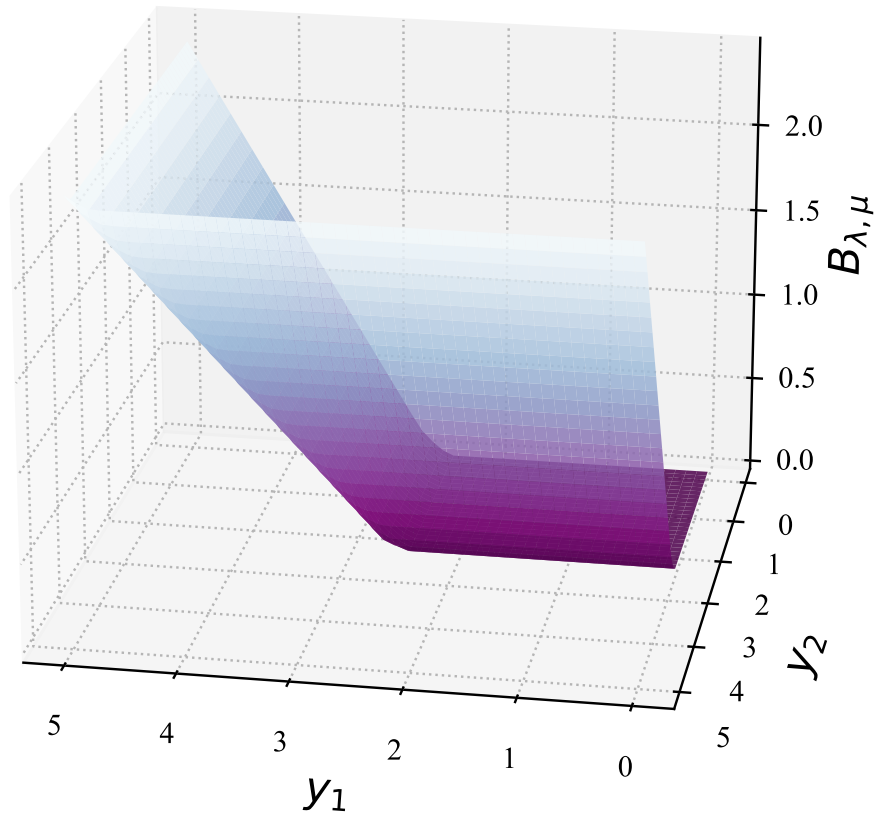


Рис. 9: Сглаживание функции выплаты опциона «call on max» $(\max(S_1, S_2) - K)^+$ преобразованием Бертсекаса: $\mu = (-0.1, -0.1)$, $\lambda = 0.3$, $K = 2.5$, поворот изображения.

3.2 Точность аппроксимации

Сглаженная функция в рассмотренном двумерном случае во всех точках не превосходит исходную. По сути она является минорантой. Построим график, на котором будет показано отклонение сглаженной функции выплат от изначально заданной (см. раздел 3.1).

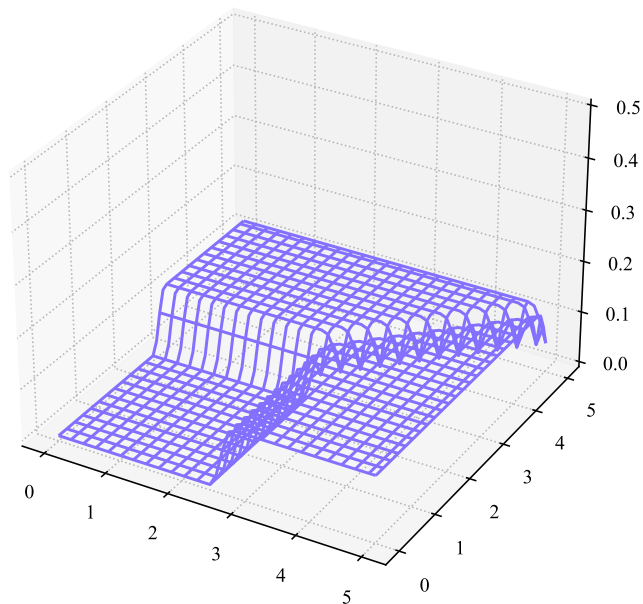


Рис. 10: Погрешность аппроксимации.

То есть практически обосновано, что разница между функцией исходной и функцией, полученной в результате сглаживающего преобразования, небольшая. При уменьшении значения λ качество приближения будет только улучшаться.

4 Заключение

В работе было рассмотрено сглаживание собственных выпуклых функций при помощи преобразований Моро, Бертсекаса в одномерном и двумерном случаях. Для первой постановки задачи было найдено аналитическое решение для каждой из рассмотренных функций-примеров. Во втором случае - сглаживание выбранной функции (выплаты раджного опциона «call on max») осуществилось численно. Получена взаимосвязь между значениями параметров и поведением сглаженной функции. Подсчитана погрешность вычислений в двумерной задаче.

В дальнейшем планируется изучение применения преобразования Бертсекаса в многомерном случае для решения задачи ценообразования и суперхеджирования⁴ обусловленных обязательств по опциону. Подробное описание задачи можно найти, обратившись к серии работ [17–20].

Причина использования сглаживания заключается не только в регуляризации численного метода, но и в наличии модельного риска. Любая модель лишь приблизительно описывает реальность — что-то описывает хорошо, что-то не очень, а что-то вообще не учитывает. В модели рассматриваемой задачи не учитываются транзакционные издержки, которые в действительности могут иметь существенное значение.

Сглаживающие преобразования в этом случае позволят обеспечить единственность и устойчивость расчета хеджирующей стратегии.

Также было бы полезно построить «хороший», с практической точки зрения, алгоритм, позволяющий решать типичные задачи суперхеджирования с торговыми ограничениями.

Для одномерного случая, в котором рассматривается только один рисковый актив, уже разработан алгоритм решения, основанный на двойном применении преобразования Фенхеля. Обзор имеющихся результатов можно найти в работах [26] и [27]. Хотелось бы обобщить полученные разработки и получить алгоритм для многомерного случая.

⁴Задача суперхеджирования (суперрепликации) заключается в таком хеджировании обусловленного обязательства по проданному опциону, т. е. проведении операций на соответствующем финансовом рынке, допустимых с учетом торговых ограничений, которое позволяет полностью покрыть обязательства (в отличие от других видов хеджирования) при минимальном начальном капитале (премии, взимаемой с покупателя опциона). Выплаты по опциону определяются на основании эволюции цен определенных торгуемых активов, фьючерсов или рыночных индексов, а сами моменты выплат и их величина определяются типом контракта.

Список литературы

- [1] *Чебышев П. Л.* О функциях, наименее уклоняющихся от нуля. Избранные труды. М.: Изд-во АН СССР, с. 579–608, 1955.
- [2] *Арутюнов А. В.* Лекции по выпуклому и многозначному анализу. М.: ФИЗМАТЛИТ, 2014.
- [3] *R. Tyrrell Rockafellar* Convex Analysis, Princeton, N.J., Princeton University Press, 1970.
- [4] *Jun Liu, Jieping Ye* Moreau-Yosida Regularization for Grouped Tree Structure Learning, Conference: Advances in Neural Information Processing Systems, 2010.
- [5] *Axel Böhm, Stephen J. Wright* Variable Smoothing for Weakly Convex Composite Functions, Journal of Optimization Theory and Applications vol. 188, pp. 628–649, 2021.
- [6] *Térence Bayen, Alain Rapaport* About Moreau-Yosida regularization of the minimal time crisis problem. Journal of Convex Analysis, vol. 23, pp.263-290, 2016.
- [7] *Geiersbach C., Hintermüller M.* Optimality Conditions and Moreau–Yosida Regularization for Almost Sure State Constraints, ESAIM: Control, Optimisation and Calculus of Variations, 2022.
- [8] *Emanuel Laude, Tao Wu, Daniel Cremers* A Nonconvex Proximal Splitting Algorithm under Moreau-Yosida Regularization, Technical University of Munich, Germany, 2017.
- [9] *Hintermüller M., Kunisch K.* Feasible and Noninterior Path-Following in Constrained Minimization with Low Multiplier Regularity, SIAM Journal on Control and Optimization, Vol. 45, No. 4, pp. 1198–1221, 2006.
- [10] *Jourani A., Thibault L., Zagrodny D.* Differential properties of the Moreau envelope, Journal of Functional Analysis, vol. 266, no. 3, pp. 1185–1237, 2014.
- [11] *Moreau J. J.* Proximité et dualité dans un espace hilbertien. Bulletin de la société mathématique de France, 93, pp.273-299, 1965.
- [12] *Bertsekas D.* Nondifferentiable optimization via approximation, Mathematical Programming Study, vol. 3, pp. 1-25, 1975.
- [13] *Bertsekas D.* Approximation Procedures Based on the Method of Multipliers, J.Optimization Theory and Applications, vol. 23, pp. 487-510, 1977.
- [14] *Parikh N. and Boyd S.* Proximal Algorithms, Foundations and Trends in Optimization, vol. 1, issue 3, pp. 135-159, 2014.

- [15] *Advani M. and Ganguli S.* Statistical mechanics of high dimensional inference, supplementary material, pp. 34-35, 2016.
- [16] *Smirnov S. N.* A Guaranteed Deterministic Approach to Superhedging—The Case of Convex Payoff, MDPI, 2019.
- [17] *Smirnov S. N.* Guaranteed deterministic approach to superhedging: market model, trading constraints and Bellman-Isaacs equations, Mathematical Games Theory and its Applications, 10:4, 59–99 (in Russian), 2018.
- [18] *Smirnov S. N.* Guaranteed deterministic approach to superhedging: «no arbitrage» properties of the market, Mathematical Games Theory and its Applications, 11:2, 68–95 (in Russian), 2019.
- [19] *Smirnov S. N.*, Guaranteed deterministic approach to superhedging: properties of semicontinuity and continuity of solutions of Bellman-Isaacs equations, Mathematical Games Theory and its Applications, 11:4, 87–115 (in Russian), 2019.
- [20] *Andreev N. A., Smirnov S. N.* Guaranteed Deterministic Approach to Superhedging: A Numerical Experiment, Computational Mathematics and Modeling, vol. 32, No. 1, 2021.
- [21] *Rubinstein M.* Somewhere over the rainbow. Risk 4.11, pp. 61-63, 1991.
- [22] *Margrabe W.* The Value of an Option to Exchange One Asset for Another. Journal of Finance, 33, pp. 177-186, 1978.
- [23] *Stulz R. M.* Options on the minimum or the maximum of two risky assets, Journal of Financial Economics, No. 1, pp. 161–185, 1982.
- [24] *Johnson H.* Options on the maximum or the minimum of several assets, Journal of Financial and Quantitative Analysis, vol. 22, is. 03, pp. 277-283, 1987.
- [25] *Ouwehand P., Graeme W.* Pricing Rainbow Options, Willmott magazine, pp. 74-80, 2006.
- [26] *Contento L., Ern A., Vermiglio R.* A linear-time approximate convex envelope algorithm using the double Legendre–Fenchel transform with application to phase separation, Computational Optimization and Applications, Springer Verlag, 2015.
- [27] *Contento L.* The Discrete Legendre-Fenchel Transform and its application to phase separation in electrolytes, University of Udine, 2012.