Методы оптимизации
Семинар 3. Выпуклые множества. Выпуклые функции

Лобанов Александр Владимирович

Московский физико-технический институт
Факультет инноваций и высоких технологий

lobanov.av@mipt.ru

15 сентября 2022 г.

Что нового?

На github появился файл с домашними заданиями:

Методы Оптимизации
МФТИ, ФИВТ
Осень 2022

Домашнее задание № 1

Матрично - векторное дифференцирование

deadline: 23:59 (Московское время), 22 сентября.

1. Найти $\nabla f(x)$ и $f''(x)$, если $f(x) = \frac{1}{2}\|Ax - b\|_2^2$, $x \in \mathbb{R}^n$.
2. Найти $\nabla f(x)$ и $f''(x)$, если $f(x) = \frac{1}{p}\|x\|_2^p$, $x \in \mathbb{R}^n \setminus \{0\}$, p — порядковый номер по списку группы (см. табл. успеваемости).
3. Найти $df(X)$ и $\nabla f(X)$, если $f(X) = \|AX - B\|_F$, $X \in \mathbb{R}^{k \times n}$.
4. Найти $df(X)$ и $\nabla f(X)$, если $f(X) = \mathrm{Tr}(AXBX^{-1})$, $X \in \mathbb{R}^{n \times n}$, $\det(X) \neq 0$.
5. Найти аналитическое выражение градиента, гессиана и сравнить с ответами, полученными любой системой автоматической дифференциации (autograd / jax / pytorch / tensorflow) для следующих функций:

1) $\boxed{f(x) = \frac{1}{2}x^\mathrm{T}Ax + b^\mathrm{T}x + c}$ 2) $\boxed{f(x) = \frac{1}{2}\|Ax - b\|_2^2}$ 3) $\boxed{f(x) = \ln(1 + \exp(ax))}$.

Формат сдачи: Файл должен быть отправлен на почту lobanov.av@mipt.ru в формате $\boxed{.pdf}$, созданном через LaTeX или через вариант печати «Сохранить как PDF» из блокнота colab \ jupyter notebook.

Аффинное множество

Аффинное множество

Множество S называется аффинным, если для любых x_1, x_2 из S прямая, проходящая через них, также лежит в S, т.е.

$$\forall \theta \in \mathbb{R},\ \forall x_1, x_2 \in S;\ \theta x_1 + (1-\theta)x_2 \in S$$

Аффинная комбинация

Пусть мы имеем $x_1, x_2, \cdots, x_k \in S$, тогда точка $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ называется аффинной комбинацией $x_1, x_2, \cdots, x_k$, если $\sum_{i=1}^{k} \theta_i = 1$.

Аффинная оболочка

Множество всех аффинных комбинаций точек множества S называется аффинной оболочкой множества S:

$$\mathbf{aff}(S) = \left\{ \sum_{i=1}^{k} \theta_i x_i \mid x_i \in S, \sum_{i=1}^{k} \theta_i = 1 \right\}$$

Аффинное множество

Аффинное множество

Множество S называется аффинным, если для любых x_1, x_2 из S прямая, проходящая через них, также лежит в S, т.е.

$$\forall \theta \in \mathbb{R},\ \forall x_1, x_2 \in S;\ \theta x_1 + (1-\theta)x_2 \in S$$

Аффинная комбинация

Пусть мы имеем $x_1, x_2, \cdots, x_k \in S$, тогда точка $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ называется аффинной комбинацией $x_1, x_2, \cdots, x_k$, если $\sum_{i=1}^{k} \theta_i = 1$.

Аффинная оболочка

Множество всех аффинных комбинаций точек множества S называется аффинной оболочкой множества S:

$$\textbf{aff}(S) = \left\{ \sum_{i=1}^{k} \theta_i x_i \mid x_i \in S, \sum_{i=1}^{k} \theta_i = 1 \right\}$$

Аффинное множество

Множество S называется аффинным, если для любых x_1, x_2 из S прямая, проходящая через них, также лежит в S, т.е.

$$\forall \theta \in \mathbb{R},\ \forall x_1, x_2 \in S;\ \theta x_1 + (1-\theta) x_2 \in S$$

Аффинная комбинация

Пусть мы имеем $x_1, x_2, \cdots, x_k \in S$, тогда точка $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ называется аффинной комбинацией $x_1, x_2, \cdots, x_k$, если $\sum_{i=1}^k \theta_i = 1$.

Аффинная оболочка

Множество всех аффинных комбинаций точек множества S называется аффинной оболочкой множества S:

$$\textbf{aff}(S) = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in S, \sum_{i=1}^k \theta_i = 1 \right\}$$

Аффинное множество

Аффинное множество

Множество S называется аффинным, если для любых x_1, x_2 из S прямая, проходящая через них, также лежит в S, т.е.

$$\forall \theta \in \mathbb{R},\ \forall x_1, x_2 \in S;\ \theta x_1 + (1-\theta) x_2 \in S$$

Аффинная комбинация

Пусть мы имеем $x_1, x_2, \cdots, x_k \in S$, тогда точка $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ называется аффинной комбинацией $x_1, x_2, \cdots, x_k$, если $\sum_{i=1}^{k} \theta_i = 1$.

Аффинная оболочка

Множество всех аффинных комбинаций точек множества S называется аффинной оболочкой множества S:

$$\textbf{aff}(S) = \left\{ \sum_{i=1}^{k} \theta_i x_i \mid x_i \in S, \sum_{i=1}^{k} \theta_i = 1 \right\}$$

Выпуклое множество

Выпуклое множество

Выпуклое множество

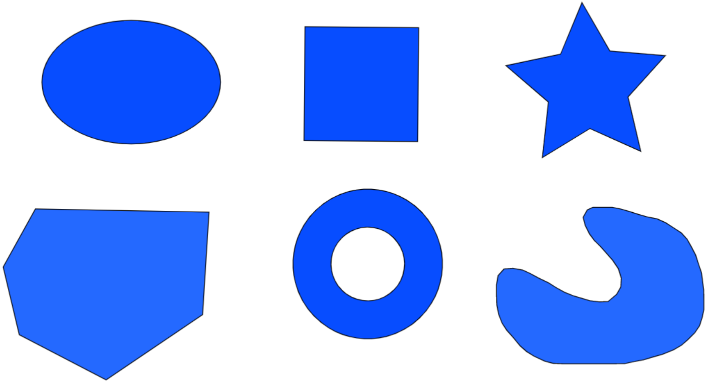
Множество S называется выпуклым, если для любых x_1, x_2 из S отрезок между ними также лежит в S, т.е.

$$\forall \theta \in [0,1],\ \forall x_1, x_2 \in S;\ \theta x_1 + (1-\theta)x_2 \in S$$

Выпуклое множество

Множество S называется выпуклым, если для любых x_1, x_2 из S отрезок между ними также лежит в S, т.е.

$$\forall \theta \in [0, 1], \ \forall x_1, x_2 \in S; \ \theta x_1 + (1 - \theta) x_2 \in S$$



Выпуклое множество

Выпуклая комбинация

Пусть мы имеем $x_1, x_2, \cdots, x_k \in S$, тогда точка $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ называется выпуклой комбинацией $x_1, x_2, \cdots, x_k$, если $\sum_{i=1}^{k} \theta_i = 1, \theta_i \geq 0$.

Выпуклое множество

Выпуклая комбинация

Пусть мы имеем $x_1, x_2, \cdots, x_k \in S$, тогда точка $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ называется выпуклой комбинацией $x_1, x_2, \cdots, x_k$, если $\sum_{i=1}^k \theta_i = 1, \theta_i \geq 0$.

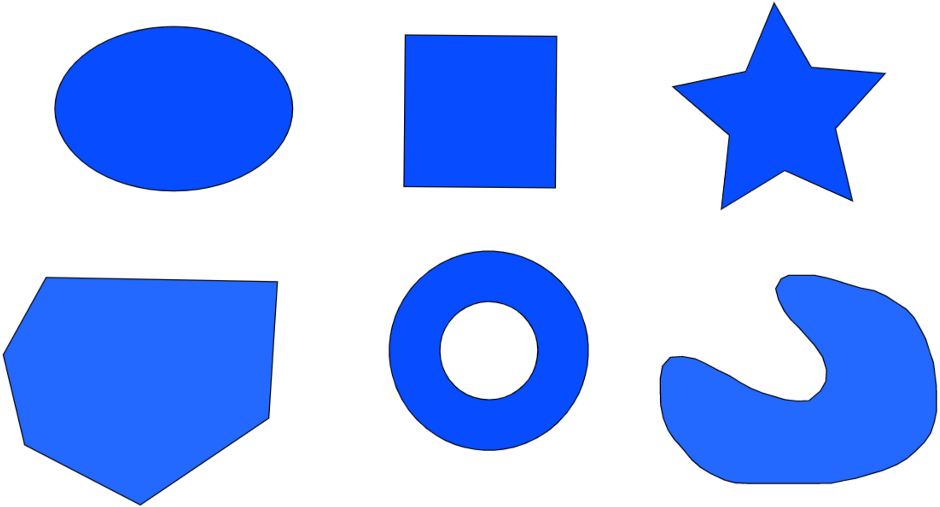
Выпуклая оболочка

Множество всех выпуклых комбинаций точек множества S называется выпуклой оболочкой множества S:

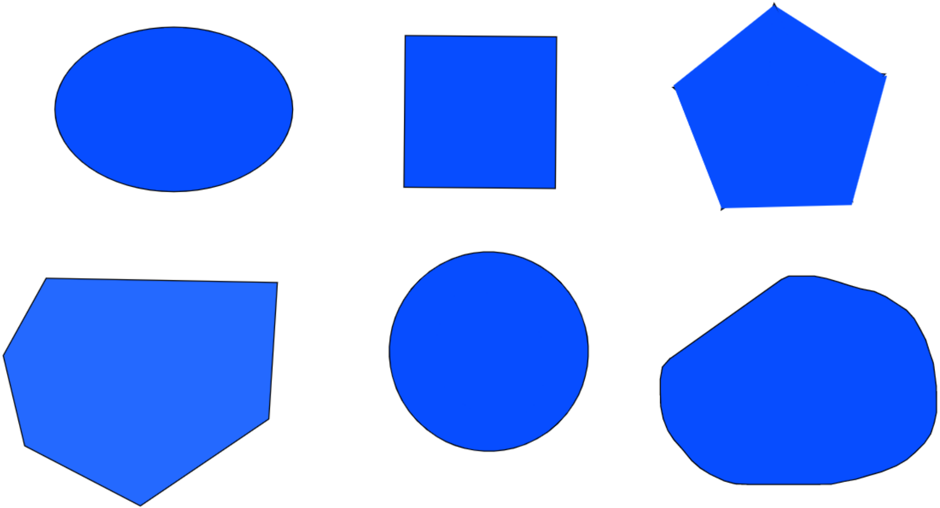
$$\textbf{conv}(S) = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in S, \sum_{i=1}^k \theta_i = 1 \right\}$$

- Множество $\textbf{conv}(S)$ – это наименьшее выпуклое множество, содержащее S.
- Множество S выпукло тогда и только тогда, когда $S = \textbf{conv}(S)$.

Выпуклое множество



Выпуклое множество



Как доказать, что множество выпукло?

По определению

$$\forall x_1, x_2 \in S,\ 0 \leq \theta \leq 1 \to \theta x_1 + (1-\theta)x_2 \in S$$

Как доказать, что множество выпукло?

По определению

$$\forall x_1, x_2 \in S,\ 0 \leq \theta \leq 1 \rightarrow \theta x_1 + (1-\theta)x_2 \in S$$

Через операции, сохраняющие выпуклость:

- Линейная комбинация выпуклых множеств — выпуклое множество:
  Пусть есть 2 выпуклых множества S_x, S_y, тогда множество S – выпукло:

$$S = \{s | s = c_1 x + c_2 y, x \in S_x, y \in S_y, c_1, c_2 \in \mathbb{R}\}$$

- Пересечение любого числа выпуклых множеств — выпуклое множество;
- Образ аффинного отображения выпуклого множества - выпуклое множество

Примеры

1. Показать, что множество выпукло тогда и только тогда, когда его пересечение с любой прямой выпукло.

Примеры

1. Показать, что множество выпукло тогда и только тогда, когда его пересечение с любой прямой выпукло.
2. Доказать, что шар в $\mathbb{R}^n$ (т.е. следующее множество $S = \{x \mid \|x - x_c\| \leq r\}$) – выпуклое множество.

Выпуклая функция

Функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, называется выпуклой на S, если

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

для всех $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$.

Выпуклая функция

Функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, называется выпуклой на S, если

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

для всех $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$.

Примеры выпуклых функций

- $f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c,\ x \in \mathbb{R}^n, A \in \mathbb{S}^n$
- $f(x) = x^p,\ p > 1, x \in \mathbb{R}_+$
- $f(x) = \|x\|^p,\ p > 1, x \in \mathbb{R}$
- $f(x) = e^{cx},\ c \in \mathbb{R}, x \in \mathbb{R}$
- $f(x) = -\ln x,\ x \in \mathbb{R}_{++}$
- $f(x) = x\ln x,\ x \in \mathbb{R}_{++}$

Как определить выпуклость функций?

Надграфик (Epigraph)

Для функции $f(x)$, определенной на $S \subseteq \mathbb{R}^n$, следующее множество:

$$\textbf{epi}\ f = \{[x, \mu] \in S \times \mathbb{R} : f(x) \leq \mu\} \subset \mathbb{R}^{n+1}$$

называется надграфиком (epigraph) функции $f(x)$.

Как определить выпуклость функций?

Надграфик (Epigraph)

Для функции $f(x)$, определенной на $S \subseteq \mathbb{R}^n$, следующее множество:

$$\textbf{epi } f = \{[x, \mu] \in S \times \mathbb{R} : f(x) \le \mu\} \subset \mathbb{R}^{n+1}$$

называется надграфиком (epigraph) функции $f(x)$.

Множество подуровней (множество Лебега)

Для функции $f(x)$, определенной на $S \subseteq \mathbb{R}^n$, следующее множество:

$$\mathcal{L}_\beta = \{x \in S : f(x) \le \beta\}$$

называется множеством подуровней функции $f(x)$.

Дифференциальный критерий первого порядка

Дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x)$$

Критерии выпуклости

Дифференциальный критерий первого порядка

Дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x)$$

Дифференциальный критерий второго порядка

Дважды дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является выпуклой тогда и только тогда, когда $\forall x \in \textbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq 0$$

Критерии выпуклости

Дифференциальный критерий первого порядка

Дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x)$$

Дифференциальный критерий второго порядка

Дважды дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является выпуклой тогда и только тогда, когда $\forall x \in \textbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq 0$$

Связь с надграфиком. Функция является выпуклой тогда и только тогда, когда ее надграфик является выпуклым множеством.

Критерии выпуклости

Дифференциальный критерий первого порядка

Дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x)$$

Дифференциальный критерий второго порядка

Дважды дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является выпуклой тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq 0$$

Связь с надграфиком. Функция является выпуклой тогда и только тогда, когда ее надграфик является выпуклым множеством.

Связь со множеством подуровней. Если $f(x)$ является выпуклой функцией, определенной на выпуклом множестве $S \subseteq \mathbb{R}^n$, тогда для любого β множество подуровней $\mathcal{L}$ является выпуклом.

Сильно выпуклая функция

Функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, называется μ-сильно выпуклой на S, если:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) - \mu\lambda(1-\lambda)\|x_1 - x_2\|^2$$

для всех $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$ для некоторой $\mu > 0$.

Сильно выпуклая функция

Сильно выпуклая функция

Функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, называется μ-сильно выпуклой на S, если:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) - \mu\lambda(1-\lambda)\|x_1 - x_2\|^2$$

для всех $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$ для некоторой $\mu > 0$.

Дифференциальный критерий первого порядка

Дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y-x) + \frac{\mu}{2}\|y-x\|^2$$

Сильно выпуклая функция

Сильно выпуклая функция

Функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, называется μ-сильно выпуклой на S, если:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) - \mu\lambda(1-\lambda)\|x_1 - x_2\|^2$$

для всех $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$ для некоторой $\mu > 0$.

Дифференциальный критерий первого порядка

Дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y-x) + \frac{\mu}{2}\|y-x\|^2$$

Дифференциальный критерий второго порядка

Дважды дифференцируемая функция $f(x)$, определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$, является μ-сильно выпуклой тогда и только тогда, когда $\forall x \in \textbf{int}(S) \neq \emptyset$: $\boxed{\nabla^2 f(x) \succeq \mu I}$

Итересные факты

- $f(x)$ называется (строго) вогнутой, если функция $-f(x)$ является (строго) выпуклой.

Итересные факты

- $f(x)$ называется (строго) вогнутой, если функция $-f(x)$ является (строго) выпуклой.
- Неравенство Йенсена для выпуклой функции:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

для $\alpha_i \geq 0$; $\sum_{i=1}^{n} \alpha_i = 1$.
Для случая бесконечной размерности:

$$f\left(\int_S x p(x) dx\right) \leq \int_S f(x) p(x) dx$$

если интеграл существует и $p(x) \geq 0$, $\int_S p(x) dx = 1$.

Итересные факты

- $f(x)$ называется (строго) вогнутой, если функция $-f(x)$ является (строго) выпуклой.
- Неравенство Йенсена для выпуклой функции:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

  для $\alpha_i \geq 0$; $\sum_{i=1}^{n} \alpha_i = 1$.
  Для случая бесконечной размерности:

$$f\left(\int_S x p(x) dx\right) \leq \int_S f(x) p(x) dx$$

  если интеграл существует и $p(x) \geq 0$, $\int_S p(x) dx = 1$.
- Если функция $f(x)$ и множество S выпуклы, то любой локальный минимум $x^* = \underset{x \in S}{\operatorname{argmin}} f(x)$ будет глобальным. Сильная выпуклость гарантирует единственность решения.

Примеры

1. Показать, что $f(x) = \|x\|$ является выпуклой функцией в $\mathbb{R}^n$.

Примеры

1. Показать, что $f(x) = \|x\|$ является выпуклой функцией в $\mathbb{R}^n$.
2. Показать, что $f(x) = c^T x + b$ является выпулой и вогнутой функцией.

Примеры

1. Показать, что $f(x) = \|x\|$ является выпуклой функцией в $\mathbb{R}^n$.
2. Показать, что $f(x) = c^T x + b$ является выпулой и вогнутой функцией.
3. Показать, что $f(x) = x^T A x$, где $A \succeq 0$ – выпукла в $\mathbb{R}^n$.

deadline: 23:59 (Московское время), 29 сентября.

1. Проверить на выпуклость множество $\{x \in \mathbb{R}^n \mid \alpha \leq a^T x \leq \beta\}$.
2. Доказать, что множество квадратных симметричных положительно определенных матриц выпукло.
3. Показать, что гиперболическое множество $\left\{x \in \mathbb{R}^n_+ \mid \prod\limits_{i=1}^{n} x_i \geq 1\right\}$ выпукло. *Подсказка:* если $a, b \geq 1$ и $0 \leq \theta \leq 1$, тогда $a^\theta b^{1-\theta} \leq \theta a + (1-\theta)b$.

Домашнее задание

<u>**deadline**: 23:59 (Московское время), 29 сентября.</u>

1. Проверить на выпуклость множество $\left\{x \in \mathbb{R}^n \mid \alpha \leq a^T x \leq \beta\right\}$.
2. Доказать, что множество квадратных симметричных положительно определенных матриц выпукло.
3. Показать, что гиперболическое множество $\left\{x \in \mathbb{R}^n_+ \mid \prod\limits_{i=1}^n x_i \geq 1\right\}$ выпукло. *Подсказка:* если $a, b \geq 1$ и $0 \leq \theta \leq 1$, тогда $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$.
4. Показать, что $f(x)$ является выпуклой функцией, используя критерии первого и второго порядка, если $f(x) = \sum_{i=1}^n x_i^4$.
5. Найти значения a, b, c, где $f(x, y, z) = x^2 + 2axy + by^2 + cz^2$ является выпуклой, строго выпуклой и сильно выпуклой.
6. Доказать, что добавление $\lambda \|x\|_2^2$ к любой выпуклой функции $g(x)$ обеспечивает сильную выпуклость результирующей функции $f(x) = g(x) + \lambda \|x\|_2^2$. Найдите константу сильной выпулости μ.