

Метод выпуклой оптимизации на квадрате

И. Курузов¹ Ф. Стонякин^{1, 2}

¹Московский Физико-Технический Институт

²Крымский Федеральный Университет имени В.И. Вернадского

62-ая конференция МФТИ, 2019

Задача:

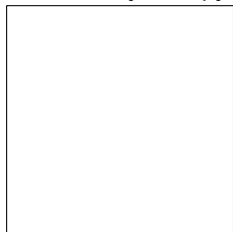
$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.

Задача:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

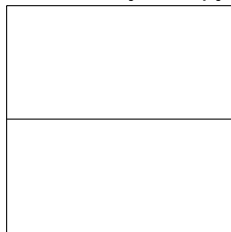
где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.



Задача:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

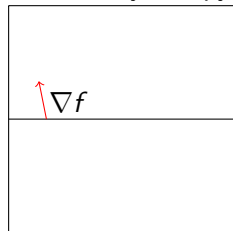
где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.



Задача:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.

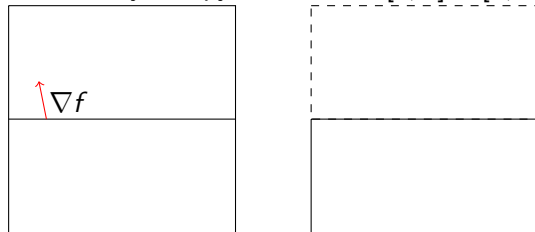


Описание метода

Задача:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.

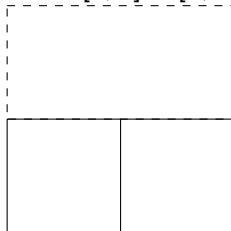
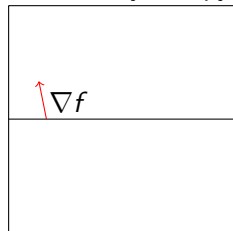


Описание метода

Задача:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.

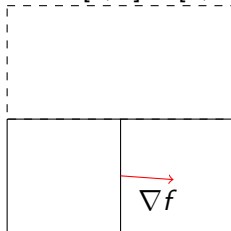
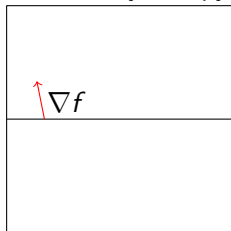


Описание метода

Задача:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

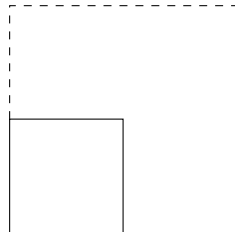
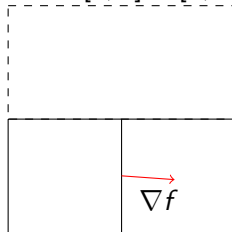
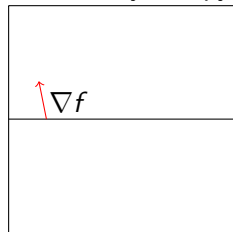
где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.



Задача:

$$\min_{(x,y)} \{f(x,y) | (x,y) \in Q\},$$

где f выпуклая функция, $Q = [a, b] \times [c, d] \in \mathbb{R}^2$.



- 1 Стратегии для вспомогательных задач
- 2 Сходимость
- 3 Двойственные задачи
- 4 Эксперименты
- 5 Обобщение

Пусть f - L -липшецева с M -липшецевым градиентом.

Доказано, что следующей точности для одномерной задачи достаточно для достижения точности ϵ по функции исходной задачи:

$$\delta \leq \frac{\epsilon}{2Ma(\sqrt{2} + \sqrt{5})(1 - \frac{\epsilon}{La\sqrt{2}})}$$

Пусть f - L -липшецева с M -липшецевым градиентом.

Доказано, что следующей точности для одномерной задачи достаточно для достижения точности ϵ по функции исходной задачи:

$$\delta \leq \frac{\epsilon}{2Ma(\sqrt{2} + \sqrt{5})(1 - \frac{\epsilon}{La\sqrt{2}})}$$

Сходимость по аргументу не гарантируется!

Хотим: выбрать прямоугольник, содержащий решение исходной задачи

Хотим: выбрать прямоугольник, содержащий решение исходной задачи
Достаточно условия на знак:

$$\text{sign } f'_y(x_0) = \text{sign } f'_y(x_{\text{current}})$$

Хотим: выбрать прямоугольник, содержащий решение исходной задачи
Достаточно условия на знак:

$$\text{sign } f'_y(x_0) = \text{sign } f'_y(x_{\text{current}})$$

Усиленное условие

$$|f'_y(x_0) - f'_y(x_{\text{current}})| \leq |f'_y(x_{\text{current}})|$$

Хотим: выбрать прямоугольник, содержащий решение исходной задачи
Достаточно условия на знак:

$$\text{sign } f'_y(x_0) = \text{sign } f'_y(x_{\text{current}})$$

Усиленное условие

$$|f'_y(x_0) - f'_y(x_{\text{current}})| \leq |f'_y(x_{\text{current}})|$$

В случае M -липшецевости градиента:

$$\delta \leq \frac{|f'_y(x_{\text{current}})|}{M}$$

Что делать, если градиент мал?

Что делать, если градиент мал?

Small Gradient

f выпуклая функция с M -липшкецевым градиентом. Тогда точка x - решение с точностью ϵ , если

$$\|\nabla f(x)\| \leq \frac{\epsilon}{a\sqrt{2}},$$

где a - размер квадрата.

Функция f выпуклая функция. Q - квадрат со стороной a .

Сходимость

Если функция f L -липшецева, то для сходимости к решению с точностью ϵ по функции достаточно следующее количество итераций:

$$N = \left\lceil \log_2 \frac{La\sqrt{2}}{\epsilon} \right\rceil. \quad (1)$$

Сходимость

Если

1. Функция f имеет M -липшецев градиент
2. $\exists \mathbf{x}^* \in Q : \nabla f(\mathbf{x}^*) = \mathbf{0}$
3. Стратегия обеспечивает сходимость по аргументу, то для сходимости к решению с точностью ϵ по функции достаточно следующее количество итераций:

$$N = \left\lceil \frac{1}{2} \log_2 \frac{Ma^2}{4\epsilon} \right\rceil. \quad (2)$$

Двойственные задачи

Задача

Задача:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \\ \text{s.t. } g_k(\mathbf{x}) \leq 0, k = \overline{1, m} \end{aligned}$$

где f μ_f -сильно выпуклая L_f -липшецева функция с M_f -липшецевым градиентом, g_k выпуклая L_{g_k} -липшецева функция с M_{g_k} -липшецевым градиентом для всех $k = \overline{1, m}$.

Двойственные задачи

Задача

Двойственная задача с точностью до знака:

$$\min_{\lambda \in \mathbb{R}_+^m} \Phi(\lambda),$$

где

$$\Phi(\lambda) = - \min_{\mathbf{x} \in \mathbb{R}^n} (f(\mathbf{x}) + \langle \lambda, g(\mathbf{x}) \rangle).$$

$$\mathbf{x}(\lambda) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} (f(\mathbf{x}) + \langle \lambda, g(\mathbf{x}) \rangle)$$

Двойственные задачи

Параметры двойственной

Теорема

Если существует $\mathbf{x}_0 \in \mathbb{R}^n : g(\mathbf{x}_0) < 0$, то

$$\|\lambda^*\|_1 \leq \frac{1}{\gamma} (f(\mathbf{x}_0) - f(\mathbf{x}^*)) = a, \text{ где } \gamma = \min_k [-g_k(\mathbf{x}_0)].$$

Двойственные задачи

Параметры двойственной

Теорема

Если существует $\mathbf{x}_0 \in \mathbb{R}^n : g(\mathbf{x}_0) < 0$, то

$$\|\lambda^*\|_1 \leq \frac{1}{\gamma} (f(\mathbf{x}_0) - f(\mathbf{x}^*)) = a, \text{ где } \gamma = \min_k [-g_k(\mathbf{x}_0)].$$

$$\min_{\lambda \in \mathbb{R}_+^m} \Phi(\lambda) = \min_{\lambda \in Q} \Phi(\lambda),$$

где $Q = [0, a]^m$.

Двойственные задачи

Параметры двойственной

$$\Phi(\lambda) = - \min_{\mathbf{x} \in \mathbb{R}^n} (f(\mathbf{x}) + \langle \lambda, g(\mathbf{x}) \rangle)$$

Градиент (теорема Демьянов-Данскин-Рубинов):

$$\nabla \Phi(\lambda) = g(\mathbf{x}(\lambda))$$

Константа Липшица для функции:

$$L = \max \|g(\mathbf{x})\|$$

Константа Липшица для градиента:

$$M = \frac{L_g^2}{\mu_f}$$

Двойственные задачи

Вычисление $x(\lambda)$

Проблемы:

1. Как сделать итерацию во вспомогательной задаче?
2. Как проверить условие остановки для вспомогательной задачи
 $\delta \leq \frac{|\Phi'_2(\lambda)|}{L}$?
3. Как выбрать прямоугольник?

Двойственные задачи

Вычисление $\mathbf{x}(\lambda)$

Нас интересует только знак:

$$1. \Phi'_1(\lambda) = g_1(\mathbf{x}(\lambda))$$

$$2. \delta - \frac{|\Phi'_2(\lambda)|}{M} = \delta - \frac{|g_2(\mathbf{x}(\lambda))|}{M}$$

$$3. \Phi'_2(\lambda) = g_2(\mathbf{x}(\lambda))$$

Двойственные задачи

Вычисление $\mathbf{x}(\lambda)$

Нас интересует только знак:

1. $\Phi'_1(\lambda) = g_1(\mathbf{x}(\lambda))$
2. $\delta - \frac{|\Phi'_2(\lambda)|}{M} = \delta - \frac{|g_2(\mathbf{x}(\lambda))|}{M}$
3. $\Phi'_2(\lambda) = g_2(\mathbf{x}(\lambda))$

$$\forall a, b \neq 0 \quad |a - b| \leq |b| \rightarrow \text{sign } a = \text{sign } b$$

Двойственные задачи

Вычисление $x(\lambda)$

Нас интересует только знак:

1. $\Phi'_1(\lambda) = g_1(x(\lambda))$
2. $\delta - \frac{|\Phi'_2(\lambda)|}{M} = \delta - \frac{|g_2(x(\lambda))|}{M}$
3. $\Phi'_2(\lambda) = g_2(x(\lambda))$

$$\forall a, b \neq 0 \quad |a - b| \leq |b| \rightarrow \text{sign } a = \text{sign } b$$

Получаем следующие условия остановки вычисления $x(\lambda)$:

1. $L_{g_1} \|x - x(\lambda)\| \leq |g_1(x)|$
2. $\frac{L_{g_2}}{M} \|x - x(\lambda)\| \leq \left| \delta - \frac{|g_2(x)|}{M} \right|$
3. $L_{g_2} \|x - x(\lambda)\| \leq |g_2(x)|$

$$f(\mathbf{x}) = \log_2 \left(1 + \sum_{k=1}^n e^{\alpha x_k} \right) + \beta \|\mathbf{x}\|_2^2 \rightarrow \min_{\mathbf{x} \in \mathbb{R}^n}$$
$$\text{s.t. } g_k(\mathbf{x}) = \langle b_k, \mathbf{x} \rangle + c_k$$

Двойственная задача с точностью до знака:

$$- \min_{\mathbf{x} \in \mathbb{R}^n} (f(\mathbf{x}) + \langle \lambda, g(\mathbf{x}) \rangle) \rightarrow \min_{\lambda \in [0, a]^2}$$

1. Метод Эллипсоидов с ϵ -субградиентом

Сходимость:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \max_{\lambda \in Q} |\Phi(\lambda)| \exp\left(-\frac{k}{8}\right) + \epsilon$$

1. Метод Эллипсоидов с ϵ -субградиентом

Сходимость:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \max_{\lambda \in Q} |\Phi(\lambda)| \exp\left(-\frac{k}{8}\right) + \epsilon$$

2. Primal Gradient Method с (δ, L) -оракулом.

Сходимость:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \frac{LR^2}{2} \exp\left(-k\frac{\mu}{L}\right) + \delta$$

1. Метод Эллипсоидов с ϵ -субградиентом

Сходимость:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \max_{\lambda \in Q} |\Phi(\lambda)| \exp\left(-\frac{k}{8}\right) + \epsilon$$

2. Primal Gradient Method с (δ, L) -оракулом.

Сходимость:

$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \frac{LR^2}{2} \exp\left(-k\frac{\mu}{L}\right) + \delta$$

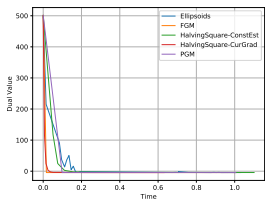
3. Fast Gradient Method с (δ, L) -оракулом.

Сходимость:

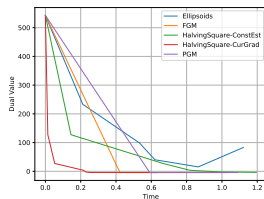
$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \min\left(\frac{4LR^2}{k^2}, LR^2 \exp\left(-\frac{k}{2}\sqrt{\frac{\mu}{L}}\right)\right) + C_k\delta,$$

Эксперименты

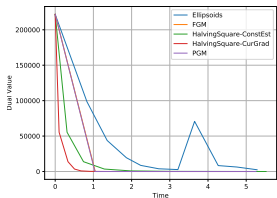
Результаты



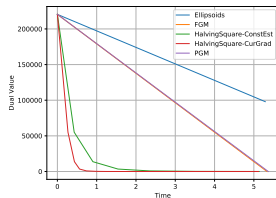
$$n = 100, \epsilon = 1e - 3$$



$$n = 100, \epsilon = 1e - 10$$



$$n = 10000, \epsilon = 1e - 3$$



$$n = 10000, \epsilon = 1e - 10$$

Обобщение на случай размерности > 2 :

Квадрат $\rightarrow n$ -мерный куб

Разделяющий отрезок $\rightarrow n - 1$ -мерный гиперкуб

Обобщение на случай размерности > 2 :

Квадрат $\rightarrow n$ -мерный куб

Разделяющий отрезок $\rightarrow n - 1$ -мерный гиперкуб

Сходимость:

$$N = \left\lceil \log_2 \frac{La\sqrt{2n}}{\epsilon} \right\rceil. \quad (3)$$

Обобщение на случай размерности > 2 :

Квадрат $\rightarrow n$ -мерный куб

Разделяющий отрезок $\rightarrow n - 1$ -мерный гиперкуб

Сходимость:

$$N = \left\lceil \log_2 \frac{La\sqrt{2n}}{\epsilon} \right\rceil. \quad (3)$$

Можно решать рекурсивно!

- Метод двумерной оптимизации
- Его приложение в решение двойственных задач
- Превосходство стратегии CurGrad
- Обобщение

Спасибо за внимание!