Метод выпуклой оптимизации на квадрате

И. Курузов[1] Ф. Стонякин[1, 2]

[1]Московский Физико-Технический Институт

[2]Крымский Федеральный Университет имени В.И. Вернандского

62-ая конференция МФТИ, 2019

Описание метода

Задача:

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

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

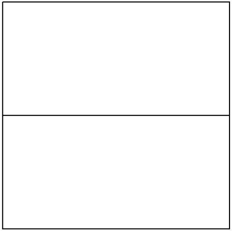
Метод был предложен Ю.Е. Нестеровым.

Описание метода

Задача:

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

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



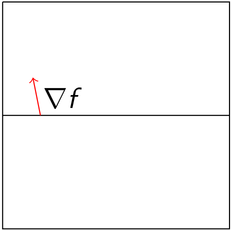
Метод был предложен Ю.Е. Нестеровым.

Описание метода

Задача:

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

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



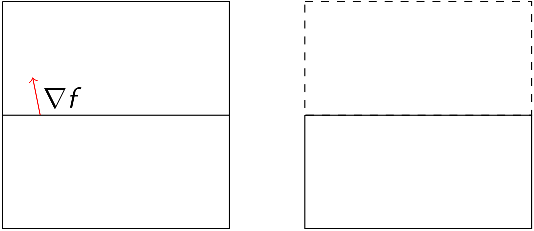
Метод был предложен Ю.Е. Нестеровым.

Описание метода

Задача:

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

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



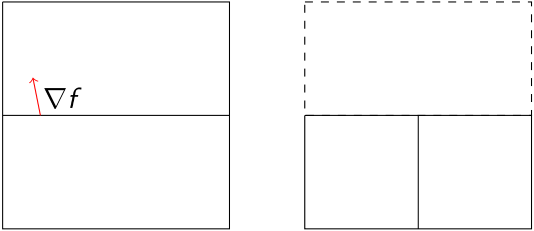
Метод был предложен Ю.Е. Нестеровым.

Описание метода

Задача:

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

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



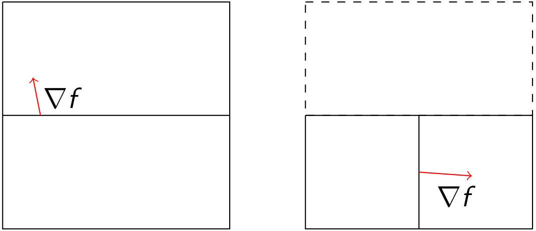
Метод был предложен Ю.Е. Нестеровым.

Описание метода

Задача:

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

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



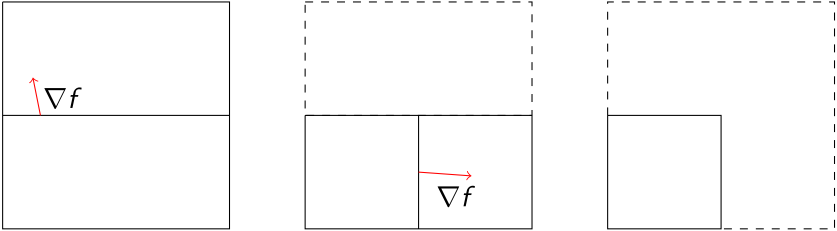
Метод был предложен Ю.Е. Нестеровым.

Описание метода

Задача:

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

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



Метод был предложен Ю.Е. Нестеровым.

План

Стратегии

Constant Estimate

Пусть f - L-липшецева с M-липшецевым градиентом.
Доказано Ф.Стонякиным и Д.Пасечнюком, что следующей точности для одномерной задачи достаточно для достижения точности ϵ по функции исходной задачи:

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

Пусть f - L-липшецева с M-липшецевым градиентом.
Доказано Ф.Стонякиным и Д.Пасечнюком, что следующей точности для одномерной задачи достаточно для достижения точности ϵ по функции исходной задачи:

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

Сходимость по аргументу не гарантируется!

Стратегии
Current Gradient

Хотим: выбрать прямоугольник, содержащий решение исходной задачи

Стратегии
Current Gradient

Хотим: выбрать прямоугольник, содержащий решение исходной задачи $x_* = \arg\min_{x \in \text{segment}} f(x)$, $x_{current}$ - текущее приближение x_* на отрезке. Достаточно условия на знак:

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

Стратегии

Current Gradient

Хотим: выбрать прямоугольник, содержащий решение исходной задачи $x_* = \arg\min_{x \in \text{segment}} f(x)$, $x_{current}$ - текущее приближение x_* на отрезке. Достаточно условия на знак:

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

Усиленное условие

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

Хотим: выбрать прямоугольник, содержащий решение исходной задачи $x_* = \arg\min_{x \in \text{segment}} f(x)$, $x_{current}$ - текущее приближение x_* на отрезке. Достаточно условия на знак:

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

Усиленное условие

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

В случае M-липшецевости градиента:

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

Что делать, если градиент мал?

Что делать, если градиент мал?

Small Gradient

f выпуклая функция с M-липшкецевым градиентом. Тогда точка $\mathbf{x}$ - решение с точностью ϵ, если

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

где a - размер квадрата.

Сходимость

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 количство итерейций:

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

Сходимость

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 количство итереций:

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

Задача:

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

$$\text{s.t. } g_k(\mathbf{x}) \le 0, k = \overline{1, m}$$

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 $k = \overline{1, m}$.

Двойственная задача с точностью до знака:

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

где

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

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

Теорема

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

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

Теорема

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

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

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

где $Q = [0, a]^m$.

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

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:

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

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. $\Phi_2'(\lambda) = g_2(\mathbf{x}(\lambda))$

Нас интересует только знак:

1. $\Phi_1'(\lambda) = g_1\left(\mathbf{x}(\lambda)\right)$
2. $\delta - \frac{|\Phi_2'(\lambda)|}{M} = \delta - \frac{|g_2(\mathbf{x}(\lambda))|}{M}$
3. $\Phi_2'(\lambda) = g_2\left(\mathbf{x}(\lambda)\right)$

$$\forall a, b \neq 0\ |a - b| \leq |b| \to \operatorname{sign} a = \operatorname{sign} b$$

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. $\Phi_2'(\lambda) = g_2(\mathbf{x}(\lambda))$

$$\forall a, b \neq 0 \; |a-b| \leq |b| \to \operatorname{sign} a = \operatorname{sign} b$$

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

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

$$f(\mathbf{x}) = \log_2\left(1 + \sum_{k=1}^{n} e^{\alpha x_k}\right) + \beta\|\mathbf{x}\|_2^2 \to \min_{\mathbf{x}\in\mathbb{R}^n}$$

$$\text{s.t. } g_k(\mathbf{x}) = \langle b_k, \mathbf{x}\rangle + c_k$$

Двойственная задача с точностью до знака:

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

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:

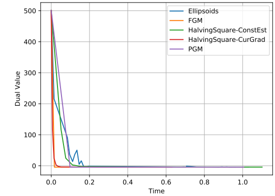
$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \frac{LR^2}{2} \exp\left(-k\frac{\mu}{L}\right) + \delta$$

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

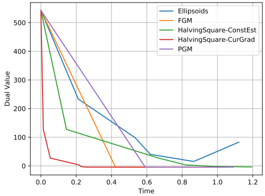
$$\min_k \Phi(\lambda_k) - \Phi(\lambda^*) \leq \min\left(\frac{4LR^2}{k^2}, LR^2 \exp\left(-\frac{k}{2}\sqrt{\frac{\mu}{L}}\right)\right) + C_k\delta,$$

Эксперименты

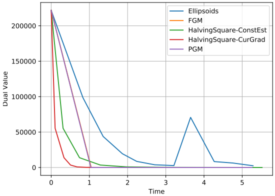
Результаты



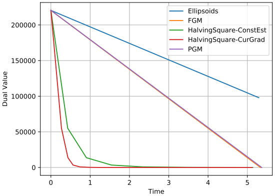
$n = 100, \epsilon = 1e-3$



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



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



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

Обобщение

Обобщение на случай размерности >2:

Квадрат $\rightarrow$ n-мерный куб

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

Обобщение

Обобщение на случай размерности >2:

$$\text{Квадрат} \to n\text{-мерный куб}$$

$$\text{Разделяющий отрезок} \to n-1\text{-мерный гиперкуб}$$

Сходимость:

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

Обобщение

Обобщение на случай размерности >2:

Квадрат $\to$ n-мерный куб

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

Сходимость:

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

Можно решать рекурсивно!

Заключение

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

Спасибо за внимание!