

Ускоренный спуск по случайному направлению для задач гладкой стохастической выпуклой оптимизации

Горбунов Эдуард

Московский Физико-Технический Институт

14 Апреля, 2017

Постановка задачи

Рассматривается следующая задача оптимизации

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \mathbb{E}_{\xi}[F(x, \xi)] = \int_{\mathcal{X}} F(x, \xi) dP(x) \right\}, \quad (1)$$

Постановка задачи

Рассматривается следующая задача оптимизации

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \mathbb{E}_{\xi}[F(x, \xi)] = \int_{\mathcal{X}} F(x, \xi) dP(x) \right\}, \quad (1)$$

где случайный вектор ξ имеет функцию распределения $P(\xi)$, $\xi \in \mathcal{X}$, и для почти всех $\xi \in \mathcal{X}$ (относительно распределения P) функция $F(x, \xi)$ является замкнутой и выпуклой.

Постановка задачи

Рассматривается следующая задача оптимизации

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \mathbb{E}_{\xi}[F(x, \xi)] = \int_{\mathcal{X}} F(x, \xi) dP(\xi) \right\}, \quad (1)$$

где случайный вектор ξ имеет функцию распределения $P(\xi)$, $\xi \in \mathcal{X}$, и для почти всех $\xi \in \mathcal{X}$ (относительно распределения P) функция $F(x, \xi)$ является замкнутой и выпуклой. Более того, мы предполагаем, что для почти всех ξ у функции $F(x, \xi)$ существует градиент $g(x, \xi)$, удовлетворяющий условию Липшица с константой $L(\xi)$ в евклидовой норме, то есть

$$\|g(x, \xi) - g(y, \xi)\|_2 \leq L(\xi) \|x - y\|_2, \quad \forall x, y \in \mathbb{R}^n, \text{ п.в. } \xi \in \mathcal{X},$$

причём будем считать, что $L_2 := \sqrt{\mathbb{E}_{\xi}[L(\xi)^2]} < +\infty$.

Постановка задачи

При сделанных предположениях $\mathbb{E}_\xi[g(x, \xi)] = \nabla f(x) \forall x \in \mathbb{R}^n$ и градиент функции $f(x)$ является липшицевым с константой L_2 в евклидовой норме.

Постановка задачи

При сделанных предположениях $\mathbb{E}_\xi[g(x, \xi)] = \nabla f(x) \forall x \in \mathbb{R}^n$ и градиент функции $f(x)$ является липшицевым с константой L_2 в евклидовой норме. Кроме того, предположим, что

$$\mathbb{E}_\xi[\|g(x, \xi) - \nabla f(x)\|_2^2] \leq \sigma^2. \quad (2)$$

Постановка задачи

Наконец, считаем, что мы имеем доступ к оракулу, который для заданной точки $x \in \mathbb{R}^n$, заданного направления $e \in S_2(1)$ с единичной евклидовой сферы в $\mathbb{R}^n$ и некоторой случайной реализации ξ выдаёт зашумлённую стохастическую аппроксимацию $\tilde{f}'(x, \xi, e)$ производной по направлению $\langle g(x, \xi), e \rangle$:

Постановка задачи

Наконец, считаем, что мы имеем доступ к оракулу, который для заданной точки $x \in \mathbb{R}^n$, заданного направления $e \in S_2(1)$ с единичной евклидовой сферы в $\mathbb{R}^n$ и некоторой случайной реализации ξ выдаёт зашумлённую стохастическую аппроксимацию $\tilde{f}'(x, \xi, e)$ производной по направлению $\langle g(x, \xi), e \rangle$:

$$\begin{aligned}\tilde{f}'(x, \xi, e) &= \langle g(x, \xi), e \rangle + \zeta(x, \xi, e) + \eta(x, \xi, e), \\ \mathbb{E}_\xi(\zeta(x, \xi, e))^2 &\leq \Delta_\zeta, \quad \forall x \in \mathbb{R}^n, \forall e \in S_2(1), \\ |\eta(x, \xi, e)| &\leq \Delta_\eta, \quad \forall x \in \mathbb{R}^n, \forall e \in S_2(1), \text{ п.в. } \xi,\end{aligned}\tag{3}$$

Постановка задачи

Наконец, считаем, что мы имеем доступ к оракулу, который для заданной точки $x \in \mathbb{R}^n$, заданного направления $e \in S_2(1)$ с единичной евклидовой сферы в $\mathbb{R}^n$ и некоторой случайной реализации ξ выдаёт зашумлённую стохастическую аппроксимацию $\tilde{f}'(x, \xi, e)$ производной по направлению $\langle g(x, \xi), e \rangle$:

$$\begin{aligned}\tilde{f}'(x, \xi, e) &= \langle g(x, \xi), e \rangle + \zeta(x, \xi, e) + \eta(x, \xi, e), \\ \mathbb{E}_\xi(\zeta(x, \xi, e))^2 &\leq \Delta_\zeta, \quad \forall x \in \mathbb{R}^n, \forall e \in S_2(1), \\ |\eta(x, \xi, e)| &\leq \Delta_\eta, \quad \forall x \in \mathbb{R}^n, \forall e \in S_2(1), \text{ п.в. } \xi,\end{aligned}\tag{3}$$

где $S_2(1)$ — единичная евклидова сфера с центром в нуле и параметры Δ_ζ , Δ_η мы можем выбирать настолько малыми, насколько потребуется.

Постановка задачи

Наконец, считаем, что мы имеем доступ к оракулу, который для заданной точки $x \in \mathbb{R}^n$, заданного направления $e \in S_2(1)$ с единичной евклидовой сферы в $\mathbb{R}^n$ и некоторой случайной реализации ξ выдаёт зашумлённую стохастическую аппроксимацию $\tilde{f}'(x, \xi, e)$ производной по направлению $\langle g(x, \xi), e \rangle$:

$$\begin{aligned}\tilde{f}'(x, \xi, e) &= \langle g(x, \xi), e \rangle + \zeta(x, \xi, e) + \eta(x, \xi, e), \\ \mathbb{E}_\xi(\zeta(x, \xi, e))^2 &\leq \Delta_\zeta, \quad \forall x \in \mathbb{R}^n, \forall e \in S_2(1), \\ |\eta(x, \xi, e)| &\leq \Delta_\eta, \quad \forall x \in \mathbb{R}^n, \forall e \in S_2(1), \text{ п.в. } \xi,\end{aligned}\tag{3}$$

где $S_2(1)$ — единичная евклидова сфера с центром в нуле и параметры $\Delta_\zeta, \Delta_\eta$ мы можем выбирать настолько малыми, насколько потребуется. Отметим, что мы используем гладкость $F(\cdot, \xi)$, когда записываем производную по направлению в виде $\langle g(x, \xi), e \rangle$, но *не предполагаем*, что у нас есть доступ к стохастическому градиенту $g(x, \xi)$ целиком.

Обозначения

Пусть $d : \mathbb{R}^n \rightarrow \mathbb{R}$ — дифференцируемая 1-сильно выпуклая по отношению к p -норме (везде далее $1 \leq p \leq 2$) функция (прокс-функция).

Обозначения

Пусть $d : \mathbb{R}^n \rightarrow \mathbb{R}$ — дифференцируемая 1-сильно выпуклая по отношению к p -норме (везде далее $1 \leq p \leq 2$) функция (*прокс-функция*). Например, для $p = 1$ можно взять

$$d(x) = \frac{en^{\frac{(\kappa-1)(2-\kappa)}{\kappa}}}{2} \|x\|_{\kappa}^2, \text{ где } \kappa = 1 + \frac{1}{\ln n}, \text{ а для } p = 2 \text{ можно выбрать}$$
$$d(x) = \frac{1}{2} \|x\|_2^2.$$

Обозначения

Пусть $d : \mathbb{R}^n \rightarrow \mathbb{R}$ — дифференцируемая 1-сильно выпуклая по отношению к p -норме (везде далее $1 \leq p \leq 2$) функция (*прокс-функция*). Например, для $p = 1$ можно взять

$d(x) = \frac{en^{\frac{(\kappa-1)(2-\kappa)}{\kappa}}}{2} \|x\|_{\kappa}^2$, где $\kappa = 1 + \frac{1}{\ln n}$, а для $p = 2$ можно выбрать $d(x) = \frac{1}{2} \|x\|_2^2$. *Дивергенцией Брегмана* по отношению к прокс-функции d будем называть следующую функцию двух аргументов:

$$V[z](x) \stackrel{\text{def}}{=} d(x) - d(z) - \langle \nabla d(z), x - z \rangle. \quad (4)$$

Обозначения

Пусть $d : \mathbb{R}^n \rightarrow \mathbb{R}$ — дифференцируемая 1-сильно выпуклая по отношению к p -норме (везде далее $1 \leq p \leq 2$) функция (прокс-функция). Например, для $p = 1$ можно взять

$d(x) = \frac{en \frac{(\kappa-1)(2-\kappa)}{\kappa}}{2} \|x\|_{\kappa}^2$, где $\kappa = 1 + \frac{1}{\ln n}$, а для $p = 2$ можно выбрать $d(x) = \frac{1}{2} \|x\|_2^2$. Дивергенцией Брегмана по отношению к прокс-функции d будем называть следующую функцию двух аргументов:

$$V[z](x) \stackrel{\text{def}}{=} d(x) - d(z) - \langle \nabla d(z), x - z \rangle. \quad (4)$$

Отметим, что из сильной выпуклости d следует

$$V[z](x) \geq \frac{1}{2} \|x - z\|_p^2, \quad x, z \in \mathbb{R}^n.$$

Ключевая лемма

Пусть e — равномерно распределённый случайный вектор на n -мерной евклидовой единичной сфере ($e \in RS_2^n(1)$).

Ключевая лемма

Пусть e — равномерно распределённый случайный вектор на n -мерной евклидовой единичной сфере ($e \in RS_2^n(1)$). Следующий факт является не только ключевым для получения оценок скорости сходимости нашего метода, но и сам по себе заслуживает отдельного внимания.

Лемма

Пусть $e \in RS_2^n(1)$, $p \in [1, 2]$ и q удовлетворяет равенству $\frac{1}{p} + \frac{1}{q} = 1$.

Тогда для $n \geq 8$ и $\rho_n = \min\{q - 1, 16 \ln n - 8\} n^{\frac{2}{q}-1}$,

$$\mathbb{E}_e \|e\|_q^2 \leq \rho_n, \quad (5)$$

$$\mathbb{E}_e [\langle s, e \rangle^2 \|e\|_q^2] \leq \frac{6\rho_n}{n} \|s\|_2^2, \quad \forall s \in \mathbb{R}^n. \quad (6)$$

Ключевая лемма

Последнее неравенство можно для $q = \infty$ записать в следующем виде (не умаляя общности, считаем, что $\|s\|_2 = 1$):

$$\mathbb{E}_e [\langle s, e \rangle^2 \|e\|_\infty^2] \lesssim \frac{1}{n} \cdot \frac{\ln n}{n} \quad \forall s \in S_2(1).$$

Ключевая лемма

Последнее неравенство можно для $q = \infty$ записать в следующем виде (не умаляя общности, считаем, что $\|s\|_2 = 1$):

$$\mathbb{E}_e [\langle s, e \rangle^2 \|e\|_\infty^2] \lesssim \frac{1}{n} \cdot \frac{\ln n}{n} \quad \forall s \in S_2(1).$$

Это можно увидеть в следующем факте. Оказывается (см. А. Blum, J. Hopcroft, R. Kannan, *Foundations of Data Science*; К. Ball, *An elementary introduction to modern convex geometry*; В. А. Зорич, *Математический анализ в задачах естествознания*), что с вероятностью хотя бы $1 - \frac{2}{c} e^{-\frac{c^2}{2}}$ будет выполнено неравенство $|\langle l, e \rangle| \leq \frac{c}{\sqrt{n-1}}$, где l — произвольный фиксированный вектор единичной евклидовой нормы.

Ключевая лемма

Последнее неравенство можно для $q = \infty$ записать в следующем виде (не умаляя общности, считаем, что $\|s\|_2 = 1$):

$$\mathbb{E}_e [\langle s, e \rangle^2 \|e\|_\infty^2] \lesssim \frac{1}{n} \cdot \frac{\ln n}{n} \quad \forall s \in S_2(1).$$

Это можно увидеть в следующем факте. Оказывается (см. A. Blum, J. Hopcroft, R. Kannan, *Foundations of Data Science*; K. Ball, *An elementary introduction to modern convex geometry*; В. А. Зорич, *Математический анализ в задачах естествознания*), что с вероятностью хотя бы $1 - \frac{2}{c} e^{-\frac{c^2}{2}}$ будет выполнено неравенство $|\langle l, e \rangle| \leq \frac{c}{\sqrt{n-1}}$, где l — произвольный фиксированный вектор единичной евклидовой нормы. Беря $c = 10$ и $l = s$, получаем, что с большой вероятностью $\langle s, e \rangle^2 \leq \frac{100}{n}$, а взяв $c = 2\sqrt{\ln n}$ и l , направленные вдоль координатных осей, можно получить, что с вероятностью хотя бы $1 - \frac{1}{n\sqrt{n}}$ выполняется неравенство $\|e\|_\infty^2 \leq \frac{4 \ln n}{n}$.

Algorithm 1 Accelerated Randomized Directional Derivative (ARDD) method

Вход: x_0 — некоторая стартовая точка; N — количество итераций; m — размер батча.

Выход: точка y_N

- 1: $y_0 \leftarrow x_0, z_0 \leftarrow x_0$
- 2: **for** $k = 0, \dots, N - 1$
- 3: $\alpha_{k+1} \leftarrow \frac{k+2}{96n^2\rho_n L_2}, \tau_k \leftarrow \frac{1}{48\alpha_{k+1}n^2\rho_n L_2} = \frac{2}{k+2}.$
- 4: Сгенерировать $e_{k+1} \in RS_2^n(1)$ независимо от предыдущих итераций и $\xi_i, i = 1, \dots, m$ — независимые реализации ξ .
- 5: Вычислить

$$\tilde{\nabla}^m f(x_{k+1}) = \frac{1}{m} \sum_{i=1}^m \tilde{f}'(x_{k+1}, \xi_i, e_{k+1}) e_{k+1}.$$

- 6: $x_{k+1} \leftarrow \tau_k z_k + (1 - \tau_k) y_k.$
 - 7: $y_{k+1} \leftarrow x_{k+1} - \frac{1}{2L_2} \tilde{\nabla}^m f(x_{k+1}).$
 - 8: $z_{k+1} \leftarrow \operatorname{argmin}_{z \in \mathbb{R}^n} \left\{ \alpha_{k+1} n \left\langle \tilde{\nabla}^m f(x_{k+1}), z - z_k \right\rangle + V[z_k](z) \right\}.$
 - 9: **end for**
 - 10: **return** y_N
-

Сходимость по функции

Теорема

Пусть метод ARDD применяется для решения задачи (1). тогда

$$\begin{aligned} \mathbb{E}[f(y_N)] - f(x^*) &\leq \frac{384\Theta_p n^2 \rho_n L_2}{N^2} + \frac{4N}{nL_2} \cdot \frac{\sigma^2}{m} + \frac{61N}{24L_2} \Delta_\zeta + \frac{122N}{3L_2} \Delta_\eta^2 \\ &\quad + \frac{12\sqrt{2n\Theta_p}}{N^2} \left(\frac{\sqrt{\Delta_\zeta}}{2} + 2\Delta_\eta \right) \\ &\quad + \frac{N^2}{12n\rho_n L_2} \left(\frac{\sqrt{\Delta_\zeta}}{2} + 2\Delta_\eta \right)^2, \end{aligned} \quad (7)$$

где $\Theta_p = V[z_0](x^*)$ определяется выбором прокс-структуры и $\mathbb{E}[\cdot] = \mathbb{E}_{e_1, \dots, e_N, \xi_{1,1}, \dots, \xi_{N,m}}[\cdot]$.

Сходимость по функции

	$p = 1$	$p = 2$
N	$O\left(\sqrt{\frac{n \ln n L_2 \Theta_1}{\varepsilon}}\right)$	$O\left(\sqrt{\frac{n^2 L_2 \Theta_2}{\varepsilon}}\right)$
m	$O\left(\max\left\{1, \sqrt{\frac{\ln n}{n}} \cdot \frac{\sigma^2}{\varepsilon^{3/2}} \cdot \sqrt{\frac{\Theta_1}{L_2}}\right\}\right)$	$O\left(\max\left\{1, \frac{\sigma^2}{\varepsilon^{3/2}} \cdot \sqrt{\frac{\Theta_2}{L_2}}\right\}\right)$
Δ_ζ	$O\left(\min\left\{n(\ln n)^2 L_2^2 \Theta_1, \frac{\varepsilon^2}{n \Theta_1}, \frac{\varepsilon^{\frac{3}{2}}}{\sqrt{n \ln n}} \cdot \sqrt{\frac{L_2}{\Theta_1}}\right\}\right)$	$O\left(\min\left\{n^3 L_2^2 \Theta_2, \frac{\varepsilon}{n \Theta_2}, \frac{\varepsilon^{\frac{3}{2}}}{n} \cdot \sqrt{\frac{L_2}{\Theta_2}}\right\}\right)$
Δ_η	$O\left(\min\left\{\sqrt{n \ln n L_2} \sqrt{\Theta_1}, \frac{\varepsilon}{\sqrt{n \Theta_1}}, \frac{\varepsilon^{\frac{3}{4}}}{\sqrt[4]{n \ln n}} \cdot \sqrt[4]{\frac{L_2}{\Theta_1}}\right\}\right)$	$O\left(\min\left\{n^{\frac{3}{2}} L_2 \sqrt{\Theta_2}, \frac{\varepsilon}{\sqrt{n \Theta_2}}, \frac{\varepsilon^{\frac{3}{4}}}{\sqrt[4]{n}} \cdot \sqrt[4]{\frac{L_2}{\Theta_2}}\right\}\right)$
O-le calls	$O\left(\max\left\{\sqrt{\frac{n \ln n L_2 \Theta_1}{\varepsilon}}, \frac{\sigma^2 \Theta_1 \ln n}{\varepsilon^2}\right\}\right)$	$O\left(\max\left\{\sqrt{\frac{n^2 L_2 \Theta_2}{\varepsilon}}, \frac{\sigma^2 \Theta_2 n}{\varepsilon^2}\right\}\right)$

Table: Оценки на число итераций, общее число вызовов оракула и параметры в методе ARDD для случаев $p = 1$ и $p = 2$.

Ускоренный безградиентный метод

Рассмотрим теперь следующую постановку. Пусть в заданной паре точек (x, y) оракул возвращает значения $(\tilde{f}(x, \xi), \tilde{f}(y, \xi))$ некоторой зашумлённой стохастической реализации функции f , где

$$\tilde{f}(x, \xi) = F(x, \xi) + \Xi(x, \xi), \quad |\Xi(x, \xi)| \leq \Delta, \quad \forall x \in \mathbb{R}^n, \text{ п.в. } \xi.$$

Ускоренный безградиентный метод

В качестве стохастической аппроксимации $\nabla f(x)$ будем использовать

$$\tilde{\nabla}^m f^t(x) = \frac{1}{m} \sum_{i=1}^m \frac{\tilde{f}(x+te, \xi_i) - \tilde{f}(x, \xi_i)}{t} e$$

Ускоренный безградиентный метод

В качестве стохастической аппроксимации $\nabla f(x)$ будем использовать

$$\begin{aligned}\tilde{\nabla}^m f^t(x) &= \frac{1}{m} \sum_{i=1}^m \frac{\tilde{f}(x+te, \xi_i) - \tilde{f}(x, \xi_i)}{t} e \\ &= \left(\left\langle g^m(x, \vec{\xi}_m), e \right\rangle + \frac{1}{m} \sum_{i=1}^m (\zeta(x, \xi_i, e) + \eta(x, \xi_i, e)) \right) e,\end{aligned}\tag{8}$$

Ускоренный безградиентный метод

В качестве стохастической аппроксимации $\nabla f(x)$ будем использовать

$$\begin{aligned}\tilde{\nabla}^m f^t(x) &= \frac{1}{m} \sum_{i=1}^m \frac{\tilde{f}(x+te, \xi_i) - \tilde{f}(x, \xi_i)}{t} e \\ &= \left(\left\langle g^m(x, \vec{\xi}_m), e \right\rangle + \frac{1}{m} \sum_{i=1}^m (\zeta(x, \xi_i, e) + \eta(x, \xi_i, e)) \right) e,\end{aligned}\quad (8)$$

где $e \in RS_2^n(1)$, ξ_i , $i = 1, \dots, m$ — независимые реализации ξ , m — размер батча, t — некоторое маленькое положительное число, которое мы будем называть *сглаживающим параметром*,

Ускоренный безградиентный метод

В качестве стохастической аппроксимации $\nabla f(x)$ будем использовать

$$\begin{aligned}\tilde{\nabla}^m f^t(x) &= \frac{1}{m} \sum_{i=1}^m \frac{\tilde{f}(x+te, \xi_i) - \tilde{f}(x, \xi_i)}{t} e \\ &= \left(\left\langle g^m(x, \vec{\xi}_m), e \right\rangle + \frac{1}{m} \sum_{i=1}^m (\zeta(x, \xi_i, e) + \eta(x, \xi_i, e)) \right) e,\end{aligned}\quad (8)$$

где $e \in RS_2^n(1)$, ξ_i , $i = 1, \dots, m$ — независимые реализации ξ , m — размер батча, t — некоторое маленькое положительное число, которое мы будем называть *сглаживающим параметром*,

$$g^m(x, \vec{\xi}_m) := \frac{1}{m} \sum_{i=1}^m g(x, \xi_i)$$

Ускоренный безградиентный метод

В качестве стохастической аппроксимации $\nabla f(x)$ будем использовать

$$\begin{aligned}\tilde{\nabla}^m f^t(x) &= \frac{1}{m} \sum_{i=1}^m \frac{\tilde{f}(x+te, \xi_i) - \tilde{f}(x, \xi_i)}{t} e \\ &= \left(\left\langle g^m(x, \vec{\xi}_m), e \right\rangle + \frac{1}{m} \sum_{i=1}^m (\zeta(x, \xi_i, e) + \eta(x, \xi_i, e)) \right) e,\end{aligned}\quad (8)$$

где $e \in RS_2^n(1)$, ξ_i , $i = 1, \dots, m$ — независимые реализации ξ , m — размер батча, t — некоторое маленькое положительное число, которое мы будем называть *сглаживающим параметром*,

$$g^m(x, \vec{\xi}_m) := \frac{1}{m} \sum_{i=1}^m g(x, \xi_i) \text{ и}$$

$$\begin{aligned}\zeta(x, \xi_i, e) &= \frac{F(x+te, \xi_i) - F(x, \xi_i)}{t} - \langle g(x, \xi_i), e \rangle, \\ \eta(x, \xi_i, e) &= \frac{\Xi(x+te, \xi_i) - \Xi(x, \xi_i)}{t}, \quad i = 1, \dots, m.\end{aligned}$$

Ускоренный безградиентный метод

$$\begin{aligned}\zeta(x, \xi_i, e) &= \frac{F(x+te, \xi_i) - F(x, \xi_i)}{t} - \langle g(x, \xi_i), e \rangle, \\ \eta(x, \xi_i, e) &= \frac{\Xi(x+te, \xi_i) - \Xi(x, \xi_i)}{t}, \quad i = 1, \dots, m.\end{aligned}$$

Из липшицевости градиента $F(\cdot, \xi)$ мы имеем $|\zeta(x, \xi, e)| \leq \frac{L(\xi)t}{2}$ для всех $x \in \mathbb{R}^n$ и $e \in S_2(1)$. Отсюда следует, что $\mathbb{E}_\xi(\zeta(x, \xi, e))^2 \leq \frac{L_2^2 t^2}{4}$ для всех $x \in \mathbb{R}^n$ и $e \in S_2(1)$.

Ускоренный безградиентный метод

$$\begin{aligned}\zeta(x, \xi_i, e) &= \frac{F(x+te, \xi_i) - F(x, \xi_i)}{t} - \langle g(x, \xi_i), e \rangle, \\ \eta(x, \xi_i, e) &= \frac{\Xi(x+te, \xi_i) - \Xi(x, \xi_i)}{t}, \quad i = 1, \dots, m.\end{aligned}$$

Из липшицевости градиента $F(\cdot, \xi)$ мы имеем $|\zeta(x, \xi, e)| \leq \frac{L(\xi)t}{2}$ для всех $x \in \mathbb{R}^n$ и $e \in S_2(1)$. Отсюда следует, что $\mathbb{E}_\xi(\zeta(x, \xi, e))^2 \leq \frac{L_2^2 t^2}{4}$ для всех $x \in \mathbb{R}^n$ и $e \in S_2(1)$. Кроме того, из $|\Xi(x, \xi)| \leq \Delta$ получаем, что $|\eta(x, \xi, e)| \leq \frac{2\Delta}{t}$ для всех $x \in \mathbb{R}^n$, $e \in S_2(1)$ и почти всех ξ .