Курбаба Денис ДЗ-3.

Во всех задачах предполагается, что в задаче

$$F(x) = f(x) + R(x) \to \min_{x\in\mathbb{R}^m}$$

функция $f(x)$ является μ-сильно выпуклой, $R(x)$ - правильная замкнутая выпуклая функция, x^* - точка минимума функции $F(x)$.

① Доказать, что $x^* = \mathrm{prox}_{\gamma R}(x^* - \gamma \nabla f(x^*))$, где $\gamma > 0$

• Определим прокс-оператор след. образом:

$$\mathrm{prox}_f(x) = \arg\min_{u\in\mathbb{R}^m} \left\{ f(u) + \frac{1}{2}\|u-x\|_2^2 \right\} \qquad (\mathbb{R}^m \to \mathbb{R}^m)$$

Св-во прокс-оператора:

$$u = \mathrm{prox}_R(x) \iff x-u \in \partial R(u) \iff$$

$$\iff \langle x-u, y-u\rangle \le R(y) - R(u) \quad \forall y \in \mathbb{R}^m$$

• Тогда используя св-во выше получим, что если произведём след. преобразования $F(x) = f(x) + R(x)$ умножив на γ и $\frac{d}{dx}$.

$$\gamma F(x) = \gamma f(x) + \gamma R(x) \quad \text{и} \quad \dot F(x) \equiv \dot f(x) + \dot R(x)$$

то можем заменить исходное ур-е $x^* = \mathrm{prox}_{\gamma R}(x^* - \gamma \nabla f(x^*))$ на эквивалентное $x^* = \mathrm{prox}_{\dot R(x)}(x^* - \nabla \dot f(x^*))$.

И теперь, используя св-во:

$$u = x^* = \mathrm{prox}_{\dot R(x)}(x^* - \nabla \dot f(x^*)) \overset{u}{\iff} x = x^* - \nabla \dot f(x^*)$$

приведение к обозначениям, использованных в определении св-ва

Итак $u = \mathrm{prox}_{\dot R(x)}(x^* - \nabla \dot f(x^*)) \iff x - u = \underline{x^*} - \nabla \dot f(x^*) - \underline{x^*} = -\nabla \dot f(x^*)$

Теперь для того, чтобы доказать что $x^* = \mathrm{prox}_{\gamma R}(x^* - \gamma \nabla f(x^*))$ необх. доказать, что $x - u = -\nabla \dot f(x^*) \in \partial \dot R(x)$.

• Итак, доказали, что $-\nabla \dot{f}(x^*) \in \delta \dot{R}(x^*)$.

Заметим вновь, что т.к. x^* - аргумент точки минимума для $F(x)$ - по условию, то по св-ву минимума: x^* является точкой минимума и для производной $\dot{F}(x)$, а это значит, что $\delta \dot{F}(x^*) = \{0\}$.

Теперь распишем значение

$$\delta \dot{F}(x^*) = \delta\left(\dot{f}(x^*) + \dot{R}(x^*)\right) = \delta \dot{f}(x^*) + \delta \dot{R}(x^*) = \{0\}$$

Ну и осталось лишь заметить, что $\dot{f}(x^*)$ - выпуклая по условию, а значит $\delta \dot{f}(x^*) = \nabla \dot{f}(x^*)$.

Соберем выражения: $\nabla \dot{f}(x^*) + \delta \dot{R}(x^*) = 0$

$$\delta \dot{R}(x^*) = -\nabla \dot{f}(x^*)$$

$$\Downarrow$$

$$-\nabla \dot{f}(x^*) \in \delta \dot{R}(x^*)$$

доказали

Значит доказали и исх. утверждение. ■

② $\exists$ $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n \mid x_i > 0,\ i = \overline{1,n}\}$ и

$$R(x) = \begin{cases} -\gamma \sum_{i=1}^{n} \ln x_i, & x \in \mathbb{R}^n_{++} \\ +\infty, & \text{иначе}, \end{cases} \quad \text{где } \gamma > 0$$

Найти $\text{prox}_R(x)$

$$\text{prox}_R(x) = \underset{y \in \mathbb{R}^n_{++}}{\arg\min} \left\{ R(y) + \frac{1}{2}\|x - y\|_2^2 \right\}$$

Разобьем ф-ю $R(x)$ на подфункции, которые действуют из $\mathbb{R}_+$ в $\mathbb{R}$, то есть теперь будем искать

$$\text{prox}\,R(x) = \left[\text{prox}\,R_1(x_1) \quad \text{prox}\,R_2(x_2) \ \dots \ \text{prox}\,R_n(x_n)\right]^T$$

$$\Updownarrow$$

$$U = [U_1 \ U_2 \dots U_n]^T$$

• Применим св-во прокси-оператора: $\left(u=\operatorname{prox}_R(x) \Leftrightarrow x-u\in\delta R(u)\right)$.

Получаем:

$$x-u\in\delta R(u) \quad \text{и} \quad \begin{cases} x_1-u_1\in\delta R_1(u_1)=\dot R_1(u_1)=-\frac{\gamma}{u_1} \\ x_2-u_2\in\delta R_2(u_2)=\dot R_2(u_2)=-\frac{\gamma}{u_2} \\ \dots \\ x_n-u_n\in\delta R_n(u_n)=\dot R_n(u_n)=-\frac{\gamma}{u_n} \end{cases}$$

т.к. $R_i\in C(\mathbb{R})$.

Итого, имеем $x_i-u_i\in\delta R_i(u_i)=\dot R_i(u_i)=-\frac{\gamma}{u_i}$

$$x_i-u_i=-\frac{\gamma}{u_i}$$

$$x_iu_i-u_i^2=-\gamma$$

$$u_i^2-x_iu_i-\gamma=0$$

подставили конкретное значение $R_i(x_i)$ из условия.

$$D=(-x_i)^2-4(-\gamma)=x_i^2+4\gamma$$

$$u_{i_1}=\frac{x_i-\sqrt{x_i^2+4\gamma}}{2}, \quad u_{i_2}=\frac{x_i+\sqrt{x_i^2+4\gamma}}{2}$$

Узнаем, какой вариант нам подходит (u_{i_1} или u_{i_2})

• Теперь, используя определение $\operatorname{prox}_R(x)=\underset{y\in\mathbb{R}^{++}_n}{\operatorname{argmin}}\left\{R(y)+\frac{1}{2}\|x+y\|_2^2\right\}=$

$=\underset{y\in\mathbb{R}^{++}_n}{\operatorname{argmin}}\left\{-\gamma\sum_{i=1}^{n}\ln y_i+\frac{1}{2}\|x-y\|_2^2\right\} \Rightarrow$ все $u_i>0$, значит

выбираем $u_{i_2}=\frac{x_i+\sqrt{x_i^2+4\gamma}}{2}$

• Итого,

$$\operatorname{prox} R(x)=\left(\frac{x_1+\sqrt{x_1^2+4\gamma}}{2}\quad\frac{x_2+\sqrt{x_2^2+4\gamma}}{2}\quad\dots\quad\frac{x_n+\sqrt{x_n^2+4\gamma}}{2}\right)^T$$

---

③ ∡ задачу $F(x)=f(x)+R(x)\to\min\limits_{x\in\mathbb{R}^n}$ (1), в которой

$$f(x)=\frac{1}{m}\sum_{i=1}^{m}f_i(x),$$

где $f_i(x)$ - выпуклые и L-гладкие функции для всех $i=\overline{1,m}$.

$\exists$ есть m компьютеров (рабочих), которые соединены с одним сервером (мастером), но не соединены между собой.

$\exists$ мы хотим решать задачу (1) в такой архитектуре обычным проксимальным градиентным спуском с шагом γ.

Тогда на каждой итерации i-й рабочий должен вычислить $\nabla f_i(x^k)$ (для всех $i=\overline{1,m}$), затем отправить $\nabla f_i(x^k)$

мастеру, который вычисляет среднее арифм. от полученных векторов, т.е. вычисляет $\nabla f(x^k)$, вычисляет $x^{k+1} = \text{prox}_{\gamma R}(x^k - \gamma \nabla f(x^k))$ и отправляет x^{k+1} всем рабочим. В данной процедуре большая нагрузка ложится на мастера, т.к. m и n могут быть большими.

Разумная попытка по устранению этого недостатка в том, чтобы вместо $\nabla f_i(x^k)$ отправлять несмещенную оценку g_i^k, которая имеет меньше битов информации. Один из таких способов - это квантизация/спарсификация.

<u>def</u> (Квантизация/спарсификация). Будем называть стохастический оператор $Q(x)$ оператором квантизации или просто квантизацией, если $\forall x \in \mathbb{R}^n$ выполняется:

$$E(Q(x)) = x, \quad E(\|Q(x) - x\|_2^2) \leq \omega \|x\|_2^2,$$

где $\omega \geq 0$.

Легко заметить, что тождественный оператор $Q(x) = x$ удовлетворяет def с константой $\omega = 0$.

Рассм. стохастический оператор

$$\text{rand}_t(x) = \frac{n}{t} \sum_{i \in S} x_i e_i,$$

где t - фиксированное число из мн-ва $\{1, \dots, n\}$, S - случайное подмн-во мн-ва $\{1, \dots, n\}$ размера t, $(e_1, \dots, e_n)$ - стандартный базис в $\mathbb{R}^n$. Иными словами, среди всех компонент вектора x выбираются t компонент случайно и равновероятно, а полученный вектор домножается на $\frac{n}{t}$, чтобы добиться несмещенности. Покажите, что данный оператор удовлетворяет def с константой $\omega = \frac{n}{t} - 1$ (hint: use $E(\|Q(x)\|_2^2) = E(\|Q(x) - x\|_2^2) + \|x\|_2^2$.

1) Докажем, что $E(Q(x)) = x$, т.е. $E(\text{rand}_t(x)) = x$

Рассм. $E(\text{rand}_t(x)) = E\left(\frac{n}{t} \sum_{i \in S} x_i e_i\right) \overset{\text{св.-ва}}{=} \frac{n}{t} \sum_{i \in S} e_i E(x_i) =$

расписали EX:

$= \frac{n}{t}\sum_{i\in S} e_i \cdot \frac{t}{n}\cdot x_i = \sum_{i\in S} e_i x_i = x$, доказали первое условие.

т.к. выбор равновероятный по усл.

2)

$\ast$ hint: $E(\|Q(x)\|_2^2) = E(\|Q(x)-x\|_2^2) + \|x\|_2^2$

в нашем случае $Q(x) = rand_t(x) = \frac{n}{t}\sum_{i\in S} x_i e_i$, заменим:

Тогда получается

$$E(\|rand_t(x)-x\|_2^2) = E\left(\left\|\frac{n}{t}\sum_{i\in S} x_i e_i\right\|_2^2\right) - \|x\|_2^2 =$$

аналогично 1)

$$= \left\|\frac{n}{t}\sum e_i\, E(x_i)\right\|_2^2 - \|x\|_2^2 \underset{\text{раскрыли } \|\,\|_2^2}{=} \frac{n^2}{t^2}\sum_{i\in S} E x_i^2 - \|x\|_2^2 =$$

расписали EX как в 1)

$$\frac{n^2}{t^2}\cdot\frac{t}{n}\cdot\sum_{i\in S} x_i^2 - \|x\|_2^2 = \frac{n}{t}\sum_{i\in S} x_i^2 - \|x\|_2^2$$

Теперь, т.к. S - случ. подмн-во размера t, а $t\in\{1,\dots,n\}$ и $t\le n$, то оценим выражение сверху:

$$E(\|rand_t(x)-x\|_2^2) = \frac{n}{t}\sum_{i\in S} x_i^2 - \|x\|_2^2 \le \frac{n}{t}\sum_{i=1}^{n} x_i^2 - \|x\|_2^2 =$$

$$= \|x\|_2^2\left(\frac{n}{t}-1\right) = \omega\|x\|_2^2$$

Итого, получили $E(\|rand_t(x)-x\|_2^2) \le \omega\cdot\|x\|_2^2$ (II условие).

∎