CONVEX OPTIMIZATION

Computer Final Project

Radmehr Karimian - Daniel Ayati

Contents

1 Theoretical Problems

1.1 Problem #1

We simply have:

$$\frac{\partial}{\partial x}(\frac{1}{2}x^T Ax - b^T x + c) = Ax - b \Rightarrow Ax^* - b = 0 \tag{1}$$

$$x^* = A^{-1}b \Rightarrow x^{(t+1)} - A^{-1}b = x^{(t)} - \eta(Ax^{(t)} - b) - A^{-1}b \tag{2}$$

$$x^{(t+1)} - A^{-1}b = (x^{(t)} - A^{-1}b)(I - \eta A) \tag{3}$$

$$\Rightarrow x^{(t)} - A^{-1}b = (x^{(0)} - A^{-1}b)(I - \eta A)^t \tag{4}$$

$$\Rightarrow ||I - \eta A||_2 < 1 \Rightarrow 0 < \lambda_i < \frac{2}{\eta} \tag{5}$$

$\blacksquare$

1.2 Problem #2

$$\frac{\partial}{\partial x}[f(x^{(t)}) + \langle \nabla(f(x^{(t)}), x - x^{(t)}\rangle + \frac{1}{2\eta}||x - x^{(t)}||_2^2] \tag{6}$$

$$= \nabla f(x^{(t)} - \frac{1}{\eta}(x^{(t+1)} - x^{(t)}) \tag{7}$$

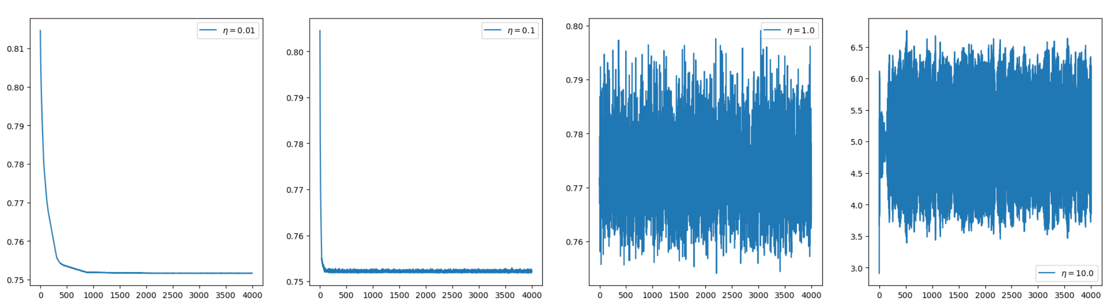
$$\Rightarrow x^{(t+1)} = x^{(t)} - \frac{1}{\eta}\nabla f(x^{(t)}) \tag{8}$$

$\blacksquare$

2 Computer Problems

2.1 Problem #1

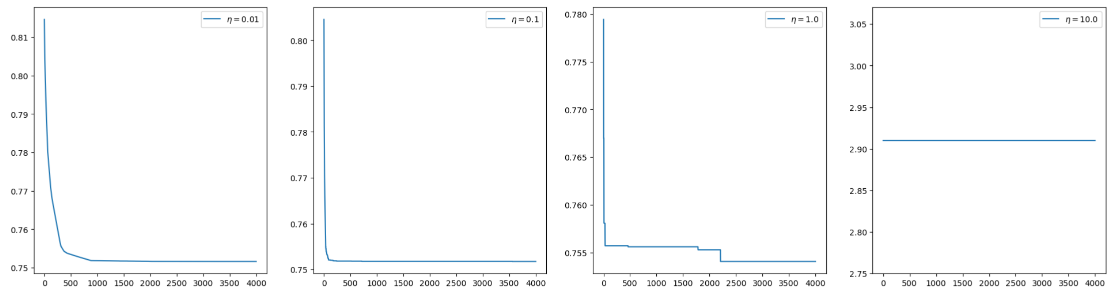
The codes are available in the ipynb and here is the output:



As you can see we have better answer and actually converge where η is small.

2.2 Problem #2

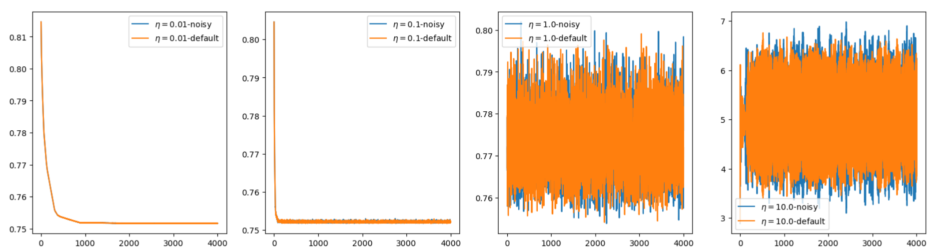
The codes are available in the ipynb and here is the output:



As you can see we don't have telorance much here but when η is small we still have better answer.

2.3 Problem #3

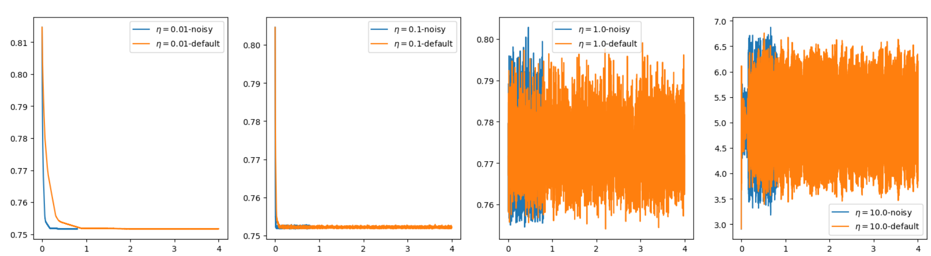
The codes are available in the ipynb and here is the output:



As you can see both answers converges to same results.

2.4 Problem #4

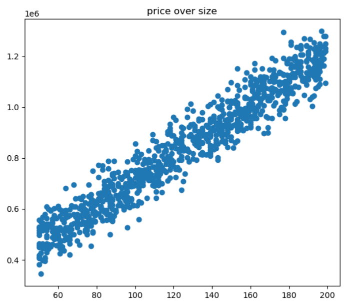
The codes are available in the ipynb and here is the output:

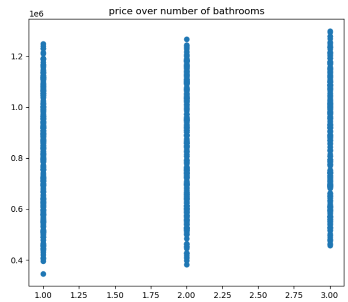


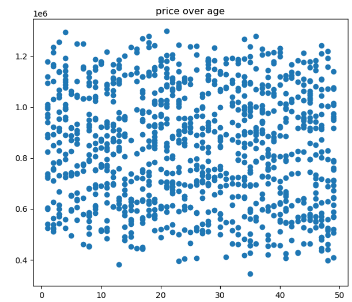
Our new method converges faster to the same result!

2.5 Problem #5

you can see some relation of price and our features here:

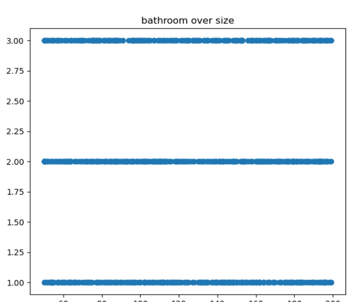


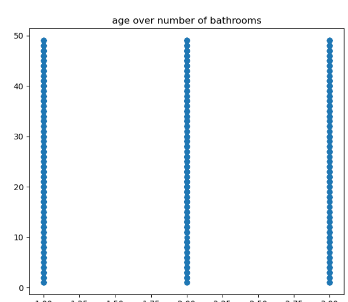


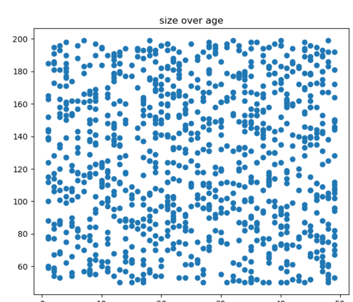


For dummy variables like city we should have another method:

next part is important and even the homework designer had forgot! We should always check that our features correlation isn't too much, because it makes some serious problem for our model specially when it's linear.







2.6 Problem #6

The codes are available in the ipynb.

2.7 Problem #7

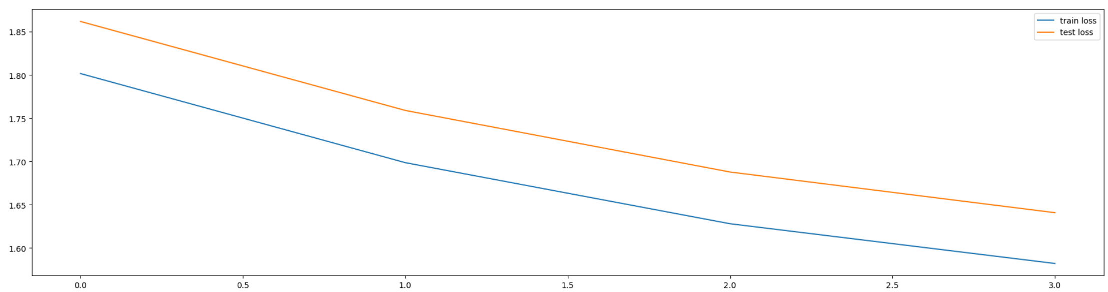
The codes are available in the ipynb.

2.8 Problem #8

The codes are available in the ipynb.

2.9 Problem #9

The codes are available in the ipynb and here is the result: With 5 Epochs and 16 Batchs and learning rate = 0.001 we have:



3 Theoretical Problems-hand written

Thanks to weak time management and other projects we haven't enough time to latex all problems

سوال ۳

$$f(x^{(t)}) \geq f(x^*) \Rightarrow f(x^*) \geq f(x) + \langle v^{(t)}, x^* - x^{(t)} \rangle$$

$$\hookrightarrow \langle v^{(t)}, x^* - x^{(t)} \rangle < 0$$

حال داریم:

$$\|x^{(t+1)} - x^*\|_2^2 = \|x^t - \eta_t v^{(t)} - x^*\|_2^2 = \|x^{(t)} - x^*\|_2^2 - 2\eta_t \langle v^{(t)}, x^t - x^* \rangle + \eta_t^2 \|v^t\|_2^2$$

$$\Rightarrow \eta_t \in \left(0, \frac{2(f(x^t) - f(x^*))}{\|v^{(t)}\|_2^2}\right)$$

سوال ٤

~~برهان خلف~~:

$$f(x^{(t+1)}) - f(x^{(t)}) \leq \rho \|x^{(t+1)} - x^{(t)}\|_2^2$$

$$\Rightarrow f(x^{(t+1)}) \geq f(x^{(t)}) + \langle x^{(t+1)} - x^{(t)}, v \rangle$$

$$\epsilon \geq 0 \Rightarrow x^{(t+1)} = x^{(t)} + \varepsilon \left(\frac{|v|}{\|v\|}\right) \Rightarrow \langle x^{(t+1)} - x^{(t)}, v \rangle \geq \varepsilon \|v\|$$

$$\Rightarrow f(x^{(t+1)}) - f(x^{(t)}) \geq \varepsilon \|v\| \Rightarrow \rho\varepsilon \geq \varepsilon \|v\|$$

$$\Rightarrow \rho \geq \|v\| \checkmark$$

سوال ۵

از سوال ۳ داریم:

$$\|x^{(t+1)} - x^*\|_2^2 = \|x^t - \eta_t v^{(t)} - x^*\|_2^2$$

$$= \|x^t - x^*\|_2^2 - 2\eta_t \langle v^t, x^t - x^* \rangle + \eta_t^2 \|v^t\|_2^2$$

$$\leq \|x^t - x^*\|_2^2 - 2\eta_t (f(x^t) - f(x^*)) + \eta_t^2 \|v^t\|_2^2$$

$\eta_2 \Rightarrow$ کمتر می‌شود $\checkmark$

سوال ۶

از سوال قبل:

$$\sum \eta_t (f(x^t) - f(x^*)) \le \sum \tfrac{1}{2}\left(\|x^{(t+1)} - x^*\|^2 - \|x^{(t)} - x^*\|^2 + \eta_t \|v^{(t)}\|^2\right)$$

$$= \sum \tfrac{\eta_t^2}{2}\|v^t\|^2 - \tfrac{1}{2}\left(\|x^T - x^*\|^2 - \|x^1 - x^*\|^2\right)$$

$$\le \tfrac{\rho^2}{2}\eta_t^2 + \tfrac{1}{2}\|x^*\|^2 \le \tfrac{\rho^2}{2}\sum \eta_t^2 + \tfrac{1}{2}B^2 \checkmark$$

سوال ۷

از اثبات قبلی داریم

$$\sum \eta_t (f(x^t) - f(x^*)) \le \tfrac{B^2}{2} + \tfrac{\rho^2}{2}\sum \eta_t^2$$

$$\Rightarrow \frac{\sum \eta_t f(x^t)}{\sum \eta_t} - f(x^*) \le \frac{B^2 + \rho^2 \sum \eta_t^2}{2\sum \eta_t}$$

از جنسن داریم

$$\frac{\sum \eta_t f(x^t)}{\sum \eta_t} \ge f\left(\frac{\sum \eta_t x^t}{\sum \eta_t}\right) = f(\bar{x}_T)$$

$$\Rightarrow f(\bar{x}_T) - f(x^*) \le \frac{B^2 + \rho^2 \sum \eta_t^2}{2\sum \eta_t}$$

سوال ۸

$$\frac{\partial}{\partial\eta}\left(\frac{B^2+\rho^2\sum\eta^2}{2\sum\eta}\right) = \frac{\partial}{\partial\eta}\left(\frac{B^2+\rho^2\eta^2\sum_0^T 1}{2\eta\sum_0^T 1}\right) = \frac{\partial}{\partial\eta}\left(\frac{B^2+\rho^2\eta^2(T+1)}{2\eta(T+1)}\right)$$

$$\left(\frac{B^2}{\eta(T+1)}+\frac{(T+1)\rho^2\eta^2}{(T+1)\eta}\right) = \frac{\rho^2}{2}\eta + \frac{B^2}{2(T+1)\eta} \Rightarrow \frac{\partial}{\partial\eta}\left(\frac{B^2}{2}\eta+\frac{B^2}{2(T+1)\eta}\right)$$

$$= \rho^2/2 - \frac{B^2}{2(T+1)\eta^2} = 0 \Rightarrow \eta = \frac{B}{\rho\sqrt{T+1}} \checkmark \Rightarrow \text{u.b}: \frac{B^2+\rho^2\eta^2(T+1)}{2\eta(T+1)}$$

$$= \frac{B^2+\cancel{\rho^2}\times\frac{B^2}{\cancel{\rho^2}\cancel{(T+1)}}\cdot\cancel{(T+1)}}{2\frac{B}{\rho}\cdot\frac{1}{\sqrt{T+1}}} = \frac{2B^2}{2B}\cdot\frac{\rho}{\sqrt{T+1}} = \frac{B\rho}{\sqrt{T+1}} \checkmark$$

سوال ۹

درست است که در زیگزاگ اداماِن نزدیکی در حال حرکت هستیم اما نکته قابل توجه دیگر این است که در قدم های خیلی کوچک است
که کاهش نداشته باشیم. برای جبران این اشتباهات که به خاطر گام و سایز از نقطه‌ای که بهترین عملکرد ما را خراب می کند.
میانگین که استدلال شکل استفاده میکنیم.

سوال ۱۰

$$\frac{\sum\eta_t f(x^t)}{\sum\eta_t} - f(x^*) \le \frac{B^2+\rho^2\sum\eta_t^2}{2\sum\eta_t}$$

$$f(x^{t^*}) \le f(x^t) \Rightarrow \frac{\sum\eta_t f(x^t)}{\sum\eta_t} \ge \frac{\sum\eta_t f(x^{t^*})}{\sum\eta_t} = f(x^{t^*})$$

$$\Rightarrow f(x^{t^*}) - f(x^*) \le \frac{B^2+\rho^2\sum\eta^2}{2\sum\eta} \checkmark$$

سوال ۸

$$\min_{y} \ \tfrac{1}{2}\|y-x\|_2^2$$
$$Ay=b$$

$$L(y,\nu) = \tfrac{1}{2}\|y-x\|_2^2 + \nu^T(Ay-b)$$

KKT

$$\frac{\partial}{\partial y}L = 0 \Rightarrow y - x + A^T\nu = 0 \Rightarrow \nu = (AA^T)^{-1}(Ax - Ay)$$

$$\Rightarrow y = x - A^T(AA^T)^{-1}(Ax-Ay) \rightarrow \Pi_C(x) = x - A^T(AA^T)^{-1}(Ax-Ay)$$

سوال ۱۲

با توجه به شرایط KKT و سوال قبل :

$$\begin{cases} Ax \geq b, & x - A^T(AA^T)^{-1}(Ay-Ax) \\ Ax \leq b, & x \end{cases}$$

سوال ۱۳

$$\min \tfrac{1}{2}\|y-x\|_2^2$$

$$L(y,\lambda) = \tfrac{1}{2}\|y-x\|_2^2 + \tfrac{\lambda}{2}(\|y\|_2^2 - b^2) \Rightarrow \nabla L = 0 \Rightarrow (\lambda+1)y - x = 0$$

$$\Rightarrow y = \frac{1}{1+\lambda} x \quad \text{KKT} \begin{cases} \|y\|_2^2 - b^2 \leq 0 \\ \lambda(\|y\|_2^2 - b^2) = 0 \end{cases} \overset{\text{حل}}{\Rightarrow} \begin{cases} \|x\| \leq b \Rightarrow \lambda = 0 \\ \|x\| > b \Rightarrow \lambda = \frac{\|x\|_2}{b} - 1 \end{cases}$$

$$\Rightarrow \Pi_C \begin{cases} \frac{b}{\|x\|_2} x & \|x\|_2 > b \\ x & \text{o.w} \end{cases}$$

سوال ۱۴

$$x^{t+1} = \underset{x\in C}{\operatorname{argmin}} \|x^t - \eta_t \nabla^t - x\|_2 = \underset{x\in C}{\operatorname{argmin}} \|x^t - \eta_t \nabla^t - x\|_2^2$$

$$= \operatorname{argmin}\{\|x^t - x\|_2^2 - 2\eta_t\langle \nabla^t, x^t - x\rangle + \eta_t^2\|\nabla^t\|_2^2\}$$

$$= \operatorname{argmin}\{\tfrac{1}{2\eta_t}\|x - x^t\|_2^2 + \langle \nabla^t, x - x^t\rangle + f(x^t)\} \checkmark$$

سوال ۱۵

projection lemma:

$$\|w-u\|_2^2 \geq \|v-u\|_2^2 \quad \text{(اثبات در کتاب هست.)}$$

۹
اگر با تصویر سازی غیر
منبسط نباشد
۸
مدیر

$$0 \leq \tfrac{1}{2}\|x^{T+1} - x^*\|_2^2 = \tfrac{1}{2}\|\Pi_C(z^{T+1}) - x^*\|_2^2$$

$$\leq \tfrac{1}{2}\|z^{T+1} - x^*\|_2^2 = \tfrac{1}{2}\|x^T - \eta_T \nabla^T - x^*\|_2^2$$

$$= \tfrac{1}{2}\|x^T - x^*\|_2^2 - \eta_T(f(x^T) - f(x^*)) + \frac{\eta_T^2}{2}\|\nabla^T\|_2^2$$

$$\leq \tfrac{1}{2}\|x^1 - x^*\|_2^2 - \sum \eta_t(f(x^t) - f(x^*)) + \sum \tfrac{1}{2}\eta_t^2\|\nabla^t\|_2^2$$

$$\le \tfrac{1}{2}B^2 + \tfrac{\rho^2}{2}\sum \eta_t^2 - \sum \eta_t \left(f(x^t) - f(x^*)\right)$$

$$\Rightarrow \eta_T \sum \left(f(x^t) - f(x^*)\right) \le \tfrac{1}{2}B^2 + \tfrac{\rho^2}{2}\sum \eta_T^2$$

ازای η_t نزولی است.

$$\frac{1}{2}B^2 + \frac{\rho^2}{2}\sum \eta_T^2 \le \tfrac{1}{2}B^2 + \frac{\rho^2\eta_T}{2}\sum \eta_t$$

$$\Rightarrow \sum \left(f(x^t) - f(x^*)\right) \le \frac{1}{2\eta_T}B^2 + \tfrac{\rho^2}{2}\sum \eta_t$$

سوال ۱۶

$$f(\bar{x}_t) - f(x^*) = f\left(\frac{1}{T}\sum f(x_t)\right) - f(x^*) \underset{\text{Jensen}}{\le} \frac{1}{T}\sum f(x_t) - f(x^*)$$

$$= \frac{1}{T}\left(\sum f(x_t) - f(x^*)\right) \underset{(15)}{\le} \frac{1}{T}\left(\frac{1}{2\eta_T}B^2 + \frac{\rho^2}{2}\sum \eta_t\right)$$

$$= \frac{1}{T}\left(\frac{\sqrt{T}}{2\alpha}B^2 + \tfrac{\rho^2}{2}\sum \tfrac{\alpha}{\sqrt{t}}\right) = \frac{B^2}{2\alpha\sqrt{T}} + \frac{\rho^2\alpha}{T}\sum \tfrac{1}{\sqrt{t}}$$

$$\le \frac{B^2}{2\alpha\sqrt{T}} + \frac{\rho^2\alpha}{2T}\sum \frac{2}{\sqrt{t}-\sqrt{t-1}} \le \frac{B^2}{2\alpha\sqrt{T}} + \frac{\rho^2\alpha}{2T}\sum \frac{2(\sqrt{t}-\sqrt{t-1})}{t-(t-1)}$$

$$= \frac{B^2}{2\alpha\sqrt{T}} + \frac{\rho^2\alpha}{2}(\sqrt{T}-1) \checkmark \Rightarrow \frac{\partial}{\partial\alpha} = 0 \Rightarrow -\frac{B^2}{2\alpha^2\sqrt{T}} + \frac{\rho^2}{2\sqrt{T}} \Rightarrow \alpha = \frac{B}{\rho}$$

$$\Rightarrow \text{u.b}: \frac{B\rho}{\sqrt{T}} \checkmark$$

سوال ۱۷

دلیل تصادفی بودن این موضوع این است که بردارهای V و X استفاده شده تصادفی هستند.

سوال ۱۸

سمت شرطی بودن امید این است که درست است خود نقطه تصادفی است اما در قسمت ۲ امید E را گرفتیم و استفاده از آن این شرط قطعی شده است.

سوال ۱۹
شکل سوال قبل

$E(\bar{v}^{(t)}) \in \partial f(x^{(t)})$

$$\frac{1}{2}\|x^{t+1}-x^*\|_2^2 = \frac{1}{2}\|\Pi_C(x^t-\eta_t v^{(t)})-x^*\|_2^2$$
$$\leq \|x^t-\eta_t v^t-x^*\|_2^2$$
$$= \frac{1}{2}\|x^t-x^*\|_2^2-\eta_t\langle \bar{v}^{(t)}, x^{(t)}-x^*\rangle + \eta_t^2\frac{1}{2}\|v^{(t)}\|_2^2$$
$$= \frac{1}{2}\|x^{(t)}-x^*\|_2^2-\eta_t\langle \bar{v}^{(t)}-E(\bar{v}^{(t)}|x^{(t)}), x^{(t)}-x^*\rangle$$
$$-\eta_t\langle E(\bar{v}^{(t)}|x^{(t)}), x^{(t)}-x^*\rangle+\eta_t^2\frac{1}{2}\|v^{(t)}\|_2^2$$
$$\leq \frac{1}{2}\|x^{(t)}-x^*\|_2^2-\eta_t\langle \xi_t, x^{(t)}-x^*\rangle-\eta_t(f(x^{(t)})-f(x^*))$$
$$+\eta_t\frac{1}{2}\|v^{(t)}\|_2^2 \checkmark$$

سوال ۲۰

$$E(\langle \xi_t, x^{(t)}-x^*\rangle) = E(\langle v^{(t)}-E(\bar{v}^{(t)}|x^{(t)}), x^{(t)}-x^*\rangle)$$
$$= \langle E(\bar{v}^{(t)}-E(\bar{v}^{(t)}|x^{(t)})), x^{(t)}-x^*\rangle$$
$$= \langle E(\bar{v}(t))-E(E(\bar{v}^{(t)}|x^{(t)})), x^{(t)}-x^*\rangle$$
$$= \langle E(v^{(t)})-E(v^{(t)}), x^{(t)}-x^*\rangle = \langle 0, x^t-x^*\rangle = 0 \checkmark$$

سوال ۱

$$E(f(\bar{X}^T)) - f(x^*) = E(f(\bar{X}^T) - f(x^*))$$

$$\underset{\text{assum.}}{\leq} E\left(\frac{1}{T}\sum (f(x^{(t)}) - f(x^*))\right)$$

$$\leq E\left(\frac{1}{T}\sum\left(\frac{1}{2\eta_t}(\|x^{(t)} - x^*\|^2 - \|x^{(t+1)} - x^*\|^2) + \frac{\eta_t}{2}\|v^{(t)}\|^2 - \langle \xi_t, x^{(t)} - x^* \rangle\right)\right)$$

$$\leq E\left(\frac{1}{T}\sum\left(\frac{1}{2\eta_t}(\|x^{(t)} - x^*\|^2 - \|x^{(t+1)} - x^*\|^2)\right)\right) + \frac{\rho^2}{2T}\sum \eta_t$$

$$\leq E\left(\frac{1}{T}\cdot\frac{\|x^{(1)} - x^*\|^2}{2\eta_1} + \frac{1}{T}\sum\left(\frac{1}{2}\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\|x^{(t+1)} - x^*\|^2\right) + \frac{1}{T\eta_T}\|x^{(T)} - x^*\|\right) + \frac{\rho^2}{2T}\sum \eta_t$$

$$\leq \frac{\|x^{(1)} - x^*\|^2}{2T\eta_1} + \frac{\rho^2}{2T}\sum \eta_t \leq B^2/2T\eta_t + \frac{\rho^2}{2T}\sum \eta_t$$

سوال ۲

$$\frac{B^2}{2T\eta_T} + \frac{\rho^2}{2T}\sum \eta_t = \frac{B^2}{2T\frac{B}{\rho\sqrt{T}}} + \frac{\rho^2}{2T}\sum \frac{B}{\rho\sqrt{t}} \leq \frac{B\rho}{2\sqrt{T}} + \frac{\rho^2}{2T}\sum \frac{B}{\rho(\sqrt{t} + \sqrt{t-1})}$$

$$= \frac{B\rho}{2}\left(\frac{1}{\sqrt{T}} + \sum \sqrt{t} - \sqrt{t-1}\right)$$

$$= \cancel{\frac{B\rho}{2}\left(\frac{1}{\sqrt{T}} + \sqrt{T}\right)}\ \frac{B\rho}{\sqrt{T}}$$

سوال ۳۳

$$\langle x^{(t)} - x^*, E(v^{(t)} \mid x^{(t)}) \rangle \geq f(x^{(t)}) - f(x^*) + \frac{\lambda}{2} \|x^{(t)} - x^*\|_2^2$$

$$\|x^t - x^*\|_2^2 - \|x^{(t+1)} - x^*\|_2^2 = \|x^{(t)} - x^*\|_2^2 - \|\Pi_C(x^t - \eta_t v^t) - x^*\|_2^2$$ از لم دوم

$$\geq \|x^{(t)} - x^*\|_2^2 - \|x^{(t)} - \eta_t v^{(t)} - x^*\|_2^2 \geq \|x^{(t)} - x^*\|_2^2 - \|x^{(t)} - \eta_t v^{(t)} - x^*\|_2^2$$

$$= 2\eta_t \langle x^{(t)} - x^*, v^{(t)} \rangle - \eta_t^2 \|v^{(t)}\|_2^2$$

$$\Rightarrow \langle x^{(t)} - x^*, E(v^{(t)} \mid x^{(t)}) \rangle \leq \frac{E(\|x^{(t)} - x^*\|_2^2 - \|x^{(t+1)} - x^*\|_2^2)}{2\eta_t} + \frac{\eta_t}{2} \rho^2$$

$$\cancel{\eta_t = \frac{1}{\lambda t} \Rightarrow} \quad E(f(x^{(t)})) - f(x^*) + \frac{\lambda}{2} E(\|x^{(t)} - x^*\|_2^2)$$

$$\leq \frac{E(\|x^{(t)} - x^*\|_2^2 - \|x^{(t+1)} - x^*\|_2^2)}{2\eta_t} + \frac{\eta_t}{2} \rho^2$$

$\eta_t = \frac{1}{\lambda t}$:

$$E(f(x^{(t)})) - f(x^*) \leq \frac{\lambda}{2} E(t\|x^{(t)} - x^*\|_2^2 - (t+1)\|x^{(t+1)} - x^*\|_2^2) + \frac{\rho^2}{2\lambda t}$$

$$\Rightarrow \sum \left(E(f(x^{(t)})) - f(x^*)\right) \leq \frac{\lambda}{2} E\left(\sum \left(t\|x^{(t)} - x^*\|_2^2 - (t+1)\|x^{(t+1)} - x^*\|_2^2\right)\right) + \frac{\rho^2}{2\lambda} \sum \frac{1}{t}$$

$$= \lambda/2\, E(-T\|x^{T+1} - x^*\|_2^2) + \frac{\rho^2}{2\lambda} \sum \frac{1}{t} = -\frac{\lambda T}{2} \|x^{T+1} - x^*\|_2^2 + \frac{\rho^2}{2\lambda} \sum \frac{1}{t}$$

$$\leq -\frac{\lambda T}{2} \|x^{T+1} - x^*\|_2^2 + \frac{\rho^2}{2\lambda}(1 + \ln(T)) \Rightarrow \times \frac{1}{T}$$ = حکم ثابت شد

سوال ۲۴

$$F(y,z) \geq F(x,z) + \langle v_{x,z}, y-x \rangle$$

امید ریاضی خطی است،

$$\Rightarrow F(y,z) - F(x,z) - \langle v, y-x \rangle \geq 0$$

$$\Rightarrow E_z\big(F(y,z) - F(x,z) - \langle v, y-x \rangle\big) \geq 0$$

$$\Rightarrow E_z\big(F(y,z)\big) - E_z\big(F(x,z)\big) - E_z\big(\langle v_{x,z}, y-x \rangle\big) \geq 0 \checkmark$$