به نام خدا


**دانشگاه صنعتی شریف**
**دانشکده مهندسی برق**


# گزارش پروژه درس بهینه سازی محدب


**مهدی عباس زاده ، امیرحسین یاری**
**۹۹۱۰۶۴۰۳ ، ۹۹۱۰۲۵۰۷**



**July ۴، ۲۰۲۳**

# Contents

## پرسش تئوری ۱

ابتدا رابطه بین $\mathbf{x}^{(t+1)}$ و $\mathbf{x}^{(t)}$ را میابیم:

$$\nabla f(\mathbf{x}^{(t)}) = \mathbf{A}\mathbf{x}^{(t)} - b$$

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

حال چون مسئله بدون قید است برای $\mathbf{x}^*$ میدانیم که

$$\nabla f(\mathbf{x}^*) = \mathbf{A}\mathbf{x}^* - b = 0$$

بنابراین میتوانیم $\mathbf{x}^*$ را از دو طرف رابطه قبلی کم کنیم و به نتایج زیر برسیم:

$$\longrightarrow \mathbf{x}^{(t+1)} - \mathbf{x}^* = (\mathbf{I} - \eta\mathbf{A})\mathbf{x}^{(t)} + \eta b - (\mathbf{x}^* - \eta(\mathbf{A}\mathbf{x}^* - b))$$

$$\longrightarrow \mathbf{x}^{(t+1)} - \mathbf{x}^* = (\mathbf{I} - \eta\mathbf{A})\mathbf{x}^{(t)} - (\mathbf{I} - \eta\mathbf{A})\mathbf{x}^* = (\mathbf{I} - \eta\mathbf{A})(\mathbf{x}^{(t)} - \mathbf{x}^*) \qquad (۱)$$

حال برای ماتریس مثبت معین $\mathbf{A}$ میدانیم که :

$$\lambda_{min}(\mathbf{A})\mathbf{I_n} \preceq \mathbf{A} \preceq \lambda_{max}(\mathbf{A})\mathbf{I_n}$$

$$\longrightarrow \mathbf{I_n} - \eta\lambda_{min}(\mathbf{A})\mathbf{I_n} \succeq \mathbf{I_n} - \eta\mathbf{A} \Longrightarrow ||\mathbf{I_n}||_2|1 - \eta\lambda_{min}| \geq ||\mathbf{I_n} - \eta\mathbf{A}||_2$$

از طرفی چون ممکن است که $|1 - \eta\lambda_{max}| > |1 - \eta\lambda_{min}|$ باشد، به نامساوی زیر میرسیم:

$$||\mathbf{I_n} - \eta\mathbf{A}||_2 \leq \max\{|1 - \eta\lambda_{max}|, |1 - \eta\lambda_{min}|\}$$

حال با استفاده از این رابطه و معادله (۱) میتوانیم به عبارت زیر برسیم (دقت کنید که برای نرم دو ماتریس داریم: $||\mathbf{A}||_2||\mathbf{B}||_2 \geq ||\mathbf{AB}||_2$) :

$$||\mathbf{x}^{(t+1)} - \mathbf{x}^*||_2 \leq ||(\mathbf{I} - \eta\mathbf{A})||_2||(\mathbf{x}^{(t)} - \mathbf{x}^*)||_2 \leq \max\{|1 - \eta\lambda_{max}|, |1 - \eta\lambda_{min}|\}||(\mathbf{x}^{(t)} - \mathbf{x}^*)||_2$$

با ادامه این روند برای $t$ های کوچکتر و با توجه به اینکه $\mathbf{x}_0 = 0$ است، به عبارت زیر میرسیم:

$$||\mathbf{x}^{(t+1)} - \mathbf{x}^*||_2 \leq \max\{|1 - \eta\lambda_{max}|, |1 - \eta\lambda_{min}|\}^{t+1}||\mathbf{x}^*||_2$$

بنابراین طبق این معادله، اگر شرط زیر برقرار باشد، وقتی $(T \to \infty)$ الگوریتم به نقطه بهینه همگرا خواهد شد:

$$\max\{|1 - \eta\lambda_{max}|, |1 - \eta\lambda_{min}|\} < 1 \qquad (۲)$$

دقت کنید که اگر دو مقدار $(m, M)$ ای یافت شود که به ترتیب از $(\lambda_{min}, \lambda_{max})$ کوچکتر و بزرگتر باشند، با جایگذاری آن ها در رابطه (۲) ، نامساوی همچنان برقرار خواهد بود و میتوان شرط همگرایی را به طور کلی تر و به صورت وجود داشتن دو عدد با شرایط ذکر شده بیان کرد.

## پرسش تئوری ۲)

از این رابطه نسبت بهز این رابطه نسبت به $\mathbf{x}$ گرادیان میگیریم و عبارت حاصل را برابر صفر قرار میدهیم:

$$\nabla f(\mathbf{x}^{(t)}) + \frac{1}{\eta}(\mathbf{x} - \mathbf{x}^{(t)}) = 0 \longrightarrow \mathbf{x} = -\eta \nabla f(\mathbf{x}^{(t)}) + \mathbf{x}^{(t)}$$

تعبیر شهودی این رابطه را میتوان اینگونه بیان کرد که ما در هر مرحله از الگوریتم، تقریب درجه دوم تابع را در نقطه $\mathbf{x}^{(t)}$ در نظر میگیریم، نقطه ای که این تابع درجه دوم را کمینه میکند پیدا میکنیم(که برابر $f(\mathbf{x}^{(t)}) + \mathbf{x}^{(t)}$ است). بنابراین گویی در هرمرحله، تقریب درجه دوم تابع در آن نقطه را کمینه میکنیم تا به نقطه کمینه خود تابع اصلی برسیم.

## پرسش تئوری ۳)

$$||\mathbf{x}^{(t+1)} - \mathbf{x}^*||^2 = ||\mathbf{x}^{(t)} - \mathbf{x}^*||^2 + \eta_t^2 ||\mathbf{v}^{(t)}||^2 - 2\eta_t < \mathbf{v}^{(t)}, \mathbf{x}^{(t)} - \mathbf{x}^* >$$

بنابراین اگر داشته باشیم:

$$\eta_t < \frac{2}{||\mathbf{v}^{(t)}||^2} < \mathbf{v}^{(t)}, \mathbf{x}^{(t)} - \mathbf{x}^* >$$

عبارت داده شده اکیدا از $||\mathbf{x}^{(t)} - \mathbf{x}^*||^2$ کمتر خواهد بود، بنابراین به ازای هر $\mathbf{v}^{(t)}$ ، میتوان طول گامی پیدا کرد که ما را به نقطه بهینه نزدیک کند.

## پرسش تئوری ۴)

یک نقطه دلخواه $\mathbf{x} \in \mathbb{R}^n$ از دامنه $f$ انتخاب میکنیم و یک بردار دلخواه زیرگرادیان $\mathbf{v} \in \partial f(\mathbf{x})$ را برمیگزینیم. عدد کوچک $\epsilon$ را جوری انتخاب میکنیم که $\mathbf{y} = \epsilon\mathbf{v} + \mathbf{x}$ درون دامنه تابع $f$ بیفتد. خاصیت زیرگرادیان را برای این دو نقطه مینویسیم:

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \epsilon ||\mathbf{v}||_2^2 \xrightarrow{\text{طرف راست نامنفی است}} |f(\mathbf{y}) - f(\mathbf{x})| \geq \epsilon ||\mathbf{v}||_2^2$$

حال از خاصیت $\rho$ ـ لیپشیتز بودن تابع استفاده میکنیم:

$$\Longrightarrow \rho ||\mathbf{y} - \mathbf{x}||_2 = \rho\epsilon ||\mathbf{v}||_2 \geq \epsilon ||\mathbf{v}||_2^2 \longrightarrow \rho \geq ||\mathbf{v}||_2$$

و حکم نتیجه میشود.

## پرسش تئوری ۵)

طبق الگوریتم داریم:

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

که $\mathbf{v}^{(t)} \in \partial f(\mathbf{x})$ است. بنابراین برای این زیر گرادیان، به ازای هر نقطه $\mathbf{x} \in \mathbb{R}^n$ من جمله نقطه $\mathbf{x}^*$ داریم:

$$< \mathbf{v}^{(t)}, \mathbf{x}^{(t)} - \mathbf{x}^* > \leq f(\mathbf{x}^{(\mathbf{t})}) - f(\mathbf{x}^*)$$

با جایگذاری این نامساوی و تقسیم کردن تمام جمله ها بر ۲، به حکم خواسته شده میرسیم:

$$\frac{1}{2}||\mathbf{x}^{(t+1)} - \mathbf{x}^*||_2^2 \leq \frac{1}{2}||\mathbf{x}^{(t)} - \mathbf{x}^*||_2^2 - \eta_t(f(\mathbf{x}^{(\mathbf{t})}) - f(\mathbf{x}^*)) + \frac{\eta_t^2}{2}||\mathbf{v}^{(t)}||_2^2 \qquad (۳)$$

## پرسش تئوری ۶)

اگر در نامساوی (۳) قرار دهیم $t = T$ و این نامساوی را به طور باز گشتی ادامه دهیم به عبارت زیر میرسیم:

$$\frac{1}{2}||\mathbf{x}^{(T+1)} - \mathbf{x}^*||_2^2 \leq \frac{1}{2}||\mathbf{x}^{(T-1)} - \mathbf{x}^*||_2^2 - \sum_{t=T-1}^{T} \eta_t(f(\mathbf{x}^{(\mathbf{t})}) - f(\mathbf{x}^*)) + \sum_{t=T-1}^{T} \frac{\eta_t^2}{2}||\mathbf{v}^{(t)}||_2^2$$

$$\longrightarrow \cdots \leq \frac{1}{2}||\mathbf{x}^{(1)} - \mathbf{x}^*||_2^2 - \sum_{t=1}^{T} \eta_t(f(\mathbf{x}^{(\mathbf{t})}) - f(\mathbf{x}^*)) + \sum_{t=1}^{T} \frac{\eta_t^2}{2}||\mathbf{v}^{(t)}||_2^2$$

حال دقت کنید که چون $\mathbf{x}^{(1)} = \mathbf{x}_0 = \mathbf{0}$ انتخاب شده است و همچنین تابع $f$ ، $\rho$ ـ لیپشیتز فرض شده است و دیگر اینکه طبق فرض دیگری برای نقطه بهینه داریم که $||\mathbf{x}^*||_2 \leq B$ ، به نامساوی زیر میرسیم:

$$\frac{1}{2}||\mathbf{x}^{(T+1)} - \mathbf{x}^*||_2^2 \leq \frac{1}{2}B^2 - \sum_{t=1}^{T} \eta_t(f(\mathbf{x}^{(\mathbf{t})}) - f(\mathbf{x}^*)) + \frac{1}{2}\rho^2 \sum_{t=1}^{T} \eta_t^2$$

حال چون $\frac{1}{2}||\mathbf{x}^{(T+1)} - \mathbf{x}^*||_2^2 \geq 0$ است، با بردن جمله دوم به سمت چپ، به نامساوی خواسته شده در صورت پرسش میرسیم:

$$\sum_{t=1}^{T} \eta_t(f(\mathbf{x}^{(\mathbf{t})}) - f(\mathbf{x}^*)) \leq \frac{1}{2}B^2 + \frac{1}{2}\rho^2 \sum_{t=1}^{T} \eta_t^2 \tag{۴}$$

## پرسش تئوری ۷)

نامساوی (۴) را بر عبارت $\sum_{t=1}^{T} \eta_t$ تقسیم میکنیم و به عبارت زیر میرسیم:

$$\frac{\sum_{t=1}^{T} \eta_t f(\mathbf{x}^{(\mathbf{t})})}{\sum_{t=1}^{T} \eta_t} - f(\mathbf{x}^*) \leq \frac{B^2 + \rho^2 \sum_{t=1}^{T} \eta_t^2}{2\sum_{t=1}^{T} \eta_t}$$

حال قرار میدهیم $\theta_t = \frac{\eta_t}{\sum_{t=1}^{T} \eta_t}$ ، چون تابع $f$ محدب است، داریم:

$$\sum_{t=1}^{T} \theta_t f(\mathbf{x}^{(t)}) \geq f(\sum_{t=1}^{T} \theta_t \mathbf{x}^{(t)})$$

بنابراین چون $\sum_{t=1}^{T} \theta_t \mathbf{x}^{(t)}$ برابر $\bar{\mathbf{x}}_T$ تعریف شده است، با جایگذاری نامساوی بالا، حکم خواسته شده نتیجه میشود:

$$f(\bar{\mathbf{x}}_T) - f(\mathbf{x}^*) \leq \frac{B^2 + \rho^2 \sum_{t=1}^{T} \eta_t^2}{2\sum_{t=1}^{T} \eta_t}$$

## پرسش تئوری ۸)

با بازنویسی کران بدست آمده، به عبارت زیر میرسیم:

$$f(\bar{\mathbf{x}}_T) - f(\mathbf{x}^*) \leq \frac{B^2}{2T\eta} + \frac{\rho^2}{2}\eta$$

مقدار بهینه کران جدید به ازای $\frac{B}{\rho\sqrt{T}} = \sqrt{\frac{\frac{B^2}{2T}}{\frac{\rho^2}{2}}} = \eta$ رخ میدهد. با جایگذاری این مقدار به جای $\eta$ ، به کران $\frac{\rho B}{\sqrt{T}}$ میرسیم.

## پرسش تئوری ۹

می دانیم که در زیرگرادیان نزولی به ازای step های خاصی کاهش رخ می دهد و امکان دارد در هر update کاهش رخ ندهد و گاها افزایش رخ دهد. برای اینکه update های اشتباه در الگوریتم جبران شود، از میانگین نقاط می توان استفاده کرد. پس به دلایل فوق می توان برای مقدار تابع در میانگین نقاط کران بالا محاسبه نمود ولی برای مقدار تابع در خود نقاط update شده، نمی توان کران بالا محاسبه کرد زیرا ممکن است نسبت به update قبلی کاهش پیدا نکرده باشد.

## پرسش تئوری ۱۰

در نامساوی (۴) ، به جای $f(\mathbf{x}^{(t)})$ ، حداقل آن ها یعنی $f(\mathbf{x}^{(t_T^*)})$ را جایگذاری میکنیم:

$$\left(f(\mathbf{x}^{(t_T^*)}) - f(\mathbf{x}^*)\right)\sum_{t=1}^{T}\eta_t \leq \frac{1}{2}B^2 + \frac{1}{2}\rho^2\sum_{t=1}^{T}\eta_t^2$$

با تقسیم کردن دو طرف بر $\sum_{t=1}^{T}\eta_t$ ، به حکم خواسته شده میرسیم:

$$f(\mathbf{x}^{(t_T^*)}) - f(\mathbf{x}^*) \leq \frac{B^2 + \rho^2\sum_{t=1}^{T}\eta_t^2}{2\sum_{t=1}^{T}\eta_t}$$

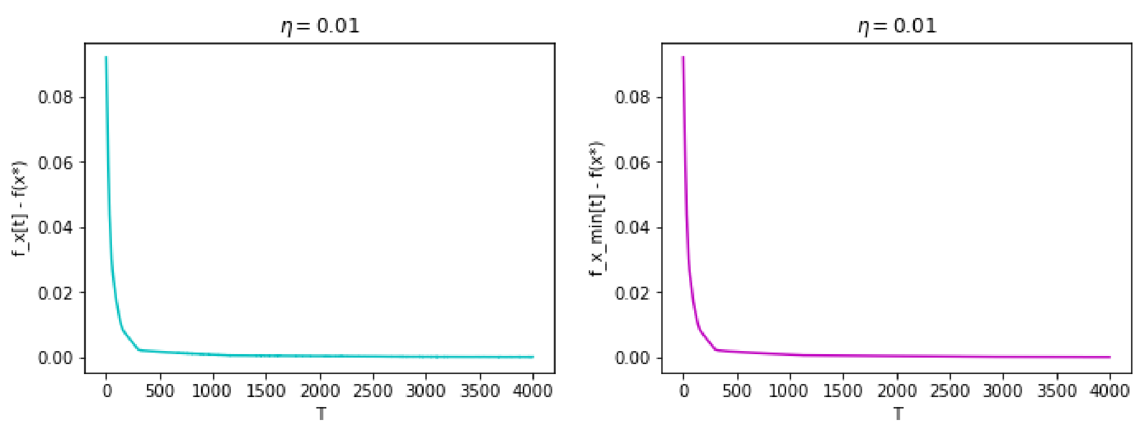
## پرسش شبیه سازی (۱و۲)

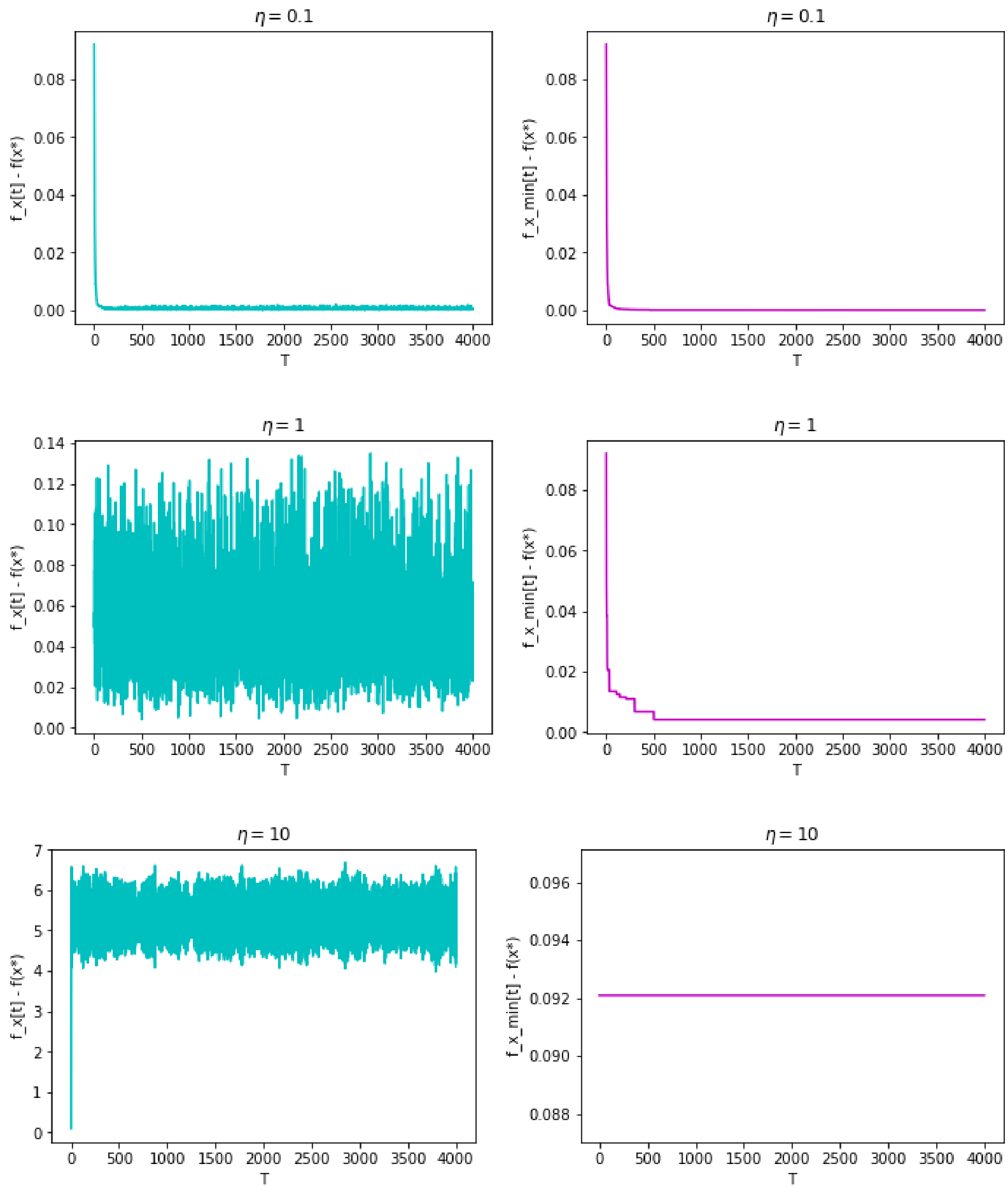
```
# tropmI deriuqer yrarbil
tropmi numpy as np
tropmi matplotlib
tropmi matplotlib.pyplot as plt
tropmi cvxpy as cp
tropmi time

# ezilaitinI retemarap
eta = [0.01, 0.1, 1, 10]
n = 10
m = 100
cov = np.identity(n)
mean = [0] * 10
Z = np.zeros((m, n))
B = np.zeros((m, 1))
T = 4000
x_0 = np.zeros(((n, 1)))
f_x_t = np.zeros((T, 1))
f_x_min_t = np.zeros((T, 1))
convergence_var = np.zeros((nel(eta),T))
convergence_var_2 = np.zeros((T, 1))
rof i ni np.arange(m):
Z[i, :] = np.random.multivariate_normal(mean, cov)
B[i] = np.random.normal(0, 1)

# gnidniF eht mumitpo eulav gnisu ypxvc
x = cp.Variable((n, 1))
objective_func = cp.Minimize(cp.norm((Z@x - B), 1))
constraints = []
prob = cp.Problem(objective_func, constraints)
prob.solve()

x_opt = x.value
f_opt = prob.value/m

# tnirP tuptuO
tnirp("mumitpo_x = ", x.value)
tnirp("f(x)* = ", f_opt)
```

x optimum = [ 0.29553518, -0.33065322, -0.12790722, 0.06042707, -0.14963684, 0.05338799, -0.03175885, 0.20391252, 0.08564487, -0.00970474]
f(x*) = 0.8247408824635432

```
# gnitnemelpmi eht mhtirogla
subgrad = np.zeros((n, 1))
elapsed_time_GD = np.zeros((nel(eta), T))
rof e ni eta:
x_t = x_0
rof t ni np.arange(T):
begin = time.time()
f_x_t[t] = np.mus(np.sba(Z@x_t - B))/m
fi (t == 0):
f_x_min_t[t] = f_x_t[t]
file (f_x_t[t] < f_x_min_t[t - 1]):
```

```
f_x_min_t[t] = f_x_t[t]
esle:
f_x_min_t[t] = f_x_min_t[t - 1]
convergence_var[eta.index(e), t] = f_x_t[t] - f_opt
end = time.time()
elapsed_time_GD[eta.index(e), t] = end - begin
convergence_var_2[t] = f_x_min_t[t] - f_opt
rof i ni np.arange(m):
tmp = np.dot(Z[i, :], x_t) - B[i]
fi (tmp > 0):
subgrad += np.reshape(Z[i , :], (n,1))
file (tmp == 0):
a = np.random.rand()
subgrad += np.reshape(a * Z[i, :] + (1 - a) * (-Z[i, :]), (n, 1))
esle:
subgrad -= np.reshape(Z[i , :], (n,1))
subgrad /= m
x_t = x_t - e*subgrad

# gnittolP
plt.figure(figsize=(10,4))
plt.subplot(1,2,1)
plt.plot(np.arange(1, T+1), convergence_var[eta.index(e),:], "c")
plt.xlabel("T")
plt.ylabel("x_f[t] - f(x)*")
plt.title((r'$\ate = $' + rts(e)))
plt.subplot(1,2,2)
plt.plot(np.arange(1, T+1), convergence_var_2, "m")
plt.xlabel("T")
plt.ylabel("nim_x_f[t] - f(x)*")
plt.title((r'$\ate = $' + rts(e)))
plt.tight_layout(pad = 2)
```




$\eta = 0.1$
f_x[t] - f(x*)
f_x_min[t] - f(x*)
$\eta = 1$
$\eta = 10$
T
0 500 1000 1500 2000 2500 3000 3500 4000

## پرسش تئوری ۱۱)

برای محاسبه ی فرم بسته تابع $\Pi_{\mathcal{C}}$ می بایست مسئله بهینه سازی زیر را حل کنیم.

$$\text{Minimize} \quad ||\mathbf{y}-\mathbf{x}||^2$$

$$\text{subject to} \quad \mathbf{A}\mathbf{y}=\mathbf{b}$$

پس لاگرانژ آن بصورت زیر است:

$$\Rightarrow \mathcal{L}(\mathbf{x},\mathbf{y},\boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x}+\mathbf{y}^T\mathbf{y}-2\mathbf{x}^T\mathbf{y}+\boldsymbol{\nu}^T(\mathbf{A}\mathbf{y}-\mathbf{b})$$

$$\frac{d\,\mathcal{L}}{d\,\mathbf{y}} = 2\mathbf{y}-2\mathbf{x}+\mathbf{A}^T\boldsymbol{\nu}=0$$

$$\mathbf{y}=\frac{1}{2}(2\mathbf{x}-\mathbf{A}^T\boldsymbol{\nu})$$

از طرفی می دانیم:

$$\mathbf{A}\mathbf{y}=\mathbf{b}$$

پس با ضرب $\mathbf{A}$ در طرفین عبارت بالا داریم:

$$\mathbf{b}=\frac{1}{2}\mathbf{A}(2\mathbf{x}-\mathbf{A}^T\boldsymbol{\nu})$$

$$\frac{1}{2}\mathbf{A}\mathbf{A}^T\boldsymbol{\nu}=\mathbf{A}\mathbf{x}-\mathbf{b}$$

$$\boldsymbol{\nu}=2(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x}-\mathbf{b})$$

با جایگذاری $\boldsymbol{\nu}$ بدست آمده خواهیم داشت:

$$\mathbf{y}=\mathbf{x}-\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x}-\mathbf{b})$$
$$\Rightarrow \Pi_{\mathcal{C}}(\mathbf{x})=\mathbf{x}-\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x}-\mathbf{b})$$

## پرسش تئوری ۱۲ )

برای محاسبه ی فرم بسته تابع $\Pi_{\mathcal{C}}$ می بایست مسئله بهینه سازی زیر را حل کنیم.

$$\text{Minimize} \quad ||\mathbf{y}-\mathbf{x}||^2$$

$$\text{subject to} \quad \mathbf{A}\mathbf{y} \preceq \mathbf{b}$$

پس لاگرانژ آن بصورت زیر است:

$$\mathcal{L}(\mathbf{y},\boldsymbol{\lambda}) = ||\mathbf{y}-\mathbf{x}||^2 + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{y}-\mathbf{b})$$
$$\Rightarrow \nabla_{\mathbf{y}}\mathcal{L} = 2\mathbf{y} - 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\lambda}$$

باتوجه به شرایط KKT داریم:

$$\begin{cases} 2\mathbf{y} - 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\lambda} = 0 \\ \mathbf{A}\mathbf{y} \preceq \mathbf{b} \\ \lambda_i(\mathbf{A}_i^T\mathbf{y} - b_i) = 0 \end{cases}$$

## پرسش تئوری ۱۳)

برای محاسبه ی فرم بسته تابع $\Pi_C$ می بایست مسئله بهینه سازی زیر را حل کنیم.

$$\text{Minimize} \quad ||\mathbf{y}-\mathbf{x}||^2$$

$$\text{subject to} \quad \|\mathbf{y}\|^2 - b^2 \leq 0$$

پس لاگرانژ آن بصورت زیر است:

$$\mathcal{L}(\mathbf{y},\lambda) = ||\mathbf{y}-\mathbf{x}||^2 + \lambda(\|\mathbf{y}\|^2 - b^2)$$
$$\Rightarrow \nabla_{\mathbf{y}}\mathcal{L} = 2(\lambda+1)\mathbf{y} - 2\mathbf{x}$$

باتوجه به شرایط KKT داریم:

$$\begin{cases} 2(\lambda+1)\mathbf{y} - 2\mathbf{x} = 0 \\ \|\mathbf{y}\|^2 - b^2 \leq 0 \\ \lambda(\|\mathbf{y}\|^2 - b^2) = 0 \end{cases}$$

باتوجه به شروط اول و دوم می توان گفت:

$$\mathbf{y} = \frac{\mathbf{x}}{\lambda+1} \Rightarrow \|\mathbf{y}^2\| - b^2 = \frac{\|\mathbf{x}\|^2}{(\lambda+1)^2} \leq 0$$

$$\Rightarrow \frac{\|\mathbf{x}\|}{b} - 1 \leq \lambda$$

پس اگر $\|\mathbf{x}\|^2 \leq b$ خواهیم داشت:

$$\lambda > 0 \Rightarrow \|\mathbf{y}^2\| - b^2 = \frac{\|\mathbf{x}\|^2}{(\lambda+1)^2} \leq 0 \Rightarrow \lambda(\|\mathbf{y}^2\| - b^2) < 0 \quad ✗$$

$$\lambda = 0 \Rightarrow \mathbf{y} = \frac{\mathbf{x}}{\lambda+1} = \mathbf{x}$$

و اگر $\|\mathbf{x}\|^2 \geq b$ خواهیم داشت:

$$\lambda \geq \frac{\|\mathbf{x}\|}{b} - 1 > 0 \Rightarrow \|\mathbf{y}^2\| - b^2\frac{\|\mathbf{x}\|^2}{(\lambda+1)^2} = 0 \Rightarrow \lambda = \frac{\|\mathbf{x}\|}{b} - 1$$

$$\lambda = 0 \Rightarrow \mathbf{y} = \frac{\mathbf{x}}{\lambda+1}\mathbf{x} = \frac{\mathbf{x}}{\|\mathbf{x}\|}b$$

## پرسش تئوری ۱۴)

$$
\begin{aligned}
&\mathbf{x}^{(t+1)} = \arg\min \|\mathbf{x}^{(t)} - \eta_t \mathbf{v}^{(t)} - \mathbf{x}\| \\
&\mathbf{x}^{(t+1)} = \arg\min \|\mathbf{x}^{(t)} - \eta_t \mathbf{v}^{(t)} - \mathbf{x}\|^2 \\
&= \arg\min \|\mathbf{x}^{(t)} - \mathbf{x}\|^2 - 2\eta_t \langle \mathbf{v}^{(t)}, \mathbf{x} - \mathbf{x}^{(t)} \rangle + \eta_t^2 \|\mathbf{v}^{(t)}\|^2 \\
&= \arg\min \frac{\|\mathbf{x}^{(t)} - \mathbf{x}\|^2}{2\eta_t} + \langle \mathbf{v}^{(t)}, \mathbf{x} - \mathbf{x}^{(t)} \rangle \\
&= \arg\min f(\mathbf{x}^{(t)}) + \frac{1}{2\eta_t} \|\mathbf{x} - \mathbf{x}^{(t)}\|^2 + \langle \mathbf{v}^{(t)}, \mathbf{x} - \mathbf{x}^{(t)} \rangle
\end{aligned}
$$

درواقع می خواهیم $\mathbf{x}$ را در نزدیکی $\mathbf{x}^{(t)}$ انتخاب نماییم به طوری که در مجموعه ی $\mathcal{C}$ نیز قرار داشته باشد. همچنین باعث مینیمم شدن hyperplane support تابع در نزدیکی نقاط $\mathbf{x}^{(t)}$ نیز خواهد شد که ضریب regularization آن $\frac{1}{\eta_t}$ می باشد.

## پرسش تئوری ۱۵)

با توجه به پرسش تئوری ۵ داریم:

$$
\frac{1}{2}||\mathbf{x}^{(t+1)} - \mathbf{x}^*||_2^2 \leq \frac{1}{2}||\mathbf{x}^{(t)} - \mathbf{x}^*||_2^2 - \eta_t(f(\mathbf{x}^{(t)}) - f(\mathbf{x}^*)) + \frac{1}{2}\eta_t^2||\mathbf{v}^{(t)}||_2^2
$$

این عبارت برای این الگوریتم نیز جواب می دهد. پس لازم است طرفین را تقسیم بر $\eta_t$ کنیم، پس:

$$
(f(\mathbf{x}^{(t)}) - f(\mathbf{x}^*)) \leq \frac{1}{2\eta_t}||\mathbf{x}^{(t)} - \mathbf{x}^*||_2^2 - \frac{1}{2\eta_t}||\mathbf{x}^{(t+1)} - \mathbf{x}^*||_2^2 + \frac{1}{2}\eta_t||\mathbf{v}^{(t)}||_2^2
$$

با جمع این نامساوی ها داریم:

$$
\sum_{t=1}^{T}(f(\mathbf{x}^{(t)}) - f(\mathbf{x}^*)) \leq \frac{1}{2\eta_1}||\mathbf{x}^{(1)} - \mathbf{x}^*||_2^2 + \sum_{t=2}^{T}((\frac{1}{2\eta_t} - \frac{1}{2\eta_{t-1}})||\mathbf{x}^{(t)} - \mathbf{x}^*||_2^2) + \sum_{t=1}^{T}(\frac{1}{2}\eta_t||\mathbf{v}^{(t)}||_2^2)
$$

با توجه به فرض ۴.۱ گفته شده در صورت سوال و پرسش تئوری ۴ و همچنین ناصعودی بودن $\eta_t$ و در نتیجه ی این ها با مثبت ماندن این ضرایب می توان دو فرض گفته شده را اعمال کرد. پس summation اول به صورت تلسکوپی خواهد بود و به $\frac{1}{2\eta_t}B^2$ خواهیم رسید. summation دوم نیز به جای هرکدام از $||\mathbf{v}^{(t)}||_2^2$ ها $\rho^2$ گذاشته و بدین صورت مسئله اثبات خواهد شد.

## پرسش تئوری ۱۶ )

باتوجه به محدب بودن داریم:

$$f(\bar{\mathbf{x}}_T) \leq \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}^{(t)})$$

پس باتوجه به نتیجه سوال قبل و تنها با تقسیم طرفین بر T ، خواهیم داشت:

$$f(\bar{\mathbf{x}}_T) - f(\mathbf{x}^*) \leq \frac{B^2}{2T\eta_t} + \frac{\rho^2}{2T}\sum_{t=1}^{T}\eta_t$$

با جایگذاری مقادیر داده شده می توان گفت:

$$\frac{B^2}{2T\eta_t} = \frac{B^2}{2\alpha\sqrt{T}} \ , \ \sum_{t=1}^{T}\eta_t = \alpha\sum_{t=1}^{T}\frac{1}{\sqrt{t}}$$

پس:

$$\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq \int_0^T \frac{1}{t}dt \leq 2\sqrt{T}$$

با جایگذاری نامساوی فوق می توان به همان رابطه سوال رسید.

$$f(\bar{\mathbf{x}}_T) - f(\mathbf{x}^*) \leq \frac{B^2}{2\alpha\sqrt{T}} + \frac{\rho^2\alpha}{2\sqrt{T}}$$

حال با برابر صفر گذاشتن مشتق عبارت فوق داریم:

$$\frac{d}{d\alpha}(\frac{B^2}{2\alpha\sqrt{T}} + \frac{\rho^2\alpha}{2\sqrt{T}}) = 0 \rightarrow \alpha_{\text{best}} = \frac{B}{\rho}$$

پس بهترین کران بالا برابر $\frac{B\rho}{\sqrt{T}}$ می باشد.

## پرسش تئوری ۱۷

منشا آن بردار $\mathbf{V}^{(t)}$ است که در update $\mathbf{X}^{(t)}$ استفاده می شود که خود این بردار تصادفی است. پس با update کردن، نتیجه تصادفی خواهد شد.

## پرسش تئوری ۱۸

می دانیم که متغیر تصادفی $\mathbf{V}^{(t)}$ از یک توزیع پیروی می کند. اما می دانیم احتمال زیرگرادیان بودن بردار $\mathbf{V}^{(t)}$ در نقاط مختلف متفاوت می باشد. پس باید از توزیع شرطی بدست آمده در هر نقطه استفاده کنیم. برای تعریف بردار زیرگرادیان تصادفی در یک نقطه می بایست میانگین از توزیع شرطی $\mathbb{E}[\mathbf{V}^{(t)}|\mathbf{X}^{(t)}]$ گرفته شود و این گونه می توان گفت که پس از مشاهده ی $\mathbf{X}^{(t)}$ باید تصمیم بگیریم که $\mathbf{V}^{(t)}$ از کدام تابع پیروی خواهد کرد.

## پرسش تئوری ۱۹

$$
\begin{aligned}
&\frac{1}{2}\|\mathbf{X}^{(t+1)} - \mathbf{x}^*\|^2 = \frac{1}{2}\|\Pi_{\mathcal{C}}(\mathbf{X}^{(t)} - \eta_t \mathbf{V}^{(t)}) - \mathbf{x}^*\|^2 \leq \|\mathbf{X}^{(t)} - \eta_t \mathbf{V}^{(t)} - \mathbf{x}^*\|^2 \\
&= \frac{1}{2}\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \eta_t\langle \mathbf{V}^{(t)}, \mathbf{X}^{(t)} - \mathbf{x}^*\rangle + \frac{\eta_t^2}{2}\|\mathbf{V}^{(t)}\|^2 \\
&= \frac{1}{2}\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \eta_t\langle \mathbf{V}^{(t)} - \mathbb{E}[\mathbf{V}^{(t)}|\mathbf{X}^{(t)}], \mathbf{X}^{(t)} - \mathbf{x}^*\rangle - \eta_t\langle \mathbb{E}[\mathbf{V}^{(t)}|\mathbf{X}^{(t)}], \mathbf{X}^{(t)} - \mathbf{x}^*\rangle + \frac{\eta_t^2}{2}\|\mathbf{V}^{(t)}\| \\
&\leq \frac{1}{2}\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \eta_t\langle \boldsymbol{\xi}_t, \mathbf{X}^{(t)} - \mathbf{x}^*\rangle - \eta_t(f(\mathbf{X}^{(t)}) - f(\mathbf{x}^*)) + \frac{\eta_t^2}{2}\|\mathbf{V}^{(t)}\|^2
\end{aligned}
$$

## پرسش تئوری ۲۰)

$$
\begin{aligned}
&\mathbb{E}[\langle \boldsymbol{\xi}_t, \mathbf{X}^{(t)} - \mathbf{x}^* \rangle] = \mathbb{E}[\langle \mathbf{V}^{(t)} - \mathbb{E}[\mathbf{V}^{(t)}|\mathbf{X}^{(t)}], \mathbf{X}^{(t)} - \mathbf{x}^* \rangle] \\
&\langle \mathbb{E}[\mathbf{V}^{(t)} - \mathbb{E}[\mathbf{V}^{(t)}|\mathbf{X}^{(t)}]], \mathbf{X}^{(t)} - \mathbf{x}^* \rangle] = \langle \mathbb{E}[\mathbf{V}^{(t)}] - \mathbb{E}[\mathbb{E}[\mathbf{V}^{(t)}|\mathbf{X}^{(t)}]], \mathbf{X}^{(t)} - \mathbf{x}^* \rangle \\
&\langle \mathbb{E}[\mathbf{V}^{(t)}] - \mathbb{E}[\mathbf{V}^{(t)}], \mathbf{X}^{(t)} - \mathbf{x}^* \rangle = 0
\end{aligned}
$$

## پرسش تئوری ۲۱)

همانند پرسش تئوری ۱۵ و باتوجه به محدب بودن تابع که در پرسش ۱۶ داریم، می توان گفت:

$$
f(\bar{\mathbf{x}}_T) - f(\mathbf{x}^*) \leq \frac{B^2}{2T\eta_t} + \frac{\rho^2}{2T}\sum_{t=1}^{T}\eta_t - \frac{1}{T}\sum_{t=1}^{T}\eta_t \left\langle \boldsymbol{\xi}_t, \mathbf{X}^{(t)} - \mathbf{x}^* \right\rangle
$$

باتوجه به اثبات پرسش تئوری ۲۰ داریم:

$$
\mathbb{E}[\langle \boldsymbol{\xi}_t, \mathbf{X}^{(t)} - \mathbf{x}^* \rangle] = 0
$$

پس با محاسبه ی امید ریاضی طرفین نامساوی فوق، ترم هایی که ثابت هستند باقی مانده و باقی ترم ها به صورت امید ریاضی متغیر خواهند شد. از طرفی به دلیل صفر بودن امید ریاضی آخرین ترم طرف راست نامساوی، خواسته ی مسئله اثبات خواهد شد.

$$
\mathbb{E}[f(\bar{\mathbf{x}}_T)] - f(\mathbf{x}^*) \leq \frac{B^2}{2T\eta_t} + \frac{\rho^2}{2T}\sum_{t=1}^{T}\eta_t
$$

## پرسش تئوری ۲۲)

باتوجه به پرسش تئوری ۲۱ داریم:

$$\mathbb{E}[f(\bar{\mathbf{x}}_T)] - f(\mathbf{x}^*) \leq \frac{B^2}{2T\eta_t} + \frac{\rho^2}{2T}\sum_{t=1}^{T}\eta_t$$

اکنون با جایگذاری $\eta_t = \frac{B}{\rho\sqrt{t}}$ می توان گفت:

$$\mathbb{E}[f(\bar{\mathbf{x}}_T)] - f(\mathbf{x}^*) \leq \frac{B\rho}{2\sqrt{T}} + \frac{B\rho}{2\sqrt{T}} = \frac{B\rho}{\sqrt{T}}$$

## پرسش تئوری ۲۳)

باتوجه به قویا محدب بودن تابع می توان گفت:

$$\langle \mathbf{X}^{(t)} - \mathbf{x}^*, \, \mathbb{E}[\mathbf{V}^{(t)} | \mathbf{X}^{(t)}] \rangle \geq f(\mathbf{X}^{(t)}) - f(\mathbf{x}^*) + \frac{\lambda}{2} \|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2$$

همچنین می دانیم:

$$\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \|\mathbf{X}^{(t+1)} - \mathbf{x}^*\|^2 = \|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \|\Pi_{\mathcal{C}}(\mathbf{X}^{(t)} - \eta_t \mathbf{V}^{(t)}) - \mathbf{x}^*\|^2$$

$$\geq \|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \|\mathbf{X}^{(t)} - \eta_t \mathbf{V}^{(t)} - \mathbf{x}^*\|^2 = 2\eta_t \left\langle \mathbf{X}^{(t)} - \mathbf{x}^*, \mathbf{V}^{(t)} \right\rangle - \eta_t^2 \|\mathbf{V}^{(t)}\|^2$$

با امیدریاضی گرفتن از طرفین نامعادله ی فوق و تقسیم بر $2\eta_t$ کردن و با درنظر گرفتن $\mathbb{E}\left[\|\mathbf{V}^{(t)}\|^2\right] \leq \rho^2$ می توان گفت:

$$\langle \mathbf{X}^{(t)} - \mathbf{x}^*, \, \mathbb{E}[\mathbf{V}^{(t)} | \mathbf{X}^{(t)}] \rangle \leq \frac{\mathbb{E}[\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \|\mathbf{X}^{(t+1)} - \mathbf{x}^*\|^2]}{2\eta_t} + \frac{\eta_t}{2}\rho^2$$

با ادغام نامساوی فوق و امیدریاضی نامساوی اول کار می توان گفت:

$$\mathbb{E}[f(\mathbf{X}^{(t)})] - f(\mathbf{x}^*) + \frac{\lambda}{2}\mathbb{E}[\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2] \leq \frac{\mathbb{E}[\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - \|\mathbf{X}^{(t+1)} - \mathbf{x}^*\|^2]}{2\eta_t} + \frac{\eta_t}{2}\rho^2$$

اگر $\eta_t = \frac{1}{\lambda t}$ را در عبارت فوق بصورت زیر جایگذاری کنیم، داریم:

$$\mathbb{E}[f(\mathbf{X}^{(t)})] - f(\mathbf{x}^*) \leq \frac{\lambda}{2}\mathbb{E}[t\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - (t+1)\|\mathbf{X}^{(t+1)} - \mathbf{x}^*\|^2] + \frac{\rho^2}{2\lambda}\frac{1}{t}$$

حال از طرفین summation میگیریم.

$$\sum_{t=1}^{T}( \mathbb{E}[f(\mathbf{X}^{(t)})] - f(\mathbf{x}^*)) \leq \frac{\lambda}{2}\mathbb{E}[\sum_{t=1}^{T}(t\|\mathbf{X}^{(t)} - \mathbf{x}^*\|^2 - (t+1)\|\mathbf{X}^{(t+1)} - \mathbf{x}^*\|^2)] + \frac{\rho^2}{2\lambda}\sum_{t=1}^{T}\frac{1}{t}$$

$$= \frac{\lambda}{2}\mathbb{E}[-T\|\mathbf{X}^{(T+1)} - \mathbf{x}^*\|^2] + \frac{\rho^2}{2\lambda}\sum_{t=1}^{T}\frac{1}{t}$$

$$= \frac{-\lambda T}{2}\|\mathbf{X}^{(T+1)} - \mathbf{x}^*\|^2 + \frac{\rho^2}{2\lambda}\sum_{t=1}^{T}\frac{1}{t} \leq \frac{\rho^2}{2\lambda}\sum_{t=1}^{T}\frac{1}{t}$$

اکنون دو تابع زیر را در نظر بگیرید.

$$h(t) = \ln(t+1) - \ln(t) \quad , \quad k(t) = \frac{1}{t+1}$$

$$\Rightarrow h(1) = \ln(2) \geq \frac{1}{2} = k(1)$$

$$\frac{\partial}{\partial t}h(t) = -\frac{1}{t(t+1)} \geq -\frac{1}{(t+1)^2} = \frac{\partial}{\partial t}k(t)$$

$$\Rightarrow t > 1 : h(t) > k(t)$$

پس می توان نوشت:

$$1 + k(1) \leq 1 + h(1)$$

$$1 + k(1) + k(2) \leq 1 + h(1) + h(2)$$

$$\sum_{t=1}^{T} \frac{1}{t} \leq 1 + \ln(T)$$

با استفاده رابطه ی فوق و رابطه ی صفحه ی قبل، می توان گفت:

$$\sum_{t=1}^{T} ( \mathbb{E}[f(\mathbf{X}^{(t)})] - f(\mathbf{x}^*)) \leq \frac{\rho^2}{2\lambda} \sum_{t=1}^{T} \frac{1}{t} \leq \frac{\rho^2}{2\lambda}(1 + \ln(T))$$

اکنون با تقسیم طرفین بر $T$ خواهیم داشت:

$$\mathbb{E}[f(\bar{\mathbf{X}}_T)] - f(\mathbf{x}^*) \leq \mathbb{E}[\frac{1}{T} \sum_{t=1}^{T} f(\mathbf{X}^{(t)})] - f(\mathbf{x}^*) \leq \frac{\rho^2}{2\lambda T}(1 + \ln(T))$$

$$\Rightarrow \mathbb{E}[f(\bar{\mathbf{X}}_T)] - f(\mathbf{x}^*) \leq \frac{\rho^2}{2\lambda} \frac{(1 + \ln(T))}{T}$$

## پرسش تئوری ۲۴)

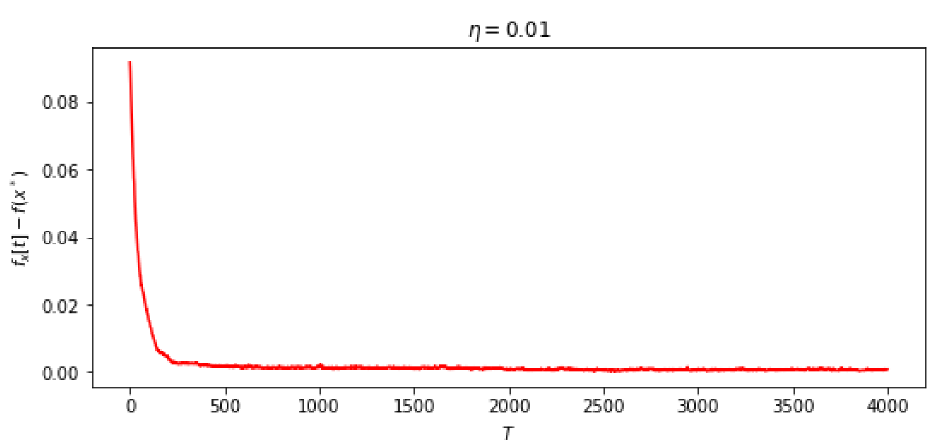
باتوجه به خطی بودن امید ریاضی می توان گفت:

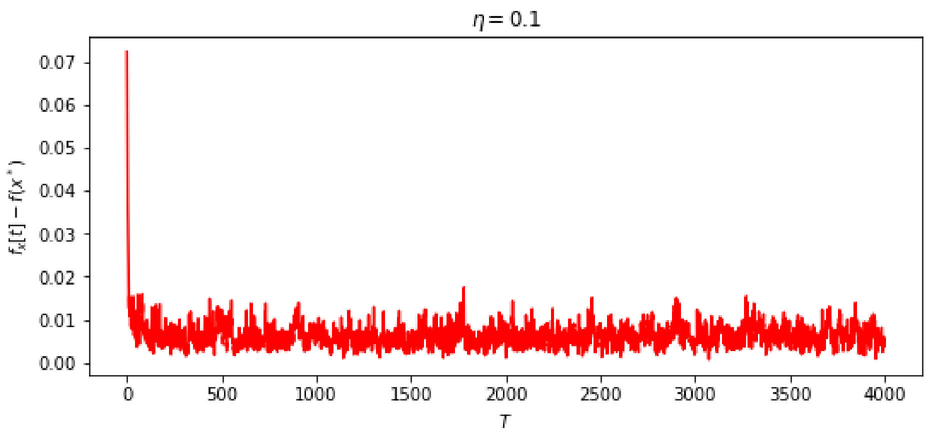
$$\mathbf{V} \in \partial_{\mathbf{x}} \mathbf{F}(\mathbf{x}, \mathbf{Z}) \Rightarrow \mathbb{E}[\mathbf{V}] \in \mathbb{E}_{\mathbf{z}}[\partial_{\mathbf{x}} \mathbf{F}(\mathbf{x}, \mathbf{Z})] = \partial\, \mathbb{E}_{\mathbf{z}}[\mathbf{F}(\mathbf{x}, \mathbf{Z})] = \partial\, f(\mathbf{x})$$

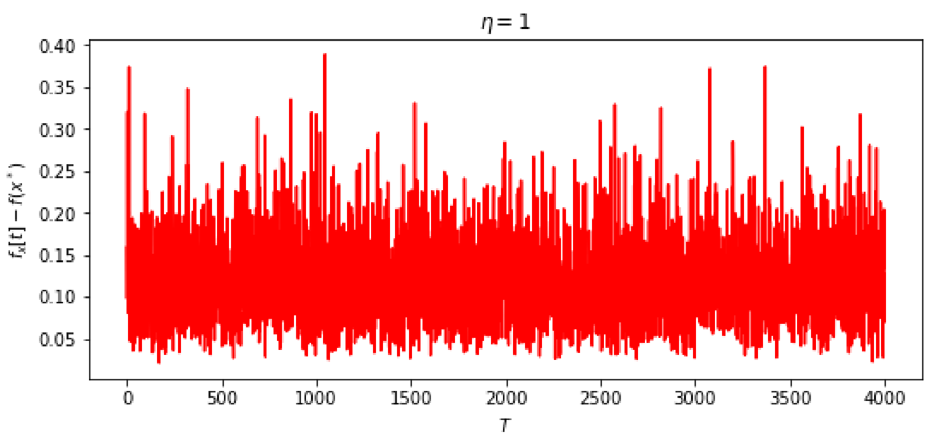
پس می توان گفت که $\mathbf{V}$ زیرگرادیان تصادفی برای $f(\mathbf{x})$ می باشد.

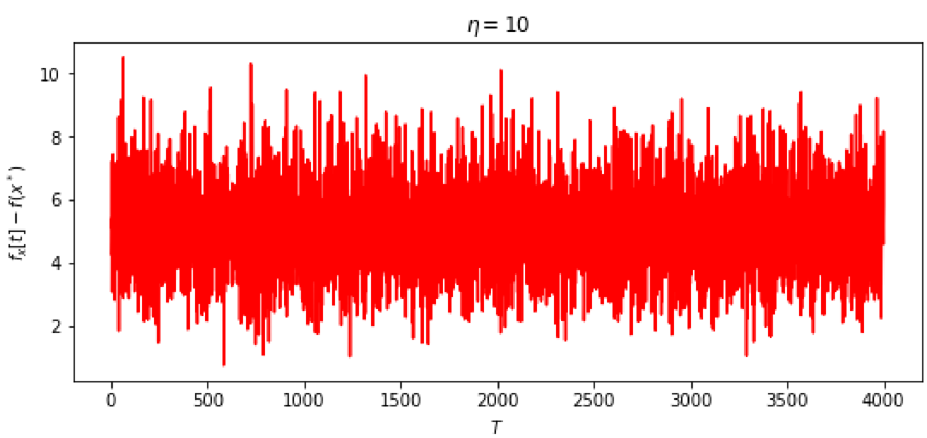
# پرسش شبیه سازی ۳)

```
fed random_subgrad(expected):
random_subgrad = np.zeros(nel(expected))
rof i ni egnar(nel(expected)):
random_subgrad[i] = np.random.normal(loc=expected[i], scale=0.13)
nruter random_subgrad

fed subgradient(x, Z, B):
m = nel(B)
subgrad = np.zeros(Z.shape[1])
rof i ni egnar(m):
fi (Z[i, :] @ x - B[i]) >= 0:
subgrad = subgrad + Z[i, :]
esle:
subgrad = subgrad - Z[i, :]
subgrad = subgrad/m
nruter subgrad

# gnitnemelpmi eht mhtirogla
eta = [0.01, 0.1, 1, 10]
convergence_var_3 = np.zeros((nel(eta),T))
elapsed_time_SGD = np.zeros((nel(eta), T))
rof e ni eta:
x_t = x_0
rof t ni np.arange(T):
begin = time.time()
expected = subgradient(x_t, Z, B)
v = random_subgrad(expected)
x_t = x_t.T - e * v
x_t = x_t.T
convergence_var_3[eta.index(e), t] = np.linalg.norm(Z @ x_t - B, dro=1) - np.linalg.norm(Z @ x_opt -
B, dro=1)
end = time.time()
elapsed_time_SGD[eta.index(e), t] = end - begin
# gnittolP
plt.figure(figsize=(8,4))
plt.plot(np.arange(1, T+1), convergence_var_3[eta.index(e),:], "r")
plt.xlabel(r'$T$')
plt.ylabel(r'x_f$[t] - f(x)*^$')
plt.title((r'$\ate = $' + rts(e)))
plt.tight_layout(pad = 2)
```



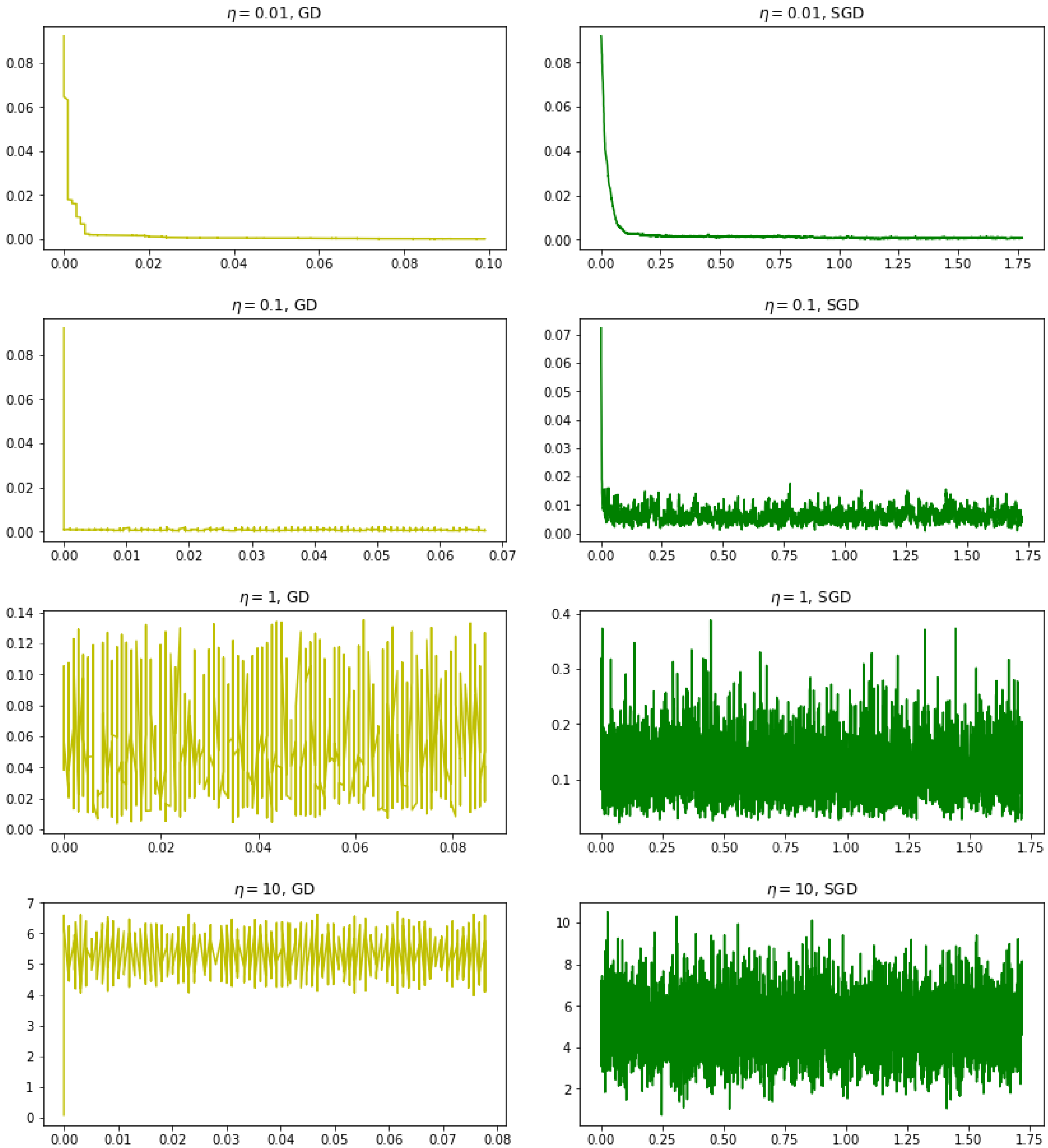

$\eta = 0.1$
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
$f_x[t] - f(x^*)$
0
500
1000
1500
2000
2500
3000
3500
4000
$T$


$\eta = 1$
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
$f_x[t] - f(x^*)$
0
500
1000
1500
2000
2500
3000
3500
4000
$T$


$\eta = 10$
10
8
6
4
2
$f_x[t] - f(x^*)$
0
500
1000
1500
2000
2500
3000
3500
4000
$T$

## پرسش شبیه سازی ۴)

```
# evitalumuc mus fo eht despale emit
time_GD = np.cumsum(elapsed_time_GD, axis=1)
time_SGD = np.cumsum(elapsed_time_SGD, axis=1)

# gnittolP
fig, axs = plt.subplots(nrows=4, ncols=2, figsize=(12, 13))
axs[0, 0].plot(time_GD[0, :], convergence_var[0, :], "y")
axs[0, 0].set_title(r'$\ate = 10.0$, DG')
axs[1, 0].plot(time_GD[1, :], convergence_var[1, :], "y")
axs[1, 0].set_title(r'$\ate = 1.0$, DG')
axs[2, 0].plot(time_GD[2, :], convergence_var[2, :], "y")
axs[2, 0].set_title(r'$\ate = 1$, DG')
axs[3, 0].plot(time_GD[3, :], convergence_var[3, :], "y")
axs[3, 0].set_title(r'$\ate = 01$, DG')
axs[0, 1].plot(time_SGD[0, :], convergence_var_3[0, :], "g")
axs[0, 1].set_title(r'$\ate = 10.0$, DGS')
axs[1, 1].plot(time_SGD[1, :], convergence_var_3[1, :], "g")
axs[1, 1].set_title(r'$\ate = 1.0$, DGS')
axs[2, 1].plot(time_SGD[2, :], convergence_var_3[2, :], "g")
axs[2, 1].set_title(r'$\ate = 1$, DGS')
axs[3, 1].plot(time_SGD[3, :], convergence_var_3[3, :], "g")
axs[3, 1].set_title(r'$\ate = 01$, DGS')
plt.tight_layout(pad = 2)
```


η = 0.01, GD
0.00 0.02 0.04 0.06 0.08
0.00 0.02 0.04 0.06 0.08 0.10
η = 0.01, SGD
0.00 0.02 0.04 0.06 0.08
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75
η = 0.1, GD
0.00 0.02 0.04 0.06 0.08
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07
η = 0.1, SGD
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75
η = 1, GD
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
0.00 0.02 0.04 0.06 0.08
η = 1, SGD
0.1 0.2 0.3 0.4
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75
η = 10, GD
0 1 2 3 4 5 6 7
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08
η = 10, SGD
2 4 6 8 10
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75

## پرسش شبیه سازی (۵ تا ۹)

```
######## noitalumis - noitseuq 5 - 9 ########
# tropmI deriuqer yrarbil
tropmi pandas as pd

# gnitropmi atad_esuoh
data = pd.read_csv('atad_esuoh.vsc')

##### gnittolp eht serugif

fig, axs = plt.subplots(nrows=2, ncols=2, figsize=(20, 13))

tmp = data.sort_values(by=['smoordeb', 'ecirp'])
axs[0, 0].plot(tmp.loc[:, "smoordeb"], tmp.loc[:, "ecirp"], "y")
axs[0, 0].set_title(r'smoordeb tceffe')

tmp = data.sort_values(by=['smoorhtab', 'ecirp'])
axs[1, 0].plot(tmp.loc[:, "smoorhtab"], tmp.loc[:, "ecirp"], "r")
axs[1, 0].set_title(r'smoorhtab tceffe')

tmp = data.sort_values(by=['ega', 'ecirp'])

axs[0, 1].plot(tmp.loc[:, "ega"], tmp.loc[:, "ecirp"], "g")
axs[0, 1].set_title(r'ega tceffe')

means_of_price = np.zeros(5)
Cities = ["narheT", "nahefsE", "dzaY", "zarihS", "zirbaT"]
rof city ni Cities:
t = (data.loc[:, city] == 1)
means_of_price[Cities.index(city)] = mus(data[t].price)/nel(t)

axs[1, 1].plot(np.arange(1, 6), means_of_price, "b")
axs[1, 1].set_xlabel(r'narheT             nahefsE                    dzaY                    zarihS
               zirbaT', fontsize=18)

plt.tight_layout(pad = 2)
```

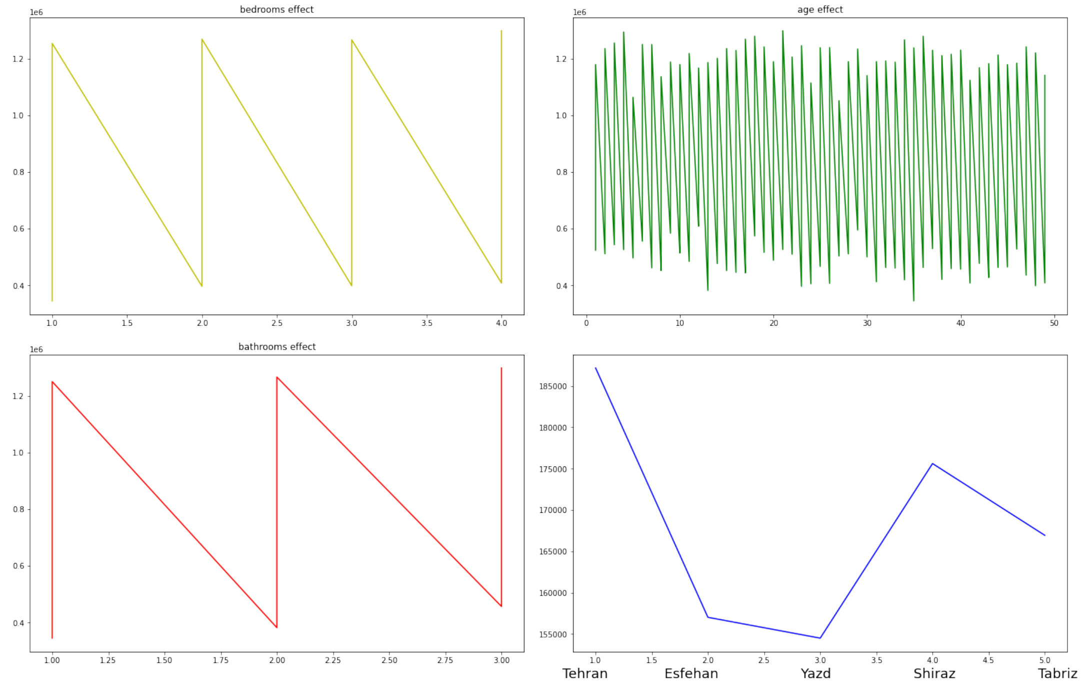
```
## tseT dna niarT gnittilps dna gnizilamron atad
Train_data = data.loc[0 : np.floor(data.shape[0] * 0.75) - 1, :]
Test_data = data.loc[np.floor(data.shape[0] * 0.75) : , :]
Train_data = (Train_data - Train_data.mean())/Train_data.std()
Test_data = (Test_data - Test_data.mean())/Test_data.std()
```

```
## ESM noitcnuf :
fed MSE_calc(data_1, data_2):
n = nel(data_1)
data_1 = np.reshape(data_1, (n, 1))
data_2 = np.reshape(data_2, (n, 1))
MSE = np.mus((data_1 - data_2)**2)/n
nruter MSE
```

```
fed subgradient(w, b, X, y):
subgrad_w = np.zeros((nel(w), 1))
subgrad_b = 0
n = nel(y)
m = nel(w)
b_tmp = np.reshape(np.array([b]*n), (n, 1))
y = np.reshape(y, (n, 1))
```

```
w = np.reshape(w, (m, 1))
tmp_var = X@w + b_tmp - y
subgrad_w = (2/n)*X.T@(tmp_var)
subgrad_b = (2/n)*np.mus(tmp_var)
nruter subgrad_w, subgrad_b


### en0 egats fo DGS mhtirogla
fed SGD_One_Stage(w, b, batch_of_data, eta):
p, q = batch_of_data.shape
w = w.reshape(9, 1)
subgrad_w, subgrad_b = subgradient(w, b, batch_of_data[:, 0: q - 1], batch_of_data[:, q - 1])
w_new = w - eta * subgrad_w
b_new = b - eta * subgrad_b
nruter w_new, b_new

```

```
#### gnitamitse (W, b) rof elpitlum sehctab####
epoch = 100
batch = 25 # %02 fo niart tes ezis
T = 4000
eta = 0.01
w = np.random.rand(9)
b = np.random.rand()
mse_train = np.zeros(100)
mse_test = np.zeros(100)
## nur eht mhtirogla
rof j ni np.arange(epoch):
Train_data = Train_data.sample(frac = 1)
rof k ni np.arange(30):
batch_of_data = Train_data.iloc[k*batch : k*batch + 25, :].to_numpy()
rof i ni np.arange(T):
w, b = SGD_One_Stage(w, b, batch_of_data, eta)
b_reshape_1 = np.reshape(np.array([b]*750), (750, 1))
train_estimate =  np.reshape(Train_data.iloc[:, 0:9].to_numpy(), (750, nel(w))) @ np.reshape(w, (9,
 1)) + b_reshape_1
b_reshape_2 = np.reshape(np.array([b]*250), (250, 1))
test_estimate = np.reshape(Test_data.iloc[:, 0:9].to_numpy(), (250, nel(w))) @ np.reshape(w, (9, 1))
 + b_reshape_2
mse_train[j] = MSE_calc(Train_data.iloc[:, 9].to_numpy(), train_estimate)
mse_test[j] = MSE_calc(Test_data.iloc[:, 9].to_numpy(), test_estimate)

```




همانطور که مشاهده میکنید، تعداد سرویس های بهداشتی و اتاق خواب ها بر روی قیمت خانه صعودی است.  در حالی که تاثیر سال ساخت برروی قیمت خانه روند مشخصی ندارد و به طور میانگین ثابت است. درباره شهر ها نیز همانطور که میبینید، قیمت شهر ها به ترتیب در تهران، شیراز، تبریز، اصفهان و یزد نزولی است.