-1

الف) داریم:

$$\nabla f(x_t) = Ax_t - b$$

$$\Longrightarrow x_{t+1} = x_t - \eta_t A x_t + \eta_t b = (I - \eta_t A)x_t + \eta_t b$$

2.

ابتدا توجه کنید که چون x^* بهینه است، داریم: $\nabla f(x^*) = 0$ پس $Ax^* - b = 0$

پس داریم: $x^* = x^* - \eta_t(Ax^* - b)$ پس می‌توان نوشت:

$$x_{t+1} - x^* = (x_t - \eta_t(Ax_t - b)) - (x^* - \eta_t(Ax^* - b))$$

$$= x_t - \eta_t A x_t - x^* + \eta_t A x^* = (I - \eta_t A)(x_t - x^*)$$

$$\Longrightarrow x_{t+1} - x^* = \left(\prod_{k=1}^{t}(I - \eta_k A)\right)(x_1 - x^*)$$

3.

می‌دانیم رابطه $\|Ax\| \le \|A\|_2 \|x\|$ برقرار است. پس داریم:

$$\|x_{t+1} - x^*\| = \left\|\left(\prod_{k=1}^{t}(I - \eta_k A)\right)(x_1 - x^*)\right\| \le \left\|\prod_{k=1}^{t}(I - \eta_k A)\right\|_2 \|x_1 - x^*\|$$

همچنین می‌دانیم که $\left\|\prod_{k=1}^{n} A_k\right\|_2 \le \prod_{k=1}^{n} \|A_k\|$ پس داریم:

$$\prod_{k=1}^{t} \|I - \eta_k A\| \|x_1 - x^*\| \le \prod_{k=1}^{t} \|I - \eta_k A\| \|x_1 - x^*\|$$

.4

چون f قویا محدب است، داریم:

$$m \text{ قویا محدب} \iff \nabla^2 f(x) \geq mI \iff A \geq mI$$

$$M \text{ هموار است} \iff \nabla^2 f(x) \leq MI \iff A \leq MI$$

$$\iff mI \leq A \leq MI \iff \frac{m}{M} I \leq \frac{1}{M} A \leq I$$

$$\Rightarrow \|I - \frac{1}{\beta} A\|_2 = 1 - \frac{\alpha}{\beta} = 1 - \frac{1}{K}$$

.5

با توجه به بخش ۳ داریم:

$$\|x_{t+1} - x^*\| \leq \prod_{K=1}^{t} \|I - \frac{1}{M} A\|_2 \|x_1 - x^*\| \leq \left(1 - \frac{1}{K}\right)^t \|x_1 - x^*\| \leq e^{-\frac{t}{K}} \|x_1 - x^*\|$$

-2

1. با توجه به m قویاً" محدب بودن داریم:

$$f(y) - f(x) \geq \langle \nabla f(x), y-x \rangle + \frac{m}{2}\|y-x\|^2$$

$$\Rightarrow \min_y f(y) - f(x) \geq \min_y \langle \nabla f(x), y-x \rangle + \frac{m}{2}\|y-x\|^2$$

$$\Rightarrow \nabla f(x) + my - mx = 0 \Rightarrow y = x - \frac{1}{m}\nabla f(x)$$

$$\Rightarrow \min_y \langle \nabla f(x), y-x \rangle + \frac{m}{2}\|y-x\|^2 = -\frac{1}{m}\|\nabla f(x)\|^2 + \frac{1}{2m}\|\nabla f(x)\|^2$$

$$= -\frac{1}{2m}\|\nabla f(x)\|^2 \Rightarrow f(x^*) - f(x) \geq -\frac{1}{2m}\|\nabla f(x)\|^2$$

$$\Rightarrow f(x) - f(x^*) \leq \frac{1}{2m}\|\nabla f(x)\|^2$$

.2

$$x_{t+1} = x_t - \nabla^2 f(x)^{-1} \nabla f(x) \qquad \Delta x_t = -\nabla^2 f(x_t)^{-1} \nabla f(x_t)$$

$$\|\nabla f(x_{t+1})\|_2 = \|\nabla f(x_t + \Delta x_t) - \nabla f(x_t) - \nabla^2 f(x_t) \Delta x_t\|^2$$

$$= \left\| \int_0^1 \left(\nabla^2 f(x_t + \beta \Delta x_t) - \nabla^2 f(x_t)\right) \Delta x_t \, d\beta \right\| \leq \frac{M}{2}\|\Delta x_t\|^2$$

$$= \frac{M}{2}\|\nabla^2 f(x_t)^{-1} \nabla f(x_t)\|_2^2 \leq \frac{M}{2}\|\nabla^2 f(x_t)\|_2^2 \|\nabla f(x_t)\|^2$$

$$\leq \left(\frac{M}{2m^2}\right)^2 \|\nabla f(x_t)\|^2$$

3.

$$f(x_K) - f(x_*) \le \frac{1}{2m}\|\nabla f(x_K)\|_2^2 \le \frac{1}{2m}\left(\frac{M}{2m^2}\right)^2 \|\nabla f(x_{K-1})\|^4$$

$$\le \frac{1}{2m}\,\frac{M}{2m^2}\,\|\nabla f(x_{K_0})\|_2^2 \le \frac{2m^3}{M^2}\left(\frac{1}{2}\right)^{K-K_0+1}$$

3-

1. فرض کنید f m قویا محدب باشد، حال فرض کنید x_1 و x_2 دو ساب‌گرادیان دلخواه از تابع f^* باشند در نقاط g_1 و g_2. پس داریم:

$$x_1 \in \partial f^*(g_1) \Longrightarrow g_1 \in \partial f(x_1)$$
$$x_2 \in \partial f^*(g_2) \Longrightarrow g_2 \in \partial f(x_2)$$

$$\xrightarrow{m \text{ قویا محدب}} \|g_2 - g_1\| \ge m\|x_2 - x_1\| \Longrightarrow \|x_2 - x_1\| \le \frac{1}{m}\|g_2 - g_1\|$$

$\Longrightarrow$ f^* $\frac{1}{m}$ لیپشیتز است

حال فرض کنید f^* $\frac{1}{m}$ - لیپشیتز باشد، g_1 و g_2 دو نقطه از f^* و x_1 و x_2 ساب‌گرادیان متناظر آن ها باشند. داریم:

$$x_1 \in \partial f^*(g_1) \Longrightarrow g_1 \in \partial f(x_1)$$
$$x_2 \in \partial f^*(g_2) \Longrightarrow g_2 \in \partial f(x_1)$$

$$\Longrightarrow \|x_1 - x_2\| \le \frac{1}{m}\|g_1 - g_2\| \Longrightarrow \|g_2 - g_1\| \ge m\|x_2 - x_1\|$$

$\Longrightarrow$ f m قویا محدب است

.2

در این الگوریتم ابتدا تابع لاگرانژین را بر حسب x کمینه و سپس بر حسب u بیشینه می‌کنیم که
این استراتژی برگرفته از مسئله دوگان است. داریم:

$$\min_{x} f(x) \quad \text{s.t. } Ax=b \;=\; \max_{u} \min_{x} f(x) + u^T(Ax-b)$$

$\Longrightarrow$ مرحله اول: کمینه نسبت به x $\longrightarrow$ $x^k = \underset{x}{\operatorname{argmin}}\, f(x) + u^T(Ax-b)$

مرحله دوم: بیشینه نسبت به u با حرکت به سمت گرادیان

$$u^{k+1} = u^k + t_k \nabla_u \left(f(x) + u^T(Ax^k - b)\right)$$

$$= u^k + t_k(Ax^k - b)$$

-4

.1

داریم:

$$\begin{cases} v + A^T w = -\nabla f(x) \\ Av = 0 \end{cases}$$

$$\Rightarrow Av + AA^T w = -A\nabla f(x) \Rightarrow AA^T w = -A\nabla f(x)$$

$$\Rightarrow w = -(AA^T)^{-1} A \nabla f(x) \Rightarrow w = \underset{y}{\operatorname{argmin}} \|A^T y + \nabla f(x)\|^2$$

$\Rightarrow A^T w$ تصویر $\nabla f(x)$ روی فضای سطری A است $\Rightarrow -\nabla f(x) - A^T w \rightarrow -\nabla f(x)$ تصویر روی فضای پوچی A

$$\Rightarrow v = -\nabla f(x) - A^T w \Rightarrow v = \underset{Au=0}{\operatorname{argmin}} \|-\nabla f(x) - u\|^2$$

2. یکی هستند. زیرا:

$$\Delta x_{pg} = \underset{Au=0}{\operatorname{argmin}} \|-f(x) - u\|_2 = -F^T \nabla f(x)$$

$$-\nabla_z \tilde{f}(z) = -\nabla_z f(Fz + \hat{x}) = -F^T \nabla_x f(x)$$

.3

اگر $F^T F = I$ باشد، و شرایط مطرح شده در قسمت 2 برقرار باشد، دیدیم که گام های این الگوریتم با گام های الگوریتم گرادیان کاهشی یکی هستند پس اگر مجموعه $\{x \mid f(x) \le f(x_0)\}$ بسته باشد و یا $f(Fz + \hat{x})$ کانوکس باشد، الگوریتم مانند گرادیان کاهشی همگرا می‌شود