IE8534 HW3

Jiacheng Yuan

5288191

1. (a)

$$RHS = f^* + \frac{1}{2t_k}\left(\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2\right) + (e^k)^T(x^{k+1} - x^k) + (e^k)^T(x^k - x^*)$$

$$= f^* + \frac{1}{2t_k}\left(\|x^k\|^2 - \|x^{k+1}\|^2 - 2x^k x^* + 2x^{k+1}x^*\right) + (e^k)^T(x^{k+1} - x^*)$$

$\because$ $e^k = \nabla f(x^k) + \frac{1}{t_k}(x^{k+1} - x^k)$

$$\therefore RHS = f^* + \frac{1}{2t_k}\left(\|x^k\|^2 - \|x^{k+1}\|^2 - 2x^{kT}x^* + 2x^{k+1T}x^*\right) + \nabla f(x^k)^T(x^{k+1} - x^*) + \frac{1}{t_k}(x^{k+1} - x^k)(x^{k+1} - x^*$$

$$= f^* + \frac{1}{2t_k}\|x^k\|^2 + \frac{1}{2t_k}\|x^{k+1}\|^2 - \frac{1}{t_k}x^{kT}x^{k+1} + \nabla f(x^k)^T(x^{k+1} - x^*)$$

$$= f^* + \nabla f(x^k)^T(x^{k+1} - x^*) + \frac{1}{2t_k}\|x^{k+1} - x^k\|^2$$

$$LHS = f(x^{k+1})$$

$$\le f(x^k) + \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{L}{2}\|x^{k+1} - x^k\|^2$$

$\because$ $f(x)$ is convex

$\therefore$ $f(x^*) \ge f(x^k) + \nabla f(x^k)^T(x^* - x^k)$

$\therefore$ $f(x^k) \le f^* + \nabla f(x^k)^T(x^k - x^*)$

$$\therefore LHS \le f^* + \nabla f(x^k)^T(x^k - x^*) + \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{L}{2}\|x^{k+1} - x^k\|^2$$

$$= f^* + \nabla f(x^k)^T(x^{k+1} - x^*) + \frac{L}{2}\|x^{k+1} - x^k\|^2$$

$\because$ $t_k \le \frac{1}{L}$

$\therefore$ $L \le \frac{1}{t_k}$

$$\therefore LHS \le f^* + \nabla f(x^k)^T(x^{k+1} - x^*) + \frac{1}{2t_k}\|x^{k+1} - x^k\|^2 = RHS$$

(b) multiply both sides of (a) with t_k

$$t_k(f(x^{k+1})-f^*) \le \frac{1}{2}(\|x^k-x^*\|^2-\|x^{k+1}-x^*\|^2)-t_k^2(e^k)^T\tilde{g}^k+t_k(e^k)^T(x^k-x^*)$$

$$\sum_{i=0}^{k} t_i(f(x^{k+1})-f^*) \le \frac{1}{2}\sum_{i=0}^{k}(\|x^i-x^*\|^2-\|x^{i+1}-x^*\|^2)-\sum_{i=0}^{k}\left[t_i(e^i)^T\tilde{g}^i+t_i(e^i)^T(x^i-x^*)\right]$$

$$\le \frac{1}{2}\|x^0-x^*\|^2-\sum_{i=0}^{k}\left[t_i^2(e^i)^T\tilde{g}^i+t_i(e^i)^T(x^i-x^*)\right]$$

$$LHS \ge k\cdot t_{min}\left(\min_{0\le i\le k} f(x^i)-f^*\right)$$

$$\therefore \min_{0\le i\le k} f(x^i)-f^* \le \frac{1}{kt_{min}}\cdot\frac{D^2}{2}-\frac{1}{kt_{min}}\sum_{i=0}^{k}\left[t_i^2(e^i)^T\tilde{g}^i+t_i(e^i)^T(x^i-x^*)\right]$$

§ 2 (a) $\text{prox}_h(a) = \underset{u}{\text{argmin}}\left(h(u) + \frac{1}{2}\|u-a\|_2^2\right) = \underset{u}{\text{argmin}}\left(\lambda\|u\|_0 + \frac{1}{2}\|u-a\|_2^2\right)$

$0 \in \lambda\|u^*\|_0 + \frac{1}{2}\|u^*-a\|_2^2$

$\Rightarrow 0 \in \lambda\partial\|u^*\|_0 + u^* - a$

$\partial\|u\|_0 = \left\{ v \in \mathbb{R}^n \,\middle|\, v_i \begin{cases} =0 & v_i \neq 0 \\ \in \mathbb{R} & v_i = 0 \end{cases} \right.$

$\therefore$ if $u_i \neq 0$ then

$u_i^* = a_i$

if $u_i = 0$ then

$0 \in \mathbb{R} + 0 - a_i$

$\Rightarrow a_i \in \mathbb{R}$

$\therefore \text{prox}_h(a)_i = \{0, a_i\}$

(b) $0 \in \lambda\partial\|u^*\|_1 + u^* - a$

$\partial\|u\| = \begin{cases} \text{sign}(u) & u \neq 0 \\ [-1, 1] & u = 0 \end{cases}$

if $u_i \neq 0$ then

$0 \in \text{sign}(u_i)\cdot\lambda + u_i^* - a_i \Rightarrow u_i^* = \begin{cases} a_i - \lambda & a_i > \lambda \\ a_i + \lambda & a_i < -\lambda \end{cases}$

if $u_i = 0$ then

$0 \in [-\lambda, \lambda] + u_i^* - a_i \Rightarrow |a_i| \leq \lambda$

$\therefore \text{prox}_h(a)_i \in \begin{cases} a_i - \lambda & a_i > \lambda \\ 0 & |a_i| \leq \lambda \\ a_i + \lambda & a_i < -\lambda \end{cases}$

(c) $0 \in \lambda \partial \|u^*\|_* + u^* - A$

$0 \in \lambda \{ y : tr((u^*)^T y) = \|u^*\|_*, \|y\| \le 1 \} + u^* - A$

$\therefore \quad u^* = A - \lambda y$

$\therefore \quad tr((A-\lambda y)^T \cdot y) = \|A - \lambda y\|_*, \quad \|y\| \le 1$

$\therefore \quad prox_h(A) = A - \lambda y$ where y can be obtained by solving

§3. (a) let $t_k^i = \rho^i t_k^0$ where $i = 1, 2, 3, \ldots$

i denotes i^{th} step in backtracking step. and t_k^i is the corresponding stepsize t

Assume at step i, $t_k^i \ge \frac{1}{L} > t_{min}$, $t_k^i \le \frac{1}{\rho L}$

If $g(x^k - t_k^i G_t(x^k)) \le g(x^k) - t_k^i \nabla g(x^k)^T G_t(x^k) + \frac{t_k^i}{2} \|G_t(x^k)\|^2$ satisfies

then $t_k = t_k^i > t_{min}$

Else:

then $t_k^{i+1} = \rho t_k^i$ and $t_k^{i+1} \in [\frac{\rho}{L} \; \frac{1}{L}]$

from the upper bound of $g(\cdot)$

$$g(x - t_k^{i+1} G_t(x)) \le g(x) + t_k^{i+1} \nabla g(x)^T G(x) + \frac{L {t_k^{i+1}}^2}{2} \|G_t(x)\|_2^2$$

$$\le g(x) + t_k^{i+1} \nabla g(x)^T G(x) + \frac{t_k^{i+1}}{2} \|G_t(x)\|_2^2$$

then $t_k = t_k^{i+1} \ge \frac{\rho}{L} \ge t_{min}$

$\therefore \quad t_k \ge t_{min} \quad \forall k$

(b) From (4) in slides 9-10, we have

$$f(x^k - t_k G_t(x^k)) \le f(z) + G_t(x^k)^T(x^k - z) - \frac{t_k}{2}\|G_t(x^k)\|_2^2 - \frac{m}{2}\|x^k - z\|_2^2$$

let $z = x^k$, then

$$f(x^k - t_k G(x^k)) \le f(x^k) - \frac{t_k}{2}\|G_t(x^k)\|_2^2 \le f(x^k)$$

$$\Rightarrow f(x^{k+1}) \le f(x^k) \quad \Rightarrow \{f(x^k)\} \text{ is nonincreasing}$$

let $z = x^*$ then

$$f(x^{k+1}) - f(x^*) \le G_t(x_k)^T(x^k - x^*) - \frac{t_k}{2}\|G_t(x^k)\|_2^2 - \frac{m}{2}\|x^k - x^*\|_2^2$$

$$\text{RHS} = \frac{1}{2t_k}\left(\|x^k - x^*\|_2^2 - \|x^k - x^* - t_k G_t(x^k)\|_2^2\right) - \frac{m}{2}\|x^k - x^*\|_2^2 \qquad ①$$

$$\le \frac{1}{2t_k}\left(\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right.$$

$$\therefore f(x^{k+1}) - f(x^*) \le \frac{1}{2t_k}\left(\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right) \quad \forall k \ge 0$$

(c)
$$\sum_{i=0}^{k}\left(f(x^{i+1}) - f(x^*)\right) \le \frac{1}{2t_k}\sum_{i=0}^{k}\left(\|x^i - x^*\|_2^2 - \|x^{i+1} - x^*\|_2^2\right)$$

$$= \frac{1}{2t_k}\left(\|x^0 - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right)$$

$$\le \frac{1}{2t_k}\|x^0 - x^*\|_2^2$$

$$\sum_{i=0}^{k}\left(f(x^{i+1}) - f(x^*)\right) \ge k\left(f(x^k) - f(x^*)\right) \Leftarrow \text{from (b), } f(x^k) \text{ nonincreasing}$$

$$\therefore f(x^k) - f(x^*) \le \frac{1}{2t_k \cdot k}\|x^0 - x^*\|_2^2$$

$$\le \frac{1}{2t_{min}k}\|x^0 - x^*\|_2^2$$

since $D = \min(\|x^0 - x\| \quad x \in X^*)$

$$\therefore f(x^k) - f(x^*) \le \frac{D^2}{2t_{min}k} \qquad \forall k \ge 1$$

(d) use (4) from slides 9–10, and ① from (b)

let $m = \sigma$ then

$$f(x^{k+1}) - f^* \le G_t(x^k)^T(x^k - x^*) - \frac{t_k}{2}\|G_t(x^k)\|^2 - \frac{\sigma}{2}\|x^k - x^*\|^2$$
$$= \frac{1}{2t_k}\left((1-\sigma t_k)\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right)$$
$$= \frac{1-\sigma t_k}{2t_k}\|x^k - x^*\|_2^2 - \frac{1}{2t_k}\|x^{k+1} - x^*\|_2^2 \quad \forall k \ge 0$$

(e) if g is strongly convex, then f is also strongly convex

$\therefore f(x^{k+1}) \ge f(x^*) + \frac{\sigma}{2}\|x^{k+1} - x^*\|_2^2$

1° from (d)

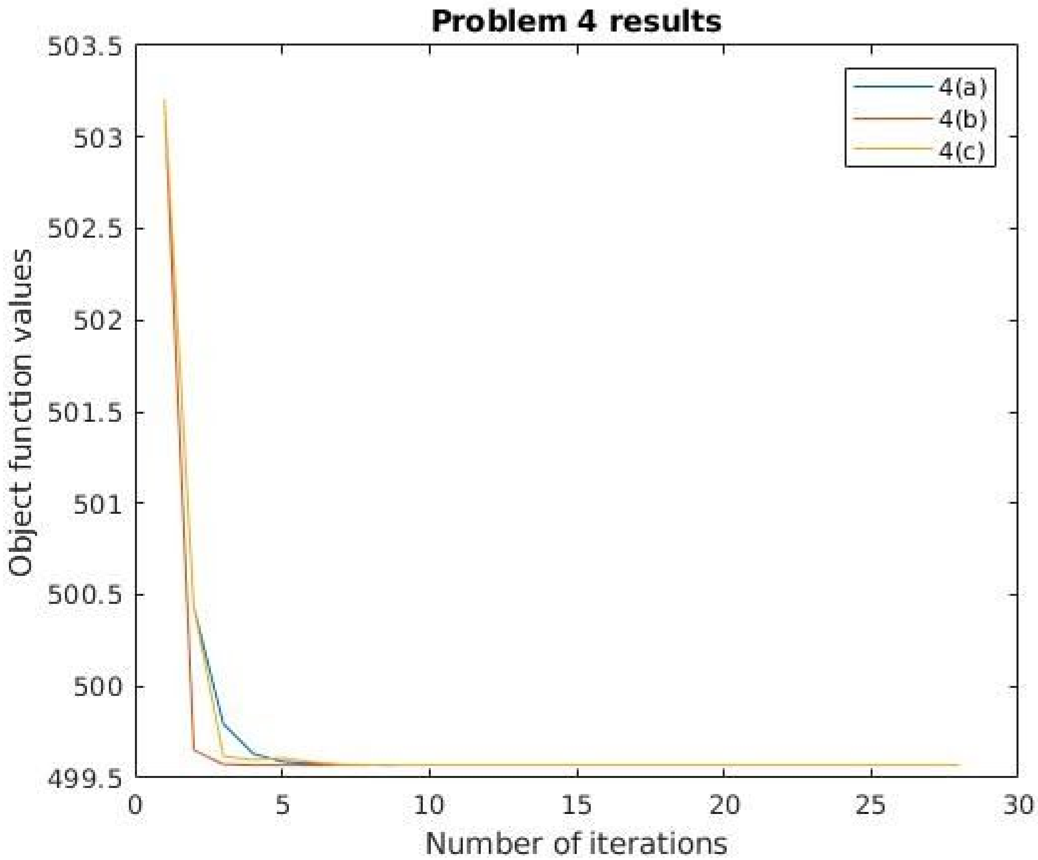
$$\frac{\sigma}{2}\|x^{k+1} - x^*\|_2^2 \le \frac{1-\sigma t_k}{2t_k}\|x^k - x^*\|_2^2 - \frac{1}{2t_k}\|x^{k+1} - x^*\|_2^2$$
$$\frac{\sigma}{2}\|x^{k+1} - x^*\|_2^2 \le \frac{1-\sigma t_{min}}{2t_{min}}\|x^k - x^*\|_2^2 - \frac{1}{2t_{min}}\|x^{k+1} - x^*\|_2^2$$

$$\therefore \|x^{k+1} - x^*\|_2^2 \le \frac{1-\sigma t_{min}}{1+\sigma t_{min}} \|x^k - x^*\|_2^2$$

$$\therefore \|x^k - x^*\|_2^2 \le \left(\frac{1-\sigma t_{min}}{1+\sigma t_{min}}\right)^k \|x^0 - x^*\|_2^2 = \left(1 - \frac{2\sigma t_{min}}{1+\sigma t_{min}}\right)^k \|x^0 - x^*\|_2^2$$

2° from (d) and (e) 1°

$$f(x^k) - f^* \le \frac{1-\sigma t_{min}}{2t_{min}}\|x^{k-1} - x^*\|_2^2$$
$$\le \frac{1-\sigma t_{min}}{2t_{min}}\left(1 - \frac{2\sigma t_{min}}{1+\sigma t_{min}}\right)^{k-1}\|x^0 - x^*\|_2^2$$
$$= \frac{1-\sigma t_{min}}{2t_{min}}\left(1 - \frac{2\sigma t_{min}}{1+\sigma t_{min}}\right)^{k-1} D^2$$



	4(a)	4(b)	4(c)
final object value	499.5697	499.5697	499.5697
number of proximal gradient evaluations	46	92	56
number of outer iterations	23	8	28
CPU time	0.08s	0.09s	0.04s