1 Questions with multiple answers

1. Answer: (a)

   Explanation:

   - We know that f is convex if and only if $f''(x) \geq 0$, for $\forall x \in \mathrm{X}$, where $f : \mathrm{X} \to \mathbb{R}$. For (c), $f''(x) = x^2(12 - 4e^{x^2}) < 0$ when $x^2 < \log 3$. Thus (c) is non-convex.
   - It is obvious that $x^4 - 5 < e^{x^2}$, for $\forall x \in \mathbb{R}$. So in (b), $f(x) = \min\{ax + b, x^4 - 5, e^{x^2}\} \Leftrightarrow f(x) = \min\{ax + b, x^4 - 5\}$. In the following case, $f(x)$ is non-convex, where $f((1-\lambda)x + \lambda y) > (1-\lambda)f(x) + \lambda f(y)$ on either $x \geq 0$ or $x \leq 0$:

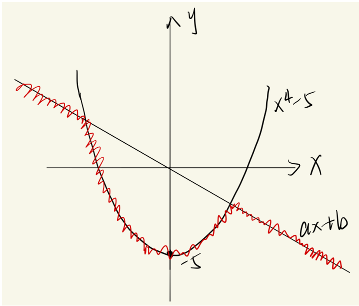

Figure 1: the red line is $f(x)$

2. Answer: (a)

   Explanation:

   We first compute $f'(x)$:

$$f'(x) = \begin{cases} -1, \ x \in]-1, 0] \\ 2x, \ x \geq 0 \end{cases} \tag{1}$$

   We know that $f(x)$ is a proper convex function on $]-1, +\infty]$. Then by proposition 3.1.8(i) in lecture 3, $\partial f(x) = [f'_-(x), f'_+(x)] \cap]-1, +\infty] = [-1, 0]$.

3. Answer: (a)

   Explanation:

   A is not necessarily symmetric. By matrix derivatives rule, $f' = A^*x + Ax + b$.

4. Answer: (b)

   Explanation:

   The Fenchel conjugate of $f(x)$: $f^*(u) = \sup_x\{\langle u, x\rangle - f(x)\}$. By differentiating $f^*(u)$ w.r.t x, we have $u = f'(x)$. Thus: $f^*(u) = \sup_x\{\langle u, x\rangle - f(x)\} = xf'(x) - f(x)$. Similarly, the Fenchel conjugate of $g(x)$: $g^*(u) = xg'(x) - g(x)$. Since $f(x) = g(2x)$, we have$f'(x) = 2g'(2x)$, and thus:

$$\begin{aligned} f^*(u) &= xf'(x) - f(x) \\ &= 2xg'(2x) - g(2x) \\ &= g^*(2u) \end{aligned} \tag{2}$$

5. Answer: (c)

   Explanation:

The problem can be converted into a constrained minimisation problem:

$$\arg\min_{w,r} \left[\frac{1}{2}||r||^2 + \frac{\lambda}{2}||w||^2\right] \text{s.t. } r = w^\top X - y\,, \tag{3}$$

whose Lagrangian is:

$$L(w, r, u) = \frac{1}{2}||r||^2 + \frac{\lambda}{2}||w||^2 + u^\top(r - w^\top X + y) \tag{4}$$

Setting derivatives w.r.t the primal variables w, r to zero, we obtain:

$$\begin{aligned} \frac{\partial L}{\partial r} &= r + u = 0,\ r = -u \\ \frac{\partial L}{\partial w} &= \lambda w - u^\top X = 0,\ w = \frac{u^\top X}{\lambda} \end{aligned} \tag{5}$$

Plugging the results obtained above back to the Lagrangian, we manage to eliminate r and w and hence obtain the dual function:

$$\begin{aligned} L(w, r, u) &= \frac{1}{2}||u||^2 + \frac{1}{2\lambda}||Xu||^2 + u^\top(-u - \frac{X^\top Xu}{\lambda} + y) \\ &= -\frac{1}{2}||u||^2 - \frac{1}{2\lambda}||Xu||^2 + u^\top y \end{aligned} \tag{6}$$

The dual problem:

$$\begin{aligned} &\arg\min_u \left[\frac{1}{2}u^\top(nId + \frac{1}{\lambda}X^\top X)u - u^\top y\right] \\ \Leftrightarrow\quad &\arg\min_u \left[\frac{1}{2}u^\top(\lambda nId + X^\top X)u - \lambda u^\top y\right] \\ \Leftrightarrow\quad &\arg\min_u \left[\frac{1}{2}u^\top(K + \lambda nId)u - \lambda u^\top y\right], \end{aligned} \tag{7}$$

where $K = X^\top X$ is the gram matrix of X. Thus the solution is obtained by setting the derivatives of the above equation to zero:

$$\begin{aligned} (K + \lambda nId)u - \lambda y = 0 \\ u = (K + \lambda nId)^{-1}\lambda y \end{aligned} \tag{8}$$

2 Theory on convex analysis and optimization

Problem 1

1. When $x > 0$, the Fenchel conjugate: $f^*(u) = \sup_x [\langle u, x\rangle - f(x)] = \sup_x[ux + \log x]$.
   By setting the derivatives w.r.t x to zero, we have:
   $$\frac{\partial f^*(u)}{\partial x} = u + \frac{1}{x} = 0,\ x = -\frac{1}{u} \tag{9}$$
   Thus $f^*(u) = -1 + \log(-\frac{1}{u}) = -1 - \log(-u),\ u < 0$. Thus the Fenchel conjugate for f:
   $$f^*(u) = \begin{cases} -\infty,\ u \geq 0 \\ -1 - \log(-u),\ u < 0 \end{cases} \tag{10}$$
2. The Fenchel conjugate: $f^*(u) = \sup_x[ux - x^2]$.
   By setting the derivatives w.r.t x to zero, we have:
   $$\frac{\partial f^*(u)}{x} = u - 2x = 0,\ x = \frac{u}{2} \tag{11}$$
   Thus the Fenchel conjugate: $f^*(u) = \frac{u^2}{4}$.
3. When $x \in [0, 1]$, the Fenchel conjugate: $f^*(u) = \sup_x[ux - 0] = \sup_x[ux]$.
   By setting the derivatives w.r.t x to zero, we have:
   $$\frac{\partial f^*(u)}{x} = u = 0 \tag{12}$$
   Thus the Fenchel conjugate:
   $$f^*(u) = \begin{cases} 0,\ u \in [0, 1] \\ -\infty,\ \text{otherwise} \end{cases} \tag{13}$$

Problem 2

1. Prove by induction:
   When $n = 2$, $f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$ is obvious, which is the definition of convexity.
   Assume the Jensen's inequality holds for n, then:
   $$\begin{aligned} f(\sum_{i=1}^{n+1} \lambda_i x_i) &= f\left(\sum_{i=1}^{n} \lambda_i x_i + \lambda_{n+1} x_{n+1}\right) \\ &= f\left((1-\lambda_{n+1})\left(\frac{1}{1-\lambda_{n+1}}\sum_{i=1}^{n} \lambda_i x_i\right) + \lambda_{n+1} x_{n+1}\right) \\ &\leq (1-\lambda_{n+1}) f\left(\frac{1}{1-\lambda_{n+1}}\sum_{i=1}^{n} \lambda_i x_i\right) + \lambda_{n+1} f(x_{n+1}), \text{ by convexity of } f \\ &\leq (1-\lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} f(x_i) + \lambda_{n+1} f(x_{n+1}), \text{ since Jensen's inequality holds for } n \\ &= \sum_{i=1}^{n} \lambda_i x_i + \lambda_{n+1} f(x_{n+1}) \\ &= \sum_{i=1}^{n+1} \lambda_i x_i \end{aligned} \tag{14}$$

2. $f''(x) = (-\log x)'' = \frac{1}{x^2} > 0$, for $x > 0$, thus $f(x)$ is twice differentiable. This indicates that $f(x)$ is stricly increasing. Thus $\langle \nabla f(x) - \nabla f(y), x - y \rangle > 0$, for $x, y \in \mathbb{R}$ and $x \neq y$. Thus by proposition 2.1.2 (iv) in lecture notes, $f(x) = -\log x$ is strictly convex.

3. By Jensen's inequality:

$$\begin{aligned} -\log\left(\sum_{i=1}^{n} \lambda_i x_i\right) &\leq -\sum_{i=1}^{n} \lambda_i \log(x_i) \\ \log\left(\sum_{i=1}^{n} \lambda_i x_i\right) &\geq \sum_{i=1}^{n} \lambda_i \log(x_i) \end{aligned} \tag{15}$$

Set $\lambda_i = \frac{1}{n}$ and take $\exp(\cdot)$ to both sides:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} x_i &\geq \exp\left(\sum_{i=1}^{n} \frac{1}{n}\log(x_i)\right) \\ &= \prod_{i=1}^{n} \exp(\frac{1}{n}\log(x_i)) \\ &= \prod_{i=1}^{n} (x_i)^{\frac{1}{n}} \\ &= \sqrt[n]{x_1 \cdots x_n} \end{aligned} \tag{16}$$

Problem 3

The polytope $C = co(a_1, \cdots, a_m)$ can be written in the form $A\lambda$, where $A = [a_1, \cdots, a_m]$, $\lambda = [\lambda_1, \cdots, \lambda_m]$ with $\lambda_i > 0$ and $\sum_i^m \lambda_i = 1$.
By convexity of f and Jensen's inequality:

$$f(A\lambda) = f(\sum_{i=1}^{m} \lambda_i a_i) \leq \sum_{i=1}^{m} \lambda_i f(a_i) \leq \max_{i=1,\cdots,m} f(a_i) \tag{17}$$

Hence the maximum of the convex function f on C is attained at one of the vertices.

Problem 4

To prove $f(x, y) = ||x - 2y||_2^2$ is jointly convex, we need to prove:

$$f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \leq \lambda f(x_1, y_1) + (1-\lambda)f(x_2, y_2) \tag{18}$$

For LHS:

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) &= ||\lambda x_1 + (1-\lambda)x_2 - 2\lambda y_1 - 2(1-\lambda)y_2||_2^2 \\ &= ||\lambda(x_1 - 2y_1) + (1-\lambda)(x_2 - 2y_2)||_2^2 \\ &= \lambda^2||x_1 - 2y_1||_2^2 + (1-\lambda)^2||x_2 - 2y_2||_2^2 + 2\lambda(1-\lambda)||(x_1 - 2y_1)(x_2 - 2y_2)||_2 \end{aligned} \tag{19}$$

For RHS:

$$\lambda f(x_1, y_1) + (1-\lambda)f(x_2, y_2) = \lambda||x_1 - 2y_1||_2^2 + (1-\lambda)||x_2 - 2y_2||_2^2 \tag{20}$$

Thus

$$
\begin{aligned}
\text{LHS} - \text{RHS} &= \lambda(\lambda-1)||x_1 - 2y_1||_2^2 + (1-\lambda)(-\lambda)||x_2 - 2y_2||_2^2 + 2\lambda(1-\lambda)||(x_1 - 2y_1)(x_2 - 2y_2)||_2 \\
&= \lambda(\lambda-1)\Bigg(||x_1 - 2y_1||_2^2 + ||x_2 - 2y_2||_2^2 - 2||(x_1 - 2y_1)(x_2 - 2y_2)||_2\Bigg) \\
&\leq \lambda(\lambda-1)\Bigg(||x_1 - 2y_1||_2^2 + ||x_2 - 2y_2||_2^2 - 2||(x_1 - 2y_1)||_2||(x_2 - 2y_2)||_2\Bigg) \\
&= \lambda(\lambda-1)\Bigg(||x_1 - 2y_1||_2 - ||x_2 - 2y_2||_2\Bigg)^2 \leq 0\,,
\end{aligned} \tag{21}
$$

since $\lambda > 0$,$(\cdot)^2 \leq 0$ and $\lambda - 1 < 0$. Thus we have LHS $\leq$ RHS.

Problem 5

The minimal sufficient conditions for existence and uniqueness:

- Existence:
  - f is continuous and bounded below with $\inf f(x) < \infty$.
- Uniqueness:
  - f is strictly convex, i.e. $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$, for $\forall x, y \in X$.
  - X is a closed, bounded and convex set.

Problem 6

1. The problem can also be written in the form:

$$
\min_{x \in \mathcal{C}} \Bigg(\frac{1}{2}||x||^2 + \mathbb{I}_{Ax-b\leq\epsilon}\Bigg) \Leftrightarrow \min_{x \in \mathcal{C}} \Bigg(f(x) + g(Ax)\Bigg)\,, \tag{22}
$$

where $f(x) = \frac{1}{2}||x||^2$ and $g(Ax) = \mathbb{I}_{Ax-b\leq\epsilon}$.
By Fenchel-Rockafellar duality theory:

$$
\min_{x \in \mathcal{C}} \Bigg(f(x) + g(Ax)\Bigg) \Leftrightarrow \max_{y} \Bigg(-f^*(-A^*y) - g^*(y)\Bigg) \Leftrightarrow \min_{y} \Bigg(f^*(-A^*y) + g^*(y)\Bigg) \tag{23}
$$

The Fenchel conjugate for f and g:

$$
\begin{aligned}
f^*(u) &= \sup_{x} \Bigg[\langle u, x\rangle - \frac{1}{2}||x||^2\Bigg] = \frac{1}{2}||u||^2 \\
g^*(u) &= \sup_{||Ax-b||<\epsilon} \Bigg[\langle u, Ax\rangle\Bigg] = \epsilon||u|| + \langle b, u\rangle
\end{aligned} \tag{24}
$$

Thus the dual problem is:

$$
\min_{u} \Bigg(\frac{1}{2}||-A^*u||^2 + \epsilon||u|| + \langle b, u\rangle\Bigg) \tag{25}
$$

2. The strong duality holds.

   We assume that $\exists\, x$ s.t. $Ax < b$, then we know that $Ax \in \operatorname{int}(\operatorname{dom} g)$ and the qualification condition $0 \in \operatorname{int}(\operatorname{dom} g - A(\operatorname{dom} f))$ hold. The we can apply Theorem 8.1.1 and 8.1.6 in notes. By theorem 8.1.6, the duality gap is 0, and thus strong duality holds.

3. The KKT conditions:

$$\begin{aligned} \hat{x} &\in \partial f^*(-A^*\hat{u}) = AA^*||\hat{u}|| \\ A\hat{x} &\in \partial g^*(\hat{u}) = b + \epsilon \\ -A^*\hat{u} &\in \partial f(\hat{x}) = ||\hat{x}|| \\ \hat{u} &\in \partial g(A\hat{x}) = \partial\, \mathbb{I}_{Ax-b\leq\epsilon} \end{aligned} \tag{26}$$

4. For FISTA, we have:

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_{\gamma,g}(y_k - \gamma\nabla f^*(y_k)) \\ y_{k+1} &= x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k) \end{aligned} \tag{27}$$

   where $t_k^2 - t_k \leq t_{k-1}^2$, $t_0 = 1$ and $t_k \geq 1$.

   By reorganising the second equation, we have:

$$\begin{aligned} y_{k+1} &= x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k) \\ &= \Big(1 - \frac{1}{t_{k+1}}\Big)x_{k+1} + \frac{1}{t_{k+1}}\Big(x_k + t_k(x_{k+1} - x_k)\Big) \\ &= \Big(1 - \frac{1}{t_{k+1}}\Big)x_{k+1} + \frac{1}{t_{k+1}}v_{k+1}\,, \end{aligned} \tag{28}$$

   where we denote $v_{k+1} = x_k + t_k(x_{k+1} - x_k)$ ($v_0 = y_0$). Thus:

$$\begin{aligned} x_{k+1} &= (1 - \frac{1}{t_k})x_k + \frac{1}{t_k}v_{k+1} \\ y_k &= \Big(1 - \frac{1}{t_k}\Big)x_k + \frac{1}{t_k}v_k \end{aligned} \tag{29}$$

   From equation (24), we know $f^* = \frac{1}{2}||-A^*u||^2$ and $g^* = \epsilon||u|| + \langle b, u\rangle$. Set $F = f^* + g^*$ (F is convex). Then we have:

$$F(x_{k+1}) + \frac{||x_{k+1} - x||^2}{2\gamma} \leq F(x) + \frac{||y_k - x||^2}{2\gamma}\ , \text{ for } \forall x \in X \tag{30}$$

   Denote $a = x_{k+1} - x$, $b = y_k - x$ and take $x = \Big(1 - \frac{1}{t_k}\Big)x_k + \frac{x^*}{t_k}$. Combining the settings with equation (29), we have:

$$a = \frac{v_{k+1} - x^*}{t_k}\ ,\ b = \frac{v_k - x^*}{t_k} \tag{31}$$

   Plugging equation (31) into (30), we have:

$$\begin{aligned} F(x_{k+1}) + \frac{||v_{k+1} - x^*||^2}{2\gamma t_k^2} &\leq (1 - \frac{1}{t_k})F(x_k) + \frac{1}{t_k}F(x^*) + \frac{||v_k - x^*||^2}{2\gamma t_k^2} \\ F(x_{k+1}) - F(x^*) + \frac{||v_{k+1} - x^*||^2}{2\gamma t_k^2} &\leq (1 - \frac{1}{t_k})(F(x_k) - F(x^*)) + \frac{||v_k - x^*||^2}{2\gamma t_k^2} \\ t_k^2(F(x_{k+1}) - F(x^*)) + \frac{||v_{k+1} - x^*||^2}{2\gamma} &\leq (t_k^2 - t_k)(F(x_k) - F(x^*)) + \frac{||v_k - x^*||^2}{2\gamma} \\ &\leq t_{k-1}^2(F(x_k) - F(x^*)) + \frac{||v_k - x^*||^2}{2\gamma} \end{aligned} \tag{32}$$

Thus we've obtained the recursive inequality. By applying the inequality for k times, we have:

$$t_{k-1}^2(F(x_k) - F(x^*)) + \frac{||v_k - x^*||^2}{2\gamma} \leq t_0^2(F(x_1) - F(x^*)) + \frac{||v_1 - x^*||^2}{2\gamma} \leq \frac{||v_0 - x^*||^2}{2\gamma} \tag{33}$$

Since $y_0 = v_0$, we have:

$$\begin{aligned} t_{k-1}^2(F(x_k) - F(x^*)) + \frac{||v_k - x^*||^2}{2\gamma} &\leq \frac{||y_0 - x^*||^2}{2\gamma} \\ t_{k-1}^2(F(x_k) - F(x^*)) &\leq \frac{||y_0 - x^*||^2}{2\gamma} \\ F(x_k) - F(x^*) &\leq \frac{||y_0 - x^*||^2}{2\gamma t_k^2}\,, \end{aligned} \tag{34}$$

which has convergence rate at $\mathcal{O}(\frac{1}{k^2})$.

3 Solving the lasso problem

The objective function value change (i.e. loss) for x and $\bar{x}$ under PSGA, and for x under RCPGA are displayed below:

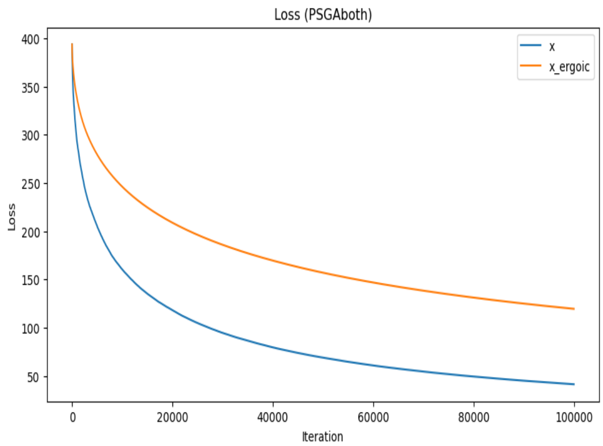

Figure 2: PSGA

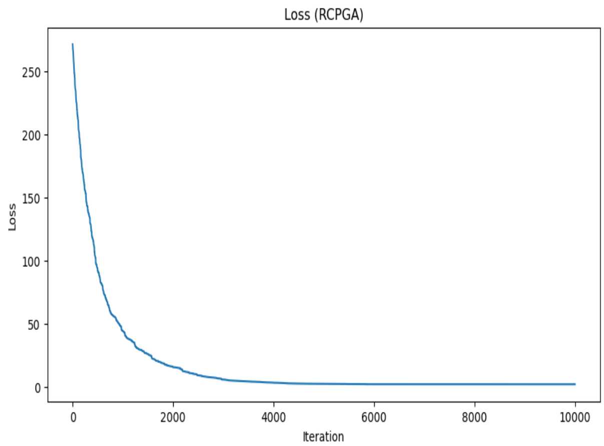

Figure 3: RCPGA

The solutions for 2 algorithms compared with sparse vector:

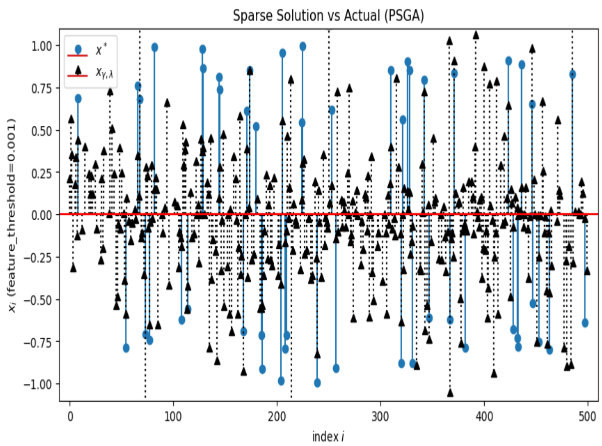

Figure 4: PSGA

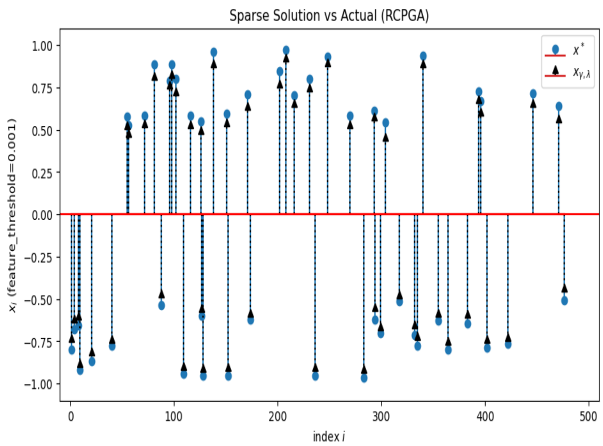

Figure 5: RCPGA

The solution of PSGA with ergodic x compared with sparse vector:

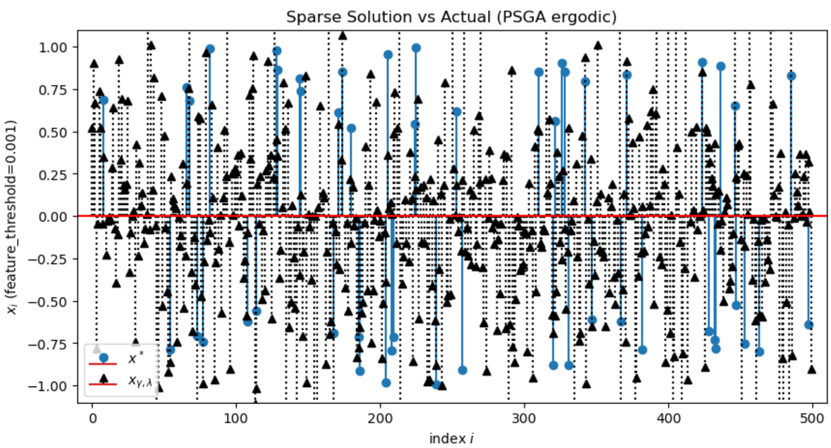

Figure 6: PSGA with ergodic

From the above figures, RCPGA can recover the sparse vector while PSGA cannot even after 10k times iterations (may still need larger iteration). Meanwhile, RCPGA performs much better than PSGA computationally with significantly faster convergence rate. When change x to the ergodic version under PSGA, the convergence rate becomes even slower.