MATH130165h: Homework 5.

20307100005.

Problem 1

[1] $f_1(x) = b^T x \quad \frac{\partial f_1(x)}{\partial x_i} = b_i \quad \therefore \nabla f_1(x) = b.$

$$\frac{\partial^2 f_1(x)}{\partial x_i \partial x_j} = 0 \quad \forall i, j \quad \therefore H(f_1) = 0$$

[2] $f_2(x) = x^T A x \quad \frac{\partial f_2(x)}{\partial x_i} = \frac{\partial (x^T A)}{\partial x_i} x + x^T A \frac{\partial x}{\partial x_i} = (x^T A)_{[i,]} \cdot x + x^T A e_i$

$$= A_{[i,]} \cdot x + (x^T A)_{[,i]} = A_{[i,]} x + A^T x_{[i,]}.$$

$$\therefore \nabla f_2(x) = A x + A^T x = (A + A^T) x.$$

$$\frac{\partial(\nabla f_2(x))}{\partial x_j} = A_{[,j]} + A^T_{[,j]} \Rightarrow \frac{\partial(\nabla f_2(x))}{\partial x} = A + A^T$$

And $H(f_2)_{[i,j]} = \frac{\partial^2 f_2(x)}{\partial x_j \partial x_i} = \frac{\partial(\nabla f_2(x))}{\partial x_j} \quad \therefore H(f_2) = A + A^T$

Problem 2

[1] $f_1(X) = tr(B^TX) = \sum_{i=1}^{n} (B^TX)_{[i,i]} = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} B^T_{ik} X_{ki} \right) = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} B_{ki} X_{ki} \right)$

$\therefore \dfrac{\partial f_1(X)}{\partial X_{ki}} = B_{ki} \qquad \therefore \nabla f_1(X) = B.$

[2] $f_2(x) = tr(X^TAX) = \sum_{i=1}^{n} (X^TAX)_{[i,i]} = \sum_{i=1}^{n} X^T_{[i,]} A X_{[,i]}.$

$\therefore \dfrac{\partial f_2(X)}{\partial X_{[,i]}} = (A^T + A) X_{[,i]} \qquad \therefore \nabla f_2(X) = A^T + A$

[3] $f_3(X) = tr(BX^TAXX^TBX^T)$

$\dfrac{\partial tr[AXB]}{\partial x_{ij}} = \dfrac{\partial \sum_{i=1}^{n} A_{[i,]} X B_{[,i]}}{\partial x_{ij}} = (BA)_{[j,i]} \qquad \therefore \dfrac{\partial tr[AXB]}{\partial X} = A^TB^T$

First, we know from the calculus rule that:

$$\frac{\partial tr[AXBXC]}{\partial X} = \frac{\partial tr[AXD]}{\partial X} + \frac{\partial tr[EXC]}{\partial X} = A^TD^T + E^TC^T$$

$$= A^T(BXC)^T + E^T(BXC)^T$$

$C = BX^T$

$D = CAX$
$= BX^TAX$

Thus, $\dfrac{\partial tr(BX^TAXX^TBX^T)}{\partial X} = AXX^TBX^TB + \underbrace{\dfrac{\partial tr(CAXX^TBX^T)}{\partial X}}$

$\hookrightarrow = (CA)^T(X^TBX^T)^T + \underbrace{\dfrac{\partial tr(DX^TBX^T)}{\partial X}}$

$\hookrightarrow = BX^TD + DX^TB$

$\therefore \dfrac{\partial tr(BX^TAXX^TBX^T)}{\partial X} = AXX^TBX^TB + (BX^TA)^T X B^TX + BX^TBX^TAX + DX^TB$

$= AXX^TBX^TB + A^TXB^TXB^TX + BX^TBX^TAX + BX^TAXX^TB.$

Problem 3

$\because g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ is convex

$\therefore g[\alpha x + (1-\alpha)y] \le \alpha g(x) + (1-\alpha) g(y)$

$\Leftrightarrow f[\alpha x + (1-\alpha)y] - \frac{\mu}{2}\|\alpha x + (1-\alpha)y\|^2 \le \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}(\alpha\|x\|^2 + (1-\alpha)\|y\|^2)$

$$\Leftrightarrow f[\alpha x + (1-\alpha)y] \le \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}\left[\alpha\|x\|^2 - \|\alpha x + (1-\alpha)y\|^2 + (1-\alpha)\|y\|^2\right]$$

$$= \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}\left[\alpha x^T x - [\alpha x^T + (1-\alpha)y^T][\alpha x + (1-\alpha)y] + (1-\alpha)y^T y\right]$$

$$= \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}\left[-\alpha(1-\alpha)(x^T y + y^T x) + \alpha(1-\alpha)x^T x + (1-\alpha)\alpha\, y^T y\right]$$

$$= \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}(1-\alpha)\alpha\left[x^T x + y^T y - x^T y - y^T x\right]$$

$$= \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}(1-\alpha)\alpha\|x-y\|^2 \quad \forall x, y,\ \forall \alpha \in [0,1]$$

$\therefore 1 \Leftrightarrow 3$.

$[\nabla f(x) - \nabla f(y)]^T (x-y) \ge \mu(x^T - y^T)(x-y)$

$\Leftrightarrow [\nabla f(x) - \nabla f(y) - \mu(x-y)]^T (x-y) \ge 0$

$\because g(x)$ is convex $\quad \therefore \forall i,\ x_i - y_i > 0 \Leftrightarrow \nabla g(x)_i - \nabla g(y)_i > 0$

$\qquad\qquad\qquad\qquad\qquad\qquad - - < 0 \Leftrightarrow - - < 0$

$\therefore [\nabla g(x) - \nabla g(y)]^T (x-y) \ge 0. \quad \because \nabla g(x) = \nabla f(x) - \mu x$

$\therefore [\nabla f(x) - \mu x - \nabla f(y) + \mu y]^T (x-y) \ge 0$

$\therefore 2 \Leftrightarrow 1$

$\therefore 2 \Leftrightarrow 1 \Leftrightarrow 3$.

$\blacksquare$

Problem 4

Suppose x_1, x_2 are the global minimizers of f.

$f(x_1) = f(x_2) \leq f(x) \quad \forall x \in D$

$\because$ f is convex

$\therefore$ $f(tx_1 + (1-t)x_2) \leq t f(x_1) + (1-t) f(x_2) \leq t f(x) + (1-t) f(x) = f(x)$

$\forall t \in [0,1], \forall x$

$\therefore$ $tx_1 + (1-t)x_2$ is a global minimizer of f.

That is, $x_1, x_2 \in M$, $\forall t \in [0,1]$, $tx_1 + (1-t)x_2 \in M$.

$\therefore$ The set of global minimizers are convex.

Problem 5

$\nabla f(x) = Ax - b$ $\therefore$ update rule : $x_{k+1} = x_k - \alpha(Ax_k - b)$

$\nabla f(x^*) = 0$ $x^* = A^{-1}b$

Let $e_k = x_k - x^*$ $e_{k+1} = x_{k+1} - x^* = (I - \alpha A)x_k + \alpha b - A^{-1}b$

$$= (I - \alpha A)x_k - A^{-1}b + \alpha A A^{-1} b$$

$$= (I - \alpha A)x_k - (I - \alpha A)A^{-1}b$$

$$= (I - \alpha A)e_k$$

Thus, to converge, the spectral radius of $I - \alpha A < 1$

$\rho(I - \alpha A) < 1$ $\Rightarrow \alpha \in \left(0, \frac{2}{\lambda_{max}}\right)$

The norm of $I - \alpha A$ is a constant, so the convergence rate is linear.

$\blacksquare$

$$\frac{\partial tr(AXX^TBCX^TXC)}{\partial X} = A^T(X^TBCX^TXC)^T + \underbrace{\frac{\partial(DX^TBCX^TXC)}{\partial X}}$$

$$\rightarrow BCX^TXCD + \underbrace{\frac{\partial(EBCX^TXC)}{\partial X}}$$

$$\rightarrow XCEBC + (EBCX^T)^T(C)^T$$

$$\therefore = A^T(X^TBCX^TXC)^T + BCX^TXCAX + XCAXX^TBC + (AXX^TBCX^T)^TC^T$$

$$D = AX \quad E = DX^T = AXX^T$$