HW4 Solution

1. Let V be a Hilbert space. Let $\boldsymbol{a} \in V$ be a given vector. The function $A\boldsymbol{x} := \langle \boldsymbol{a}, \boldsymbol{x} \rangle$ can be viewed as a linear transformation from V to $\mathbb{R}$. Find the operator norm $\|A\|$.

Solution: By definition, $\|A\| = \sup_{\|x\|=1} |\langle a, x \rangle|$

By Cauchy-Schwartz Inequality,

$$|\langle a, x \rangle| \leq \|a\| \cdot \|x\|$$

Hence $\|A\| \leq \sup_{\|x\|=1} \|a\| \cdot \|x\| = \|a\|$. $\quad \cdots \quad (1)$

On the other hand, choose $x = \frac{a}{\|a\|}$. Then

$$\|x\| = \left\| \frac{a}{\|a\|} \right\| = \frac{\|a\|}{\|a\|} = 1.$$

Hence $\sup_{\|x\|=1} |\langle a, x \rangle| \geq \langle a, \frac{a}{\|a\|} \rangle = \|a\|$

Together with (1), we have

$$\|A\| = \sup_{\|x\|=1} |\langle a, x \rangle| = \|a\|.$$

2. Let $f_1, f_2, \ldots, f_n$ are differentiable functions from $V \mapsto \mathbb{R}$ with V a Hilbert space. Define $F : V \mapsto \mathbb{R}^n$ by

$$F(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_n(\boldsymbol{x}) \end{bmatrix}, \quad \forall\, \boldsymbol{x} \in V.$$

Prove that

$$DF(\boldsymbol{x})(\boldsymbol{y}) = \begin{bmatrix} \langle \nabla f_1(\boldsymbol{x}), \boldsymbol{y} \rangle \\ \langle \nabla f_2(\boldsymbol{x}), \boldsymbol{y} \rangle \\ \vdots \\ \langle \nabla f_n(\boldsymbol{x}), \boldsymbol{y} \rangle \end{bmatrix}, \quad \forall\, \boldsymbol{x} \in V.$$

Solution: Since $f_1, f_2, \cdots, f_n$ are differentiable functions from $V \mapsto \mathbb{R}$.

$$f_i(x) = f_i(x^{(0)}) + Df_i(x^{(0)})(x - x^{(0)}) + o(\|x - x^{(0)}\|),$$

$i = 1, 2, \cdots, n$

then $F(x) = F(x^{(0)}) + DF(x^{(0)})(x - x^{(0)}) + o(\|x - x^{(0)}\|)$

$$= \begin{pmatrix} f_1(x^{(0)}) \\ \vdots \\ f_n(x^{(0)}) \end{pmatrix} + \begin{pmatrix} Df_1(x^{(0)}) \\ \vdots \\ Df_n(x^{(0)}) \end{pmatrix} (x - x^{(0)}) + o(\|x - x^{(0)}\|)$$

$$DF(x)(y) = \begin{pmatrix} Df_1(x) \\ \vdots \\ Df_n(x) \end{pmatrix} (y)$$

Since V is a Hilbert space, we have

$$Df_i(x)(y) = \langle \nabla f_i(x), y \rangle, \quad i = 1, 2, \cdots, n$$

Thus

$$DF(x)(y) = \begin{pmatrix} \langle \nabla f_1(x), y \rangle \\ \vdots \\ \langle \nabla f_n(x), y \rangle \end{pmatrix}$$

3. Find $\nabla f(\boldsymbol{x})$ and $\nabla^2 f(\boldsymbol{x})$.

   (a) $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, and $\lambda > 0$ are given.
   (b) $f(\boldsymbol{X}) = \boldsymbol{b}^T\boldsymbol{X}\boldsymbol{c}$, where $\boldsymbol{X} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$.
   (c) $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{R}^n$, and $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is non-symmetric.
   (d) $f(\boldsymbol{X}) = \boldsymbol{b}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{c}$, where $\boldsymbol{X} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$.
   (e) $f(\boldsymbol{X}) = \text{trace}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})$, where $\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$.

Solution: (a) $f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_2^2$

$$= \frac{1}{2}\|A(x - x^{(0)}) + Ax^{(0)} - b\|_2^2 + \lambda\|x - x^{(0)} + x^{(0)}\|_2^2$$

$$= \frac{1}{2}\|A(x - x^{(0)})\|_2^2 + \langle A(x - x^{(0)}), Ax^{(0)} - b \rangle$$
$$+ \frac{1}{2}\|Ax^{(0)} - b\|_2^2 + \lambda\|x - x^{(0)}\|_2^2 + 2\lambda\langle x - x^{(0)}, x^{(0)} \rangle$$
$$+ \|x^{(0)}\|_2^2$$

$$= \frac{1}{2}\|A(x-x^{(0)})\|_2^2 + \lambda\|x-x^{(0)}\|_2^2$$
$$+ \langle A(x-x^{(0)}), Ax^{(0)}-b\rangle + 2\lambda\langle x-x^{(0)}, x^{(0)}\rangle$$
$$+ \frac{1}{2}\|Ax^{(0)}-b\|_2^2 + \|x^{(0)}\|_2^2$$
$$= f(x^{(0)}) + \langle A^T(Ax^{(0)}-b) + 2\lambda x^{(0)}, x-x^{(0)}\rangle$$
$$+ o(\|x-x^{(0)}\|_2)$$

Therefore $\nabla f(x) = A^T(Ax - b) + 2\lambda x$

Furthermore, $\nabla f(x) = A^T(A(x-x_0+x_0)-b) + 2\lambda(x-x_0+x_0)$
$$= A^TAx_0 - b + 2\lambda x_0 + A^TA(x-x_0) + 2\lambda(x-x_0)$$
$$= \nabla f(x^{(0)}) + (A^TA + 2\lambda)(x-x_0)$$

Hence $\nabla(\nabla f(x^{(0)})) = A^TA + 2\lambda I$

That is, $\nabla^2 f(x) = A^TA + 2\lambda I$

(b) $f(X) = b^TXc = b^T(X - X^{(0)} + X^{(0)})c$
$$= b^T[(X-X^{(0)})c + X^{(0)}c]$$
$$= b^T(X-X^{(0)})c + b^TX^{(0)}c$$
$$= f(X^{(0)}) + b^T(X-X^{(0)})c$$
$$= f(X^{(0)}) + \langle b, (X-X^{(0)})c\rangle$$
$$= f(X^{(0)}) + \langle bc^T, X-X^{(0)}\rangle$$

Hence, $\nabla f(X) = bc^T$.

Since bc^T is a constant matrix, $\nabla^2 f(X) = 0$

(c) $f(x) = x^TAx = (x-x^{(0)}+x^{(0)})^TA(x-x^{(0)}+x^{(0)})$
$$= (x-x^{(0)})^TA(x-x^{(0)}+x^{(0)}) + (x^{(0)})^TA(x-x^{(0)}+x^{(0)})$$
$$= (x-x^{(0)})^TA(x-x^{(0)}) + (x-x^{(0)})^TAx^{(0)}$$
$$+ (x^{(0)})^TA(x-x^{(0)}) + (x^{(0)})^TAx^{(0)}$$

$$= f(x^{(0)}) + (x-x^{(0)})^T A x^{(0)} + (x-x^{(0)})^T A^T x^{(0)}$$
$$+ (x-x^{(0)})^T A (x-x^{(0)})$$
$$= f(x^{(0)}) + (x-x^{(0)})^T (A+A^T) x^{(0)} + (x^{(0)})^T A x^{(0)}$$

Hence, $\nabla f(x) = (A+A^T)x$

Furthermore, $\nabla f(x) = (A+A^T)(x-x^{(0)}+x^{(0)})$
$$= (A+A^T)(x-x^{(0)}) + (A+A^T)x^{(0)}$$
$$= \nabla f(x^{(0)}) + (A+A^T)(x-x^{(0)})$$

Hence $\nabla^2 f(x) = A+A^T$

(d) $f(X) = b^T(X-X^{(0)}+X^{(0)})^T(X-X^{(0)}+X^{(0)})c$
$$= b^T(X-X^{(0)})^T(X-X^{(0)}+X^{(0)})c + b^T(X^{(0)})^T(X-X^{(0)}+X^{(0)})c$$
$$= b^T(X-X^{(0)})^T(X-X^{(0)})c + b^T(X-X^{(0)})^T X^{(0)} c$$
$$+ b^T(X^{(0)})^T(X-X^{(0)})c + b^T(X^{(0)})^T X^{(0)} c$$
$$= f(X^{(0)}) + b^T(X-X^{(0)})^T X^{(0)} c + b^T(X^{(0)})^T(X-X^{(0)})c$$
$$+ b^T(X-X^{(0)})^T(X-X^{(0)})c$$
$$= f(X^{(0)}) + \langle X^{(0)}c, (X-X^{(0)})b\rangle + \langle X^{(0)}b, (X-X^{(0)})c\rangle$$
$$+ \langle (X-X^{(0)})b, (X-X^{(0)})c\rangle$$
$$= f(X^{(0)}) + \langle X^{(0)}cb^T, X-X^{(0)}\rangle + \langle X^{(0)}bc^T, X-X^{(0)}\rangle$$
$$+ \langle (X-X^{(0)})b, (X-X^{(0)})c\rangle$$
$$= f(X^{(0)}) + \langle X^{(0)}(cb^T+bc^T), X-X^{(0)}\rangle + o(X-X^{(0)})$$

Hence, $\nabla f(X) = X(cb^T+bc^T)$

Furthermore, $\nabla f(X) = (X-X^{(0)}+X^{(0)})(cb^T+bc^T)$
$$= (X-X^{(0)})(cb^T+bc^T) + X^{(0)}(cb^T+bc^T)$$
$$= \nabla f(X^{(0)}) + (X-X^{(0)})(cb^T+bc^T)$$

Hence $\nabla^2 f(X)(Y) = Y(cb^T+bc^T)$

(e) $f(X) = \mathrm{trace}\,(XAXB)$

$$= \langle A^T X^T, XB \rangle$$

$$= \langle A^T (X - X_0 + X_0)^T, (X - X_0 + X_0) B \rangle$$

$$= \langle A^T (X - X_0)^T + A^T X_0^T, (X - X_0) B + X_0 B \rangle$$

$$= \langle A^T (X - X_0)^T, (X - X_0) B + X_0 B \rangle + \langle A^T X_0^T, (X - X_0) B + X_0 B \rangle$$

$$= \langle A^T (X - X_0)^T, (X - X_0) B \rangle + \langle A^T (X - X_0)^T, X_0 B \rangle$$
$$+ \langle A^T X_0^T, (X - X_0) B \rangle + \langle A^T X_0^T, X_0 B \rangle$$

$$= f(X_0) + \langle A^T (X - X_0)^T, X_0 B \rangle + \langle A^T X_0^T, (X - X_0) B \rangle + \langle A^T (X - X_0)^T, (X - X_0) B \rangle$$

$$= f(X_0) + \mathrm{trace}((X - X_0) A X_0 B\rangle + \mathrm{trace}(X_0 A (X - X_0) B\rangle + \langle A^T (X - X_0)^T, (X - X_0) B \rangle$$

Note that

$$\mathrm{trace}(X_0 A (X - X_0) B) = \mathrm{trace}(B X_0 A (X - X_0)) = \mathrm{trace}((X - X_0) B X_0 A)$$

Hence $f(X) = f(X_0) + \mathrm{trace}((X - X_0) A X_0 B) + \mathrm{trace}((X - X_0) B X_0 A) + \langle A^T (X - X_0)^T, (X - X_0) B \rangle$

$$= f(X_0) + \mathrm{trace}((X - X_0)(A X_0 B + B X_0 A)) + \langle A^T (X - X_0)^T, (X - X_0) B \rangle$$

Hence $\nabla f(X) = (AXB + BXA)^T = B^T X^T A^T + A^T X^T B^T$

Since $X \mapsto \nabla f(X)$ is a bounded linear mapping,

$$\nabla^2 f(X)(Y) = B^T Y^T A^T + A^T Y^T B^T$$