Non-convex Optimization for Analyzing Big Data

Assignment 2

Zhiwei Han

May 22, 2016

Task 1

1.1

Assume that $\mathbf{A}$ is symmetric.

(1) conjugate symmetry
$\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbf{A}} = tr(\mathbf{XAY^T}) = tr((\mathbf{XAY^T})^{\mathbf{T}})) = tr(\mathbf{YAX^T}) = \langle \mathbf{Y}, \mathbf{X} \rangle_{\mathbf{A}} = \overline{\langle \mathbf{Y}, \mathbf{X} \rangle_{\mathbf{A}}}$

(2) non-negativity
$\langle \mathbf{X}, \mathbf{X} \rangle_{\mathbf{A}} = tr(\mathbf{XAX^T}) = \sum_{i=1}^{m} \mathbf{x_i^T A x_i} > 0, \mathbf{X} \neq \mathbf{0}$
$\langle \mathbf{X}, \mathbf{X} \rangle_{\mathbf{A}} = 0 \quad \text{iff} \quad \mathbf{X} = \mathbf{0}$

(3) linearity in the first argument
$\langle \alpha\mathbf{X} + \beta\mathbf{Y}, \mathbf{Z} \rangle_{\mathbf{A}} = tr((\alpha\mathbf{X} + \beta\mathbf{Y})\mathbf{AZ^T}) = tr(\alpha\mathbf{XAZ^T} + \beta\mathbf{YAZ^T})$
$= \alpha\langle \mathbf{X}, \mathbf{Z} \rangle_{\mathbf{A}} + \beta\langle \mathbf{Y}, \mathbf{Z} \rangle_{\mathbf{A}}$

1.2

First it's easy to check that $H_1 \oplus H_2$ is an Abelian group together with scalar product over field $\mathbb{Z}$ which satisfies the corresponding axioms. So $H_1 \oplus H_2$ is a vector space.
Next show that $\langle \cdot, \cdot \rangle_{H_1 \oplus H_2}$ is an inner product.

(1) conjugate symmetry
$\langle (x_1, x_2), (y_1, y_2) \rangle_{H_1 \oplus H_2} = \langle x_1, y_1 \rangle_{H_1} + \langle x_2, y_2 \rangle_{H_2} = \overline{\langle y_1, x_1 \rangle}_{H_1} + \overline{\langle y_2, x_2 \rangle}_{H_2}$
$= \overline{\langle (y_1, y_2), (x_1, x_2) \rangle}_{H_1 \oplus H_2}$

(2) non-negativity
$\langle (x_1, x_2), (x_1, x_2) \rangle_{H_1 \oplus H_2} = \langle x_1, x_1 \rangle_{H_1} + \langle x_2, x_2 \rangle_{H_2} \geq 0$
$\langle (x_1, x_2), (x_1, x_2) \rangle_{H_1 \oplus H_2} = 0 \quad \text{iff} \quad (x_1, x_2) = (0, 0)$

(3) linearity in the first argument
$\langle \alpha(x_1, x_2) + \beta(y_1, y_2), (z_1, z_2) \rangle_{H_1 \oplus H_2} = \langle \alpha x_1 + \beta y_1, z_1 \rangle_{H_1} + \langle \alpha x_2 + \beta y_2, z_2 \rangle_{H_2}$
$= \alpha(\langle x_1, z_1 \rangle_{H_1} + \langle x_2, z_2 \rangle_{H_2}) + \beta(\langle y_1, z_1 \rangle_{H_1} + \langle y_2, z_2 \rangle_{H_2})$
$= \alpha\langle (x_1, x_2), (z_1, z_2) \rangle_{H_1 \oplus H_2} + \beta\langle (y_1, y_2), (z_1, z_2) \rangle_{H_1 \oplus H_2}$

The orthogonal complement of $S = \{(x_1, 0) : x_1 \in H_1\}$ is $\{(0, x_2) : x_2 \in H_2\}$.

proof: Let $S^\perp = \{(x, y) \in H_1 \oplus H_2 : \langle (x, y), (x_1, 0)\rangle_{H_1 \oplus H_2} = 0,\ \forall x_1 \in H_1\}$,
$S_1 = \{(0, x_2) : x_2 \in H_2\}$, show that $S_1 = S^\perp$.

(1) $S_1 \subseteq S^\perp$
$\forall (0, x_2) \in H_2, \quad \langle (x_1, 0), (0, x_2)\rangle_{H_1 \oplus H_2} = \langle x_1, 0\rangle_{H_1} + \langle 0, x_2\rangle_{H_2},\ \forall x_1 \in H_1$

(2) $S^\perp \subseteq S_1$
$\forall (x, y) \in H_1 \oplus H_2$, since $\langle (x, y), (x_1, 0)\rangle_{H_1 \oplus H_2} = \langle x, x_1\rangle_{H_1} + \langle y, 0\rangle_{H_2} =$
$\langle x, x_1\rangle_{H_1} = 0,\ \forall x_1 \in H_1$, it must hold that $x = 0$

1.3

a)
P_2 forms an Abelian multiplicative group and this group together with scalar product forms a vector space.

(1) conjugate symmetry
$\langle p, q\rangle = \sum_{i=-1}^{1} p(i)q(i) = \sum_{i=-1}^{1} q(i)p(i) = \overline{\langle q, p\rangle}$

(2) non-negativity
Suppose $p(x) = c_0 + c_1 x + c_2 x^2$,
$\langle p, q\rangle = (c_0 - c_1 + c_2)^2 + c_0^2 + (c_0 + c_1 + c_2)^2 \geq 0$
$\langle p, p\rangle = 0 \quad \text{iff} \quad c_0 = c_1 = c_2 = 0$, i.e. $p = 0$

(3) linearity in the first argument
$\langle \alpha p + \beta q, r\rangle = \sum_{i=-1}^{1} (\alpha p + \beta q)(i)r(i) = \sum_{i=-1}^{1} (\alpha p(i) + \beta q(i))r(i)$
$= \alpha \sum_{i=-1}^{1} p(i)r(i) + \beta \sum_{i=-1}^{1} q(i)r(i) = \alpha\langle p, r\rangle + \beta\langle q, r\rangle$

b)
Suppose $p \in P_3,\ p(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$
$\langle p, p\rangle = (c_0 - c_1 + c_2 - c_3)^2 + c_0^2 + (c_0 + c_1 + c_2 + c_3)^2 = 0 \nRightarrow c_0 = c_1 = c_2 = c_3 = 0$
Hence $\langle p, p\rangle = 0 \nRightarrow p = 0$, the inner product defined in (2) is not an inner product for P_3.

Task 2

2.1

First show that the directional derivatives exist for every $(u, w) \in \mathbb{R}^2$ at $(0, 0)$.

(1) $w = 0$
$\frac{d}{dt}|_{t=0} f(0 + tu, 0 + tw) = \frac{d}{dt}|_{t=0} f(tu, 0) = u$

(2) $w > 0$
$\frac{d}{dt}|_{t=0^+} f(0 + tu, 0 + tw) = \lim_{t\to 0^+} \frac{f(0+tu,0+tw)-f(0,0)}{t} = \lim_{t\to 0^+} \frac{t\sqrt{u^2+w^2}}{t} = \sqrt{u^2 + w^2}$
$\frac{d}{dt}|_{t=0^-} f(0+tu, 0+tw) = \lim_{t\to 0^-} \frac{f(0+tu,0+tw)-f(0,0)}{t} = \lim_{t\to 0^-} \frac{-|t|\sqrt{u^2+w^2}}{t} = \sqrt{u^2 + w^2}$
$\frac{d}{dt}|_{t=0} f(0 + tu, 0 + tw) = \sqrt{u^2 + w^2}$

(3) $w < 0$
Similarly $\frac{d}{dt}|_{t=0} f(0 + tu, 0 + tw) = -\sqrt{u^2 + w^2}$

Hence every directional derivative of f exists at $(0, 0)$.

$$\frac{d}{dt}|_{t=0} f(0 + tu, 0 + tw) = \begin{cases} u, & w = 0 \\ \sqrt{u^2 + w^2}, & w > 0 \\ -\sqrt{u^2 + w^2}, & w < 0 \end{cases}$$

Next show that f is not differentiable at $(0, 0)$.

Let $x_0 = (0, 0)$, $v_1 = (u, w)$, $v_2 = (u, -w)$, $u, w \in \mathbb{R}$, $u, w > 0$. Suppose f is differentiable at x_0 with derivative $Df(x_0)$. The directional derivative of f at x_0 along $v_1 + v_2 = (2u, 0)$ is

$$D_{v_1+v_2} f(x_0) = 2u > 0$$

However, using the linearity of $Df(x_0)$ we can derive that

$$D_{v_1+v_2} f(x_0) = Df(x_0)[v_1 + v_2] = Df(x_0)[v_1] + Df(x_0)[v_2]$$
$$= D_{v_1} f(x_0) + D_{v_2} f(x_0) = \sqrt{u^2 + w^2} - \sqrt{u^2 + w^2} = 0$$

which leads to a contradiction. So f is not differentiable at $x_0 = (0, 0)$

2.2

$f(x, y)$ is continuous on $\mathbb{R}^2 \backslash (0, 0)$, so it suffice to prove that it is also continuous at $(0, 0)$. At $(0, 0)$, $\forall \epsilon > 0$ exists an open neighbourhood $\{(x, y) | \sqrt{x^2 + y^2} < 2\epsilon\}$ such that

$$|f(x, y) - f(0, 0)| = |\tfrac{xy}{\sqrt{x^2+y^2}} - 0| \leq \tfrac{\frac{1}{2}(x^2+y^2)}{\sqrt{x^2+y^2}} < \epsilon$$

So f is continuous on $\mathbb{R}^2$.

$$\frac{\partial f}{\partial x} = \begin{cases} \frac{\partial}{\partial x} f(x, 0), \ y = 0 \\ \frac{\partial}{\partial x} f(x, y), \ y \neq 0 \end{cases} = \begin{cases} 0, & y = 0 \\ y^3 (x^2 + y^2)^{-\frac{3}{2}}, & y \neq 0 \end{cases}$$

Similarly $\frac{\partial f}{\partial y} = \begin{cases} 0, & x = 0 \\ x^3 (x^2 + y^2)^{-\frac{3}{2}}, & x \neq 0 \end{cases}$

So f is partially differentiable.

Finally we show that f is not differentiable at $(0, 0)$ so it cannot be differentiable. Suppose f is differentiable at $(0, 0)$, then

$$\nabla f(0, 0) = \begin{bmatrix} \frac{\partial f}{\partial x}|_{(0,0)} \\ \frac{\partial f}{\partial y}|_{(0,0)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\lim_{(x,y)\to(0,0)} \frac{f(x,y)-f(0,0)-\nabla^T f(0,0)(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right])}{\|\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]-\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]\|_2}$
$= \lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$

If we take sequence $\{(x_n, y_n)\} = \{(\frac{1}{n}, \frac{1}{n})\} \to (0,0)$, then

$$\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2} = \frac{1}{2} \neq 0$$

So f cannot be well-approximated by a linear map at $(0,0)$, hence it is not differentiable at $(0,0)$.

2.3

Choose inner product $\langle \mathbf{A}, \mathbf{B} \rangle = tr(\mathbf{B^H A})$

$\frac{d}{dt}|_{t=0} f(\mathbf{U}, \mathbf{V} + t\mathbf{H})$

$= \frac{d}{dt}|_{t=0}[\,\langle \mathbf{U}(\mathbf{V}+t\mathbf{H})^\mathbf{T} - \mathbf{Y}, \mathbf{U}(\mathbf{V}+t\mathbf{H})^\mathbf{T} - \mathbf{Y}\rangle + \frac{\lambda}{2}\langle \mathbf{U}, \mathbf{U}\rangle + \frac{\lambda}{2}\langle \mathbf{V}+t\mathbf{H}, \mathbf{V}+t\mathbf{H}\rangle\,]$

$= 2\langle \mathbf{UH^T}, \mathbf{UV^T} - \mathbf{Y}\rangle + \lambda\langle \mathbf{H}, \mathbf{V}\rangle$

$= 2\langle \mathbf{H}, (\mathbf{UV^T} - \mathbf{Y})^\mathbf{T}\mathbf{U}\rangle + \lambda\langle \mathbf{H}, \mathbf{V}\rangle$

$= \langle \mathbf{H}, \mathbf{2}(\mathbf{UV^T} - \mathbf{Y})^\mathbf{T}\mathbf{U} + \lambda\mathbf{V}\rangle$

2.4

$D_{[u,v,w]}\mathbf{f}(x_0, y_0, z_0) = \frac{d}{dt}|_{t=0}\mathbf{f}(x_0+tu, y_0+tv, z_0+tw)$

$= \begin{bmatrix} \frac{d}{dt}|_{t=0}[e^{x_0+tu+z_0+tw} + (x_0+tu)^2(y_0+tv)] \\ \frac{d}{dt}|_{t=0}[(y_0+tv)^2 + (z_0+tw)^2 + (x_0+tu)] \end{bmatrix}$

$= \begin{bmatrix} e^{x_0+z_0}(u+w) + 2ux_0y_0 + vx_0^2 \\ 2vy_0 + 2wz_0 + u \end{bmatrix}$

$= \begin{bmatrix} e^{x_0+z_0} + 2x_0y_0 & x_0^2 & e^{x_0+z_0} \\ 1 & 2y_0 & 2z_0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$

$$D\mathbf{f}(x_0, y_0, z_0) = \begin{bmatrix} e^{x_0+z_0} + 2x_0y_0 & x_0^2 & e^{x_0+z_0} \\ 1 & 2y_0 & 2z_0 \end{bmatrix}$$

It's easier to to obtain $D\mathbf{f}$ by calculating the Jacobian of $\mathbf{f}$.

$$D\mathbf{f}(x_0, y_0, z_0) = \mathbf{J}|_{x_0,y_0,z_0} = [\frac{d\mathbf{f}}{dx}\ \frac{d\mathbf{f}}{dy}\ \frac{d\mathbf{f}}{dz}]|_{(x_0,y_0,z_0)} = \begin{bmatrix} e^{x_0+z_0} + 2x_0y_0 & x_0^2 & e^{x_0+z_0} \\ 1 & 2y_0 & 2z_0 \end{bmatrix}$$

2.5

Consider the directional derivative of $\mathbf{f}$ along direction $\mathbf{V} \in \mathbb{R}^{m\times n}$.

$\frac{d}{dt}|_{t=0}\mathbf{f}(\mathbf{X} + t\mathbf{V})$

$= \frac{d}{dt}|_{t=0}[(\mathbf{X} + t\mathbf{V})\mathbf{A}(\mathbf{X} + t\mathbf{V})^{\mathbf{T}} - (\mathbf{X} + t\mathbf{V})\mathbf{B}]$

$= \mathbf{VAX^T} + \mathbf{XAV^T} - \mathbf{VB}$

$D\mathbf{f} : \mathbb{R}^{m\times n} \to \mathcal{L}(\mathbb{R}^{m\times n}, \mathbb{R}^{m\times m}),\ \mathbf{X} \mapsto D\mathbf{f}(\mathbf{X})$

$D\mathbf{f}(\mathbf{X})[\,\cdot\,] = [\,\cdot\,]\mathbf{AX^T} + \mathbf{XA}[\,\cdot\,]^{\mathbf{T}} - [\,\cdot\,]\mathbf{B}$

Check that $D\mathbf{f}$ is indeed the differential map.

$\lim_{\mathbf{X}\to\mathbf{X}_0} \frac{\mathbf{f}(\mathbf{X}) - \mathbf{f}(\mathbf{X}_0) - D\mathbf{f}(\mathbf{X}_0)[\mathbf{X} - \mathbf{X}_0]}{\|\mathbf{X} - \mathbf{X}_0\|}$

$= \lim_{\mathbf{X}\to\mathbf{X}_0} \frac{(\mathbf{XAX^T} - \mathbf{XB}) - (\mathbf{X}_0\mathbf{AX}_0^{\mathbf{T}} - \mathbf{X}_0\mathbf{B}) - [(\mathbf{X} - \mathbf{X}_0)\mathbf{AX}_0^{\mathbf{T}} + \mathbf{X}_0\mathbf{A}(\mathbf{X} - \mathbf{X}_0)^{\mathbf{T}} - (\mathbf{X} - \mathbf{X}_0)\mathbf{B}]}{\|\mathbf{X} - \mathbf{X}_0\|}$

$= \lim_{\mathbf{X}\to\mathbf{X}_0} \frac{(\mathbf{X} - \mathbf{X}_0)\mathbf{AX^T} - (\mathbf{X} - \mathbf{X}_0)\mathbf{AX}_0^{\mathbf{T}}}{\|\mathbf{X} - \mathbf{X}_0\|}$

$= \mathbf{0}$