MSBD HW3 Solution

1. This question is about the inner product representation of bounded linear functions.

   (a) Consider the function $E_{st} : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined by $E_{st}(\boldsymbol{X}) = x_{st}$, where $\boldsymbol{X} = [x_{ij}]_{i,j=1}^n$, i.e., E_{st} obtains the (s,t)-entry of a matrix. Find a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ such that $E_{st}(\boldsymbol{X}) = \langle \boldsymbol{A}, \boldsymbol{X}\rangle$ for all $\boldsymbol{X} \in \mathbb{R}^{n\times n}$.

   (b) Consider the trace function $\mathrm{Tr} : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined by $\mathrm{Tr}(\boldsymbol{X}) = \sum_{i=1}^n x_{ii}$, where $\boldsymbol{X} = [x_{ij}]_{i,j=1}^n$. Find a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ such that $\mathrm{Tr}(\boldsymbol{X}) = \langle \boldsymbol{A}, \boldsymbol{X}\rangle$ for all $\boldsymbol{X} \in \mathbb{R}^{n\times n}$.

   (c) Given $\boldsymbol{a} \in \mathbb{R}^n$, consider the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\boldsymbol{x}) = |\langle \boldsymbol{a}, \boldsymbol{x}\rangle|^2$ for any $\boldsymbol{x} \in \mathbb{R}^n$. Obviously f is NOT linear. Nevertheless, we can convert it to a linear function on the "lifted" matrix $\boldsymbol{x}\boldsymbol{x}^T \in \mathbb{R}^{n\times n}$. More precisely, there exists a linear function $F : \mathbb{R}^{n\times n} \to \mathbb{R}$ satisfying $f(\boldsymbol{x}) = F(\boldsymbol{x}\boldsymbol{x}^T)$. Find the inner product representation of F (i.e., find $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ such that $f(\boldsymbol{x}) = F(\boldsymbol{x}\boldsymbol{x}^T) = \langle \boldsymbol{A}, \boldsymbol{x}\boldsymbol{x}^T\rangle$.) (*This "lifting" technique is quite useful in, e.g., imaging and signal processing, machine learning.*)

Solution: (a) Let $A = (a_{ij})_{i,j=1}^n$.

Note that $\langle A, X\rangle = \sum_{ij} a_{ij}x_{ij}$.

In order to make $E_{st}(X) = \langle A, X\rangle$, A should satisfy

$$\sum_{ij} a_{ij}x_{ij} = x_{st}. \qquad (*)$$

Since X is an arbitrary matrix in $\mathbb{R}^{n\times n}$, we choose

$$x_{ij} = \begin{cases} 1, & i=s,\ j=t \\ 0, & \text{otherwise} \end{cases}$$

From $(*)$, we get $a_{st} = x_{st} = 1$

For any $i_0 \neq s$, $j_0 \neq t$, we choose

$$x_{ij} = \begin{cases} 1, & i=i_0,\ j=j_0 \\ 0, & \text{otherwise} \end{cases}$$

and substitute it into $(*)$, we have

$$a_{i_0 j_0} = 0$$

Therefore, $A = (a_{ij})_{i,j=1}^n$ with

$$a_{ij} = \begin{cases} 1, & i=s,\ j=t \\ 0, & \text{otherwise} \end{cases}$$

(b) Let $A = (a_{ij})_{i,j=1}^{n}$. For any $X \in \mathbb{R}^{n\times n}$

$Tr(X) = \langle A, X\rangle$ implies

$$\sum_i x_{ii} = \sum_{i,j} a_{ij} x_{ij} \quad \cdots\cdots \quad (**)$$

$$= \sum_{i=j} a_{ii} x_{ii} + \sum_{i\neq j} a_{ij} x_{ij}.$$

Choose $x_{ij} = \begin{cases} 1, & i=j=1 \\ 0, & \text{otherwise} \end{cases}$

We have $1 = x_{11} = a_{11}$ from $(**)$

Similarly, for any integer $k = 2, \cdots, n$, if we choose

$x_{ij} = \begin{cases} 1, & i=j=k \\ 0, & \text{otherwise} \end{cases}$

we get $a_{kk} = x_{kk} = 1$ from $(**)$

For any i_0, j_0 and $i_0 \neq j_0$, we choose

$x_{ij} = \begin{cases} 1, & i=i_0,\ j=j_0 \\ 0, & \text{otherwise} \end{cases}$

then we get $0 = \sum_i x_{ii} = a_{i_0j_0} x_{i_0j_0}$

that is $a_{i_0j_0} = 0$.

Altogether, if we choose

$a_{ij} = \begin{cases} 1, & i=j \\ 0, & \text{otherwise} \end{cases}$

We have $tr(X) = \langle A, X\rangle$

(c) Note that $xx^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} (x_1, \cdots, x_n)$

$$= \begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 \end{pmatrix} = (x_ix_j)_{i,j=1}^{n}$$

We have $\langle A, xx^T \rangle = \sum_{i,j} A_{ij} x_i x_j$.

On the other hand,

$$f(x) = |\langle a, x \rangle|^2 = |\sum_i a_i x_i|^2$$
$$= \sum_i a_i x_i \cdot \sum_j a_j x_j$$
$$= \sum_{i,j} a_i a_j x_i x_j$$

In order to let $f(x) = \langle A, xx^T \rangle$ for any $x \in \mathbb{R}^n$,

we have $\sum_{i,j} a_{ij} x_i x_j = \sum_{i,j} a_i a_j x_i x_j$, for any $x \in \mathbb{R}^n$

That is, we need to choose

$$a_{ij} = a_i a_j$$

2. Let V be a Hilbert space. Let S_1 and S_2 be two hyperplanes in V defined by

$$S_1 = \{\boldsymbol{x} \in V \mid \langle \boldsymbol{a}_1, \boldsymbol{x} \rangle = b_1\}, \quad S_2 = \{\boldsymbol{x} \in V \mid \langle \boldsymbol{a}_2, \boldsymbol{x} \rangle = b_2\}.$$

Let $\boldsymbol{y} \in V$ be given. We consider the projection of $\boldsymbol{y}$ onto $S_1 \cap S_2$, i.e., the solution of

$$\min_{\boldsymbol{x} \in S_1 \cap S_2} \|\boldsymbol{x} - \boldsymbol{y}\|. \tag{1}$$

(a) Prove that $S_1 \cap S_2$ is a plane, i.e., if $\boldsymbol{x}, \boldsymbol{z} \in S_1 \cap S_2$, then $(1+t)\boldsymbol{z} - t\boldsymbol{x} \in S_1 \cap S_2$ for any $t \in \mathbb{R}$.

(b) Prove that $\boldsymbol{z}$ is a solution of (1) if and only if $\boldsymbol{z} \in S_1 \cap S_2$ and

$$\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{x} \rangle = 0, \quad \forall \boldsymbol{x} \in S_1 \cap S_2. \tag{2}$$

(c) Find an explicit solution of (1).

(d) Prove the solution found in part (c) is unique.

Solution: (a) Proof: Since $x, z \in S_1 \cap S_2$, then

$\langle a_1, x \rangle = b_1$, $\langle a_2, x \rangle = b_2$

$\langle a_1, z \rangle = b_1$, $\langle a_2, z \rangle = b_2$

Hence, $\langle a_1, (1+t)z - tx \rangle$

$$= (1+t)\langle a_1, z \rangle - t\langle a_1, x \rangle$$
$$= (1+t)b_1 - tb_1$$
$$= b_1$$

That is, $(1+t)z - tx \in S_1$

Furthermore, $\langle a_2, (1+t)z - tx \rangle$

$$= (1+t)\langle a_2, z\rangle - t\langle a_2, x\rangle$$
$$= (1+t)b_2 - tb_2$$
$$= b_2$$

That is, $(1+t)z - tx \in S_2$

Therefore, $(1+t)z - tx \in S_1 \cap S_2$

(b) If $z \in S_1 \cap S_2$ and $\langle z-y, z-x\rangle = 0$ for $\forall x \in S_1 \cap S_2$, then for any $x \in S_1 \cap S_2$,

$$\|x-y\|^2 = \|x-z+z-y\|^2$$
$$= \|x-z\|^2 + 2\langle x-z, z-y\rangle + \|z-y\|^2$$
$$= \|x-z\|^2 - 2\langle z-x, z-y\rangle + \|z-y\|^2$$
$$= \|x-z\|^2 + \|z-y\|^2$$
$$\geq \|z-y\|^2$$

Hence $z = \underset{x \in S_1 \cap S_2}{\arg\min} \|x-y\|^2$

On the other hand, if z is the solution of $\min_{x \in S_1 \cap S_2} \|x-y\|_2$, we can prove $\langle z-y, z-x\rangle = 0$

for any $x \in S_1 \cap S_2$.

Actually, for $x \in S_1 \cap S_2$ and $z = \underset{x \in S_1 \cap S_2}{\arg\min} \|x-y\|^2$, we have $(1+t)z - tx \in S_1 \cap S_2$ and

$$\|z-y\|^2 \leq \|(1+t)z - tx - y\|^2 = \|z-y + t(z-x)\|^2$$
$$= \|z-y\|^2 + 2t\langle z-y, z-x\rangle + t^2\|z-x\|^2$$

Hence $t\langle z-y, z-x\rangle \geq -\frac{t^2}{2}\|z-x\|^2$

when $t > 0$, $\langle z-y, z-x\rangle \geq -\frac{t}{2}\|z-x\|^2$

Letting $t \to 0+$ gives $\langle z-y, z-x \rangle \geq 0$, $\forall x \in S_1 \cap S_2$

when $t<0$, $\langle z-y, z-x \rangle \leq -\frac{t}{2}\|z-x\|^2$

Letting $t \to 0-$ gives $\langle z-y, z-x \rangle \leq 0$ $\forall x \in S_1 \cap S_2$

So z satisfies $\langle z-y, z-x \rangle = 0$

(c) Let $z \in S_1 \cap S_2$, $x \in S_1 \cap S_2$, then

$$\langle z, a_1 \rangle = b_1, \quad \langle x, a_1 \rangle = b_1$$

$$\langle z, a_2 \rangle = b_2 \quad \langle x, a_2 \rangle = b_2$$

Hence $\langle z-x, a_1 \rangle = 0$ (1)

$\langle z-x, a_2 \rangle = 0$. (2)

If z is a solution to $\min\limits_{x \in S_1 \cap S_2} \|x-y\|$, then

$$\langle z-y, z-x \rangle = 0$$

Together with (1) and (2), we have

$$z-y \in \text{span}\{a_1, a_2\}.$$

Let $z = y + \alpha a_1 + \beta a_2$, $\alpha, \beta \in \mathbb{R}$.

Then $\begin{cases} \langle z, a_1 \rangle = b_1 \\ \langle z, a_2 \rangle = b_2 \end{cases}$. That is,

$$\begin{cases} \langle y+\alpha a_1+\beta a_2, a_1 \rangle = b_1 \\ \langle y+\alpha a_1+\beta a_2, a_2 \rangle = b_2 \end{cases}$$

$$\alpha \|a_1\|^2 + \beta \langle a_2, a_1 \rangle = b_1 - \langle y, a_1 \rangle$$

$$\alpha \langle a_1, a_2 \rangle + \beta \|a_2\|^2 = b_2 - \langle y, a_2 \rangle$$

$$\Rightarrow \begin{cases} \alpha^* = \dfrac{\begin{vmatrix} b_1 - \langle y, a_1 \rangle & \langle a_2, a_1 \rangle \\ b_2 - \langle y, a_2 \rangle & \|a_2\|^2 \end{vmatrix}}{\begin{vmatrix} \|a_1\|^2 & \langle a_2, a_1 \rangle \\ \langle a_1, a_2 \rangle & \|a_2\|^2 \end{vmatrix}} \end{cases}$$

$$\beta^* = \frac{\begin{vmatrix} \|a_1\|^2 & b_1 - \langle y, a_1\rangle \\ \langle a_1, a_2\rangle & b_2 - \langle y, a_2\rangle \end{vmatrix}}{\begin{vmatrix} \|a_1\|^2 & \langle a_2, a_1\rangle \\ \langle a_1, a_2\rangle & \|a_2\|^2 \end{vmatrix}}$$

Then $z = y + \alpha^* a_1 + \beta^* a_2$ is the solution.

(d) Assume there are two solutions z_1 and z_2. Then

$z_1 \in S_1 \cap S_2$, $z_2 \in S_1 \cap S_2$ and

$\langle z_1 - y, z_1 - x\rangle = 0, \quad \forall x \in S_1 \cap S_2$ ①

$\langle z_2 - y, z_2 - x\rangle = 0, \quad \forall x \in S_1 \cap S_2$ ②

Choose $x = z_2$ in ①, we get

$\langle z_1 - y, z_1 - z_2\rangle = 0$ ③

Choose $x = z_1$ in ②, we get

$\langle z_2 - y, z_2 - z_1\rangle = 0$ ④

③ − ④, we get

$\langle z_1 - z_2, z_1 - z_2\rangle = 0$

$\|z_1 - z_2\|^2 = 0$

Hence $z_1 = z_2$

That is, the solution is unique.

3. Let $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ be given with $\boldsymbol{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. Assume $N < n$, and $\boldsymbol{x}_i$, $i = 1, 2, \ldots, N$, are linearly independent. Consider the ridge regression

$$\min_{\boldsymbol{a}\in\mathbb{R}^n} \sum_{i=1}^N \left(\langle \boldsymbol{a}, \boldsymbol{x}_i\rangle - y_i\right)^2 + \lambda\|\boldsymbol{a}\|_2^2,$$

where $\lambda \in \mathbb{R}$ is a regularization parameter, and we set the bias $b = 0$ for simplicity.

(a) Prove that the solution must be in the form of $\boldsymbol{a} = \sum_{i=1}^N c_i \boldsymbol{x}_i$ for some $\boldsymbol{c} = [c_1, c_2, \ldots, c_N]^T \in \mathbb{R}^N$. (*Hint: Similar to the proof of the representer theorem.*)

(b) Re-express the minimization in terms of $\boldsymbol{c} \in \mathbb{R}^N$, which has fewer unknowns than the original formulation.

Solution: Let $a = a_s + \sum_{j=1}^{N} c_j x_j$, where $a_s \perp \mathrm{span}\{x_1, \cdots, x_N\}$.

Then
$$\begin{aligned}
\langle a, x_i \rangle &= \langle a_s + \sum_{j=1}^{N} c_j x_j, x_i \rangle \\
&= \langle a_s, x_i \rangle + \langle \sum_{j=1}^{N} c_j x_j, x_i \rangle \\
&= 0 + \sum_{j=1}^{N} c_j \langle x_j, x_i \rangle \\
&= \sum_{j=1}^{N} c_j \langle x_j, x_i \rangle
\end{aligned}$$

$$\begin{aligned}
\|a\|_2^2 &= \langle a_s + \sum_{j=1}^{N} c_j x_j, a_s + \sum_{j=1}^{N} c_j x_j \rangle \\
&= \|a_s\|^2 + 2\langle \sum_{j=1}^{N} c_j x_j, a_s \rangle + \sum_{i,j} c_i c_j \langle x_i, x_j \rangle \\
&= \|a_s\|^2 + \sum_{i,j} c_i c_j \langle x_i, x_j \rangle
\end{aligned}$$

So
$$\begin{aligned}
&\sum_{i=1}^{N} (\langle a, x_i \rangle - y_i)^2 + \lambda \|a\|_2^2 \\
&= \sum_{i=1}^{N} \left(\sum_{j=1}^{N} c_j \langle x_j, x_i \rangle - y_i \right)^2 + \lambda \left[\|a_s\|^2 + \sum_{i,j} c_i c_j \langle x_i, x_j \rangle \right] \\
&\geq \sum_{i=1}^{N} \left(\sum_{j=1}^{N} c_j \langle x_j, x_i \rangle - y_i \right)^2 + \lambda \sum_{i,j} c_i c_j \langle x_i, x_j \rangle
\end{aligned}$$

Hence
$$\min_{a \in \mathbb{R}^n} \sum_{i=1}^{N} (\langle a, x_i \rangle - y_i)^2 + \lambda \|a\|_2^2$$

$$\Leftrightarrow \min_{c \in \mathbb{R}^N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} c_j \langle x_j, x_i \rangle - y_i \right)^2 + \lambda \sum_{i,j} c_i c_j \langle x_i, x_j \rangle$$

and $a_s = 0$

That is, the solution must be in the form of
$$a = \sum_{j=1}^{N} c_j x_j .$$

(b) $$\min_{c \in \mathbb{R}^N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} c_j \langle x_j, x_i \rangle - y_i \right)^2 + \lambda \sum_{i,j} c_i c_j \langle x_i, x_j \rangle$$

$$\Longleftrightarrow \min_{c \in \mathbb{R}^N} \|Kc - y\|^2 + \lambda c^T K c$$

where $K = \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \cdots & \langle x_1, x_N \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_2, x_N \rangle \\ \cdots & & & \\ \langle x_N, x_1 \rangle & \langle x_N, x_2 \rangle & \cdots & \langle x_N, x_N \rangle \end{pmatrix}$