MSBD 5004 Mathematical Methods for Data Analysis
Homework 3

Due date: 10 April, 11:59pm, Friday

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.*)

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.

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.