# Homework 5

*Posted: April 17, 2024*
*Due: May 1, 2024*

1. Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two vectors and let $\boldsymbol{a}x = \boldsymbol{b}$ be an one-indeterminate linear system. (the indeterminate is of course the scalar $x$). The system is, in general incompatible, if $m > 1$.
   Prove that the best approximation of the solution (which minimizes $\sum_{i=1}^{m}(a_i x - b_i)^2$) is:
   $$x = \frac{(\boldsymbol{a}, \boldsymbol{b})}{\parallel \boldsymbol{a} \parallel_2^2}.$$
2. Let $X \in \mathbb{R}^{m\times n}$ be a centered data sample matrix whose covariance matrix is $cov(X) = aI_n + bJ_{n,n}$. Prove that $a \geqslant 0$ and $b \geqslant 0$.
3. Let $X \in \mathbb{R}^{m\times n}$ be a centered data sample matrix and let $p, q \in \mathbb{N}$ be such that $p + q = n$ and $X = (U\ V)$, where $U \in \mathbb{R}^{m\times p}$ and $V \in \mathbb{R}^{m\times q}$. Define the *mixed covariance* of $U$ and $V$ as $cov(U, V) = \frac{1}{m-1}U'V \in \mathbb{R}^{p\times q}$. Prove that
   $$cov(X) = \begin{pmatrix} cov(U) & cov(U, V) \\ cov(U, V)' & cov(V) \end{pmatrix}.$$
4. Let $X \in \mathbb{R}^{m\times n}$ be a data matrix involving $m$ experiments and $n$ variables, $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. The result of each experiment is capured by a vector $\boldsymbol{y} \in \mathbb{R}^m$ and we to express these results as a linear combination of the columns of $X$. In other words, we seek the best approximation of $\boldsymbol{y}$ as a linear combination of the colums of $X$, that is, $X\boldsymbol{w}$. Thus $\boldsymbol{y} - X\boldsymbol{w}$ is orthogonal to all vectors in $X$. Prove that:
   (a) the optimal linear combination $X\boldsymbol{w}$ of the colums of $X$ is obtained taking $\boldsymbol{w} = (X'X)^{-1}X'\boldsymbol{y}$; thus, this linear combination is $X\boldsymbol{w} = X(X'X)^{-1}X'\boldsymbol{y}$;
   (b) if $G = X(X'X)^{-1}X'$, the matrix $G$ is symmetric and idempotent.