1 Trace Derivatives

(a) Let $P$ be a $p \times q$ matrix and $Q$ be a $q \times p$ matrix. Compute $\frac{\partial \text{trace}(PQ)}{\partial P}$.

(b) Let $P$ be a $p \times q$ matrix and $Q$ be a $q \times q$ matrix. Compute $\frac{\partial \text{trace}(PQP^\top)}{\partial P}$ at $P = U$.

2 Unitary invariance

(a) Prove that the regular Euclidean norm (also called the $\ell^2$-norm) is unitary invariant; in other words, the $\ell^2$-norm of a vector is the same, regardless of how you apply a rigid linear transformation to the vector (i.e., rotate or reflect). Note that rigid linear transformation of a vector $v \in \mathbb{R}^d$ means multiplying by an orthogonal $U \in \mathbb{R}^{d \times d}$.

(b) Now show that the Frobenius norm of matrix $A$ is unitary invariant. The Frobenius norm is defined as $\|A\|_F = \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2} = \sqrt{\text{tr}(A^\top A)}$.

3 Least Squares (using vector calculus)

(a) In ordinary least-squares linear regression, we typically have $n > d$ so that there is no $w$ such that $Xw = y$ (these are typically overdetermined systems — too many equations given the number of unknowns). Hence, we need to find an approximate solution to this problem. The residual vector will be $r = Xw - y$ and we want to make it as small as possible. The most common case is to measure the residual error with the standard Euclidean $\ell^2$-norm. So the problem becomes:

$$\min_w \|Xw - y\|_2^2$$

Where $X \in \mathbb{R}^{n \times d}$, $w \in \mathbb{R}^d$, $y \in \mathbb{R}^n$. Derive using vector calculus an expression for an optimal estimate for $w$ for this problem assuming $X$ is full rank.

(b) How do we know that $X^\top X$ is invertible?

(c) What should we do if $X$ is not full rank?