Proof of Gauss Markov Theorem

We want to find an unbiased estimator of σ^2 for the Gauss Markov Model,

$$\boldsymbol{\vec{y} = X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \left(\overrightarrow{0}, \sigma^2 I_n\right)}$$

Motivated by the expression of unbiased estimator of σ^2 in case of iid $X_1, X_2, ...X_n$ we might guess something of the form $C||\vec{y} - X\hat{\vec{\beta}}||^2$ where C is a constant.

So let us calculate the expected value of $||\vec{y} - X\hat{\vec{\beta}}||^2$

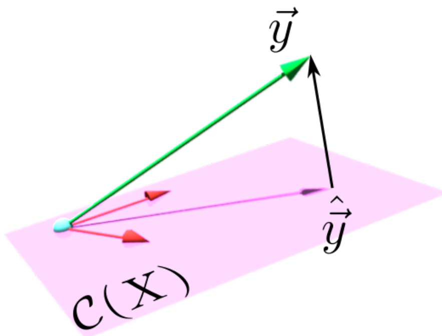

Figure 1:

We note that the expression of $\hat{\vec{\beta}}$ involves a complex combination of X, X^T and $\vec{y}$. So to calculate the expected value, we take a simpler approach. Note that the term $\vec{y} - X\hat{\vec{\beta}}$ is a vector orthogonal to $C(X)$. Thus, it is basically the projection of $\vec{y}$ on the orthogonal complement of $C(X)$.

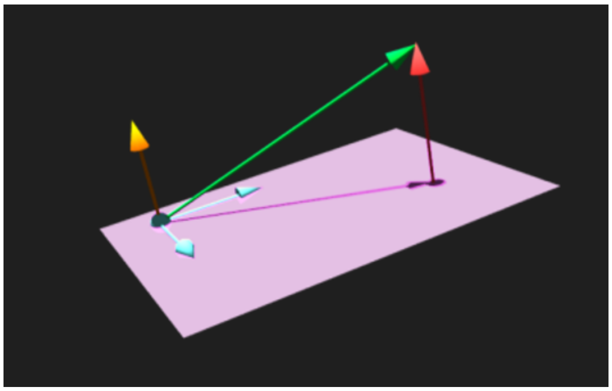

Figure 2:

It might be a good idea to look at things from an orthonormal basis, so we consider the orthonormal basis of $C(X)$. Let it be $\vec{v_1}, \vec{v_2}, \vec{v_3}.....\vec{v_k}$, where $k = rank(C(X))$. Let us extend it to an orthonormal basis of R^n by including orthonormal vectors $\vec{v_{k+1}}, \vec{v_{k+2}}, \vec{v_{k+3}}.....\vec{v_n}$.

In Fig.2 the cyan vectors are orthonormal basis of $C(X)$ and the yellow one is that of orthogonal complement of $C(X)$. One important result about orthonormal basis is that the dot product of a vector with one vector of the orthonormal basis gives the coordinate corresponding to that vector in the orthonormal basis representation of the original vector. This is because, for any vector v having representation as $\Sigma_{i=1}^{n}\alpha_i\vec{v_i}$, where $\vec{v_i}$ are orthonormal basis of R^n,

$\vec{v_i}^T\vec{v} = \alpha i||\vec{v_i}||^2 = \alpha_i$ (since $||\vec{v_i}||^2 = 1$ and $\vec{v_i}^T\vec{v_j} = 0$ for $i \neq j$)

Also note that $||\vec{v}||^2 = \Sigma_{i=1}^{n}\alpha_i^2$ (since $||\vec{v_i}||^2 = 1$ and $\vec{v_i}^T\vec{v_j} = 0$ for $i \neq j$)

Let us now go back to the problem at hand. Note that $\vec{y} - X\vec{\hat{\beta}}$ can be written as a linear combination of vectors $\vec{v_{k+1}}, \vec{v_{k+2}}, \vec{v_{k+3}}.....\vec{v_n}$. Since it belongs to orthogonal complement of $C(X)$.

Thus $||\vec{y} - X\vec{\hat{\beta}}||^2 = \Sigma_{i=k+1}^{n}\alpha_i^2$. This looks good, if we can find the Expected value of each term, we might make some progress. On that note, let us try to calculate the expected value of α_i^2.

See that $\alpha_i = \vec{v_i}^T\vec{y}$. Thus using some results about random vectors, we can easily see that

$E[\alpha_i] = E[\vec{v_i}^T\vec{y}] = \vec{v_i}^T E[\vec{y}] = \vec{v_i}^T X\vec{\beta} = 0$ (for, $k+1 \leq i \leq n$, since all these vectors are orthogonal to $C(X)$)

also, $Var[\alpha_i] = \vec{v_i}^T Var[\vec{y}]\vec{v_i} = \sigma^2\vec{v_i}^T\vec{v_i} = \sigma^2$ (since $\vec{v_i}$ are unit vectors)

Therefore $E[\alpha_i^2] = 0 + \sigma^2 = \sigma^2$ (for, $k+1 \leq i \leq n$). But we know that

$E[||\vec{y} - X\vec{\hat{\beta}}||^2] = E[\Sigma_{i=k+1}^{n}\alpha_i^2] = \Sigma_{i=k+1}^{n}E[\alpha_i^2] = (n-k)\sigma^2$.

Thus $\dfrac{||\vec{y} - X\vec{\hat{\beta}}||^2}{n-k}$, where $k = rank(C(X)$ is an unbiased estimator of σ^2 for the Gauss Markov Model.