ECE531 Screencast 6.2: The Principle of Orthogonality

D. Richard Brown III

Worcester Polytechnic Institute

Normed Vector Spaces and Euclidean Geometry Review

A "normed vector space" is a set that has some special properties like

- Closed under addition
- Closed under scalar multiplication
- etc.

and a norm satisfying certain properties like $\|u\| > 0$ unless $u = 0$, $\|\alpha u\| = |\alpha|\|u\|$ for real α, and a triangle inequality.

A well-known example of a normed vector space is $\mathbb{R}^n$. Given $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$, we can define a norm in this vector space as $\|u\| = \sqrt{u^\top u}$ and

- the squared length of a vector is given by the inner product of the vector with itself, i.e. $u^\top u$ and $v^\top v$
- the vectors u and v are orthogonal if their inner product $u^\top v = 0$.
- the subspace spanned by u and v is all possible coordinates formed by linear combinations of u and v

Vector Spaces of Random Variables

Zero-mean random variables can also be thought of as "vectors" in some vector space. Denoting $\mathcal{H}$ as the set of all zero-mean random variables, it can be shown that $\mathcal{H}$ has all the properties of a vector space, e.g. closed under addition and scalar multiplication.

Moreover, we can use expectation as a norm in this vector space. Specifically, for U and V in $\mathcal{H}$, we can say

- the squared "length" of a random variable is given by the inner product of the "vector" with itself, i.e. $\mathrm{E}[U^2] = \mathrm{var}(U)$ and $E[V^2] = \mathrm{var}(V)$
- the random variables U and V are orthogonal if their inner product $\mathrm{E}[UV] = \mathrm{cov}(U, V) = 0$.
- the subspace spanned by U and V is all possible coordinates formed by linear combinations of U and V

Geometric Interpretation of Scalar LMMSE (1/2)

First, we assume that the parameter Θ and the observations Y are zero mean. If this is not true for your model, since the means are assumed to be known, you can form a new parameter and observation model as

$$\Theta' = \Theta - \mathrm{E}[\Theta]$$
$$Y' = Y - \mathrm{E}[Y]$$

and proceed from here without loss of generality.

Under this assumption, we know $c = 0$ for our LMMSE estimator and the MSE is is only a function of A:

$$J(A) = \mathrm{E}\left[\left(A^\top Y - \Theta\right)^2\right] = \mathrm{E}\left[\epsilon^2\right] = \mathrm{var}(\epsilon)$$

Under our geometrical interpretation, the MSE is the “length” of the estimation error. We seek to find the estimator that minimizes this length.

Geometric Interpretation of Scalar LMMSE (2/2)

Now, since we are concerned here with linear estimators of the form

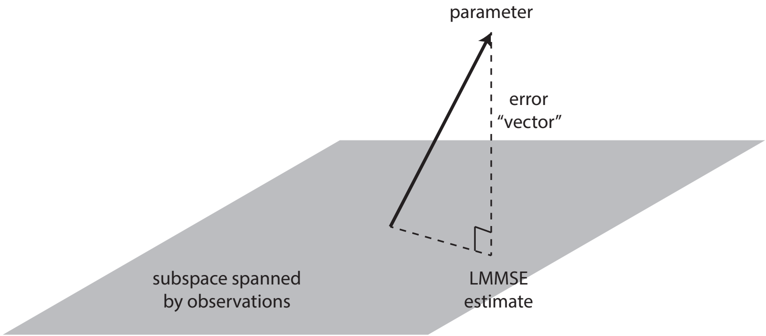
$$\hat{\theta}(Y) = A^\top Y$$

the estimate must be in the subspace spanned by the observations.

Remarks:

- If the unknown parameter is also in the subspace spanned by the observations, we can make the MSE equal to zero.
- Usually, the unknown parameter is not in the subspace spanned by the observations. How can we minimize the "length" of the estimation error in this case?

The Principle of Orthogonality: Intuition



To minimize the MSE, the estimation error "vector" must be **orthogonal** to the subspace spanned by the observations. This means the LMMSE estimator must satisfy

$$\mathrm{E}\left\{\epsilon Y_k\right\} = \mathrm{E}\left\{\left(\hat{\theta}(Y) - \Theta\right) Y_k\right\} = 0$$

for all $k = 0, \dots, n-1$.

The Principle of Orthogonality

Theorem

A linear estimator of the scalar parameter Θ is an LMMSE estimator if and only if

$$\mathrm{E}\{\hat{\theta}(Y)\} = \mathrm{E}\{\Theta\}$$

and

$$\mathrm{E}\left\{\left(\hat{\theta}(Y) - \Theta\right) Y_k\right\} = 0$$

for all k.

This result can be used to directly derive the LMMSE estimator and provides a geometric way to understand sequential LMMSE estimation. It is also often handy for solving problems related to LMMSE estimation.