Transformation of Random Vectors: Continued

White Random Vectors

A random vector $\mathbf{X}$ in an underlying sample space $\mathbf{S} \subseteq \mathbf{R}^n$ is said to be *white* if the components of the random vector are statistically independent of each other, i.e., the joint n^{th}–order PDF of the components is separable, i.e,

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

As a consequence, the components are also pairwise independent of each other, i.e.,

$$f_{X_i,X_j}(x_i, x_j) = f_{X_i}(x_i) f_{X_j}(x_j), \quad i \neq j.$$

Since the components are pairwise independent they are also pairwise uncorrelated. This implies that the covariance matrix associated with a white random vector is diagonal, i.e.,

$$\mathbf{C}_x = \mathbf{\Lambda} \equiv diag(\lambda_1, \lambda_2, \ldots \lambda_n).$$

If the uncorrelated random variables are further identically distributed, i.e., possess identical statistical characteristics then the random vector is said to be a *i.i.d* random vector. In this case the covariance matrix becomes an identity matrix:

$$\mathbf{C}_x = \sigma^2 \mathbf{I},$$

where σ is the common standard deviation of the components. A random vector is said to be *weakly white* if the components are just statistically uncorrelated. Henceforth when we refer to a white random vector it will mean white in the weak sense.

In certain digital communication applications such as Vitterbi decoding, used to remove *inter symbol interference* (ISI) introduced by a channel with finite memory, a white channel noise model is required. In some cases this may not be true and it will be necessary to whiten the noise before the application of the Vitterbi algorithm.

Whitening of a Random Vector:

If we transform a random vector $\mathbf{X}$ using a linear transformation $\mathbf{A}$ then the mean vector and covariance matrix of the transformed random vector $\mathbf{Y}$ are given by:

$$\mathbf{m}_y = \mathbf{A}\mathbf{m}_x, \quad \mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T$$

In linear algebra terminology we seek a transformation $\mathbf{A}$ that transforms the random vector $\mathbf{X}$ into another random vector $\mathbf{Y} = \mathbf{A}\mathbf{X}$ that has a diagonal covariance matrix $\mathbf{C}_y$, i.e., we require:

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T = \mathbf{D},$$

where $\mathbf{D}$ is some diagonal, positive semi–definite matrix. The solution to this whitening problem is to choose the linear transformation $\mathbf{A}$ to be equal to $\mathbf{V}^T$, where $\mathbf{V}$ is the unitary matrix of eigenvectors of the covariance matrix $\mathbf{C}_x$, i.e.,

$$\mathbf{Y} = \mathbf{V}^T\mathbf{X} \Longleftrightarrow \mathbf{C}_y = \mathbf{V}^T\mathbf{C}_x\mathbf{V} = \mathbf{\Lambda}.$$

This process of whitening the random vector $\mathbf{X}$ using the eigenvectors of its covariance matrix is also called as the *Karhunen Loeve Transform* (KLT).

IID Whitening:

If we require that the transformed vector be not only white but also i.i.d then we require:

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T = \sigma^2\mathbf{I}.$$

The solution to this i.i.d whitening problem is given by:

$$\mathbf{A} = \sigma\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^T,$$

where $\mathbf{\Lambda}^{\frac{1}{2}}$ refers to the self–adjoint matrix square–root of $\mathbf{\Lambda}$, $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues of $\mathbf{C}_x$ and $\mathbf{V}$ is the unitary matrix of eigenvectors of $\mathbf{C}_x$. The fact that the covariance of the transformed vector $\mathbf{Y}$ is identity can be verified via:

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T = \sigma^2\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{C}_x\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}} = \sigma^2\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Lambda}\mathbf{\Lambda}^{-\frac{1}{2}} = \sigma^2\mathbf{I}.$$

Geometrical Interpretation of Whitening

A unitary transformation $\mathbf{A}$ is a rotation in $\mathbf{R}^n$ if the coloums are orthonormal, i.e .,

$$\mathbf{A}^T\mathbf{A} = \mathbf{I} \Longleftrightarrow \mathbf{A}^{-1} = \mathbf{A}^T.$$

For a deterministic vector, the length of a vector remains invariant under a unitary operation, i.e.,

$$||\mathbf{y}|| = ||A\mathbf{x}|| = ||\mathbf{x}||.$$

In the space of random vectors $\mathbf{X} \in \mathbf{S} \subseteq \mathbf{R}^n$ this implies that the average power is preserved, i.e.,

$$< \mathbf{Y}, \mathbf{Y} >=< \mathbf{AX}, \mathbf{AX} >= E\{\mathbf{X}^T\mathbf{A}^T\mathbf{AX}\} =< \mathbf{X}, \mathbf{X} >= P_{\text{ave}}.$$

The KLT in the first case, i.e., the weakly white case therefore corresponds to a rotation of the random vector $\mathbf{X}$. In the second case, i.e., the i.i.d white case the KLT corresponds to first a rotation by $\mathbf{V}^T$ followed by inverse scaling of the axes by $\sigma\mathbf{\Lambda}^{-\frac{1}{2}}$. In a sense this is analogous to the process of reducing a general quadratic form to the standard quadratic form that we encounter in coordinate geometry.