1 Principal Component Analysis

In this document, we discuss Principal Component Analysis (PCA), or Empirical Orthogonal Functions (EOFs). The idea behind the method is to define a new set of variables that are linear functions of the original data, that are better suited to investigating the data with. These new "variables" are determined by applying Singular Value Decomposition to a trajectory matrix, which we now define.

We define the m dimensional trajectory matrix $X \in \mathbb{R}^m$ to be the matrix whose rows are the m dimensional unit delay time delay vectors of a scalar observation s_j. We have that i^{th} row of X is

$$\mathbf{X}_i \quad = \quad [s_{i+(m-1)} \; s_{i+(m-2)} \; \ldots \; s_i] \tag{1}$$

The entire trajectory matrix is then

$$X = \begin{pmatrix} s_{m-1} & s_{m-2} & \cdots & s_1 \\ s_m & s_{m-1} & \cdots & s_2 \\ \vdots & \vdots & \ldots & \vdots \\ s_{i+m-1} & s_{i+m-2} & \cdots & s_i \\ \vdots & \vdots & \ldots & \vdots \\ s_N & s_{N-1} & \cdots & s_{N-(m-1)} \end{pmatrix}.$$

This matrix contains every m length pattern observed in the data.

The matrix X is like any other real valued array and so we can apply singular value decomposition to it. We have

$$X \quad = \quad U \Sigma V^T. \tag{2}$$

How should we interpret the singular values and singular vectors when X is a trajectory matrix and not a linear propagator?

The SVD applied to a trajectory matrix identifies the directions $\mathbf{v}_i$ that contain the most variance as quantified by the singular value σ_i. The direction $\mathbf{v}_i$ is a linear combination of the original variables, in our case lagged values of the scalar time series. As a result we can project the data onto the singular vectors to get a new representation of the data. This new representation has the properties that the variables are ordered according to how much variance they contribute to the signal. The method identifies functional shapes in the data that are orthogonal (hence the name Empirical Orthogonal Functions). Moreover, the new variables are ordered according to their variance content and it is sometimes the case that a small fraction of the projected variables can account for a large proportion of the variance.

To see that this is the case consider

$$XV = U\Sigma. \tag{3}$$

The left hand side of equation 3 is the projection of X onto V, the array of singular vectors. On the right hand side of equation 3 we have the orthogonal matrix U and the diagonal matrix Σ. The matrix U contains the projected data (it has the same number of rows as X), but by definition the columns of U are orthogonal. The diagonal matrix Σ scales each column of $\mathbf{u}_i$ by σ_i. The share of the variance described by projecting the data onto the i^{th} principal component is given by $\sigma_i / \sum \sigma_i$.

The trajectory matrix is often singular, having many more rows than columns. We can reduce the computation burden of a singular value decomposition by considering the square of the trajectory matrix, since

$$\begin{aligned} X^{\mathrm{T}}X &= (U\Sigma V^{\mathrm{T}})^{\mathrm{T}}(U\Sigma V^{\mathrm{T}}) \\ &= V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} \\ &= V\Sigma^2 V^{\mathrm{T}} \end{aligned}$$

Hence the SVD of $X^{\mathrm{T}}X$ has the same singular vectors as X and the singular values are σ^2.

1.1 Example: Noise Ball

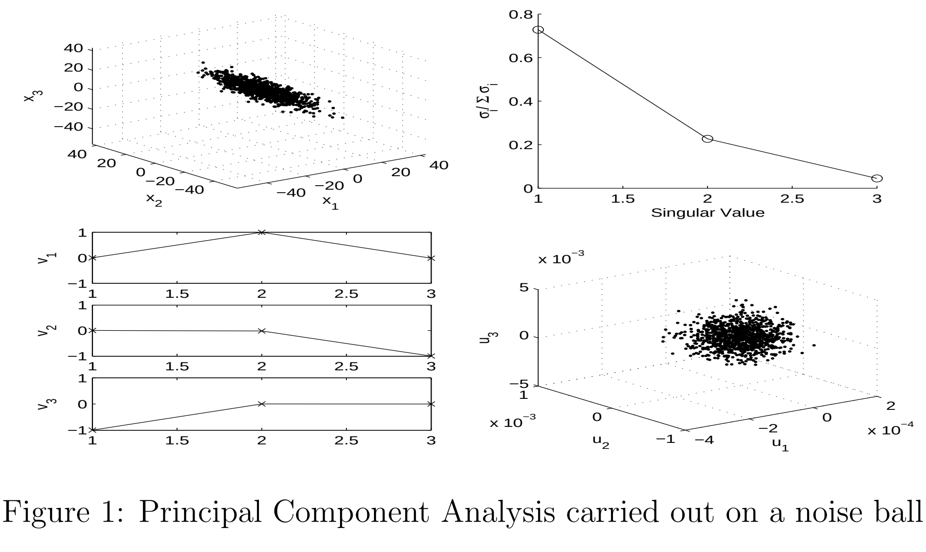

Figure 1: Principal Component Analysis carried out on a noise ball.

1.2 Example: Lorenz

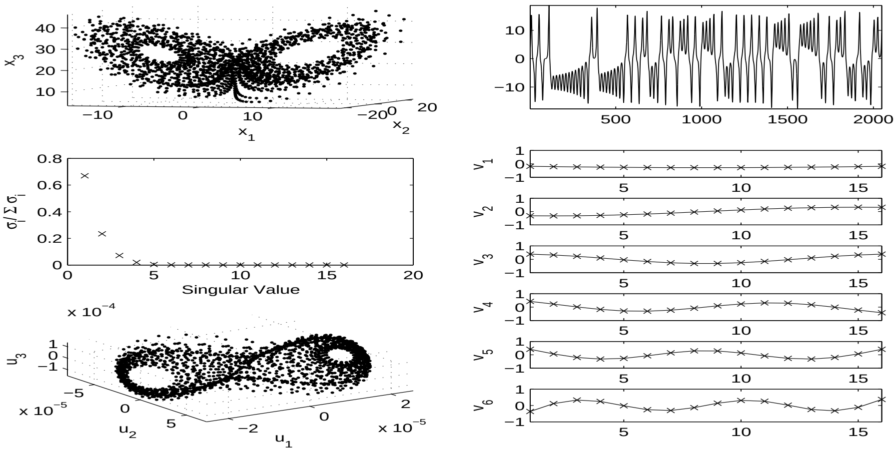

Figure 2: Principal Component Analysis carried out on the x component of Lorenz.