# Principal Component Analysis 推导

## 1° Maxium Variance Formulation

考虑一个 D 维的欧式空间的数据集$\{\mathrm{x_n}\}$.现在 PCA 的目的就是在最大化方差的情况下完成降维。现在我们假设将数据从 D 维降到一个已经给定的 M 维。

为了推导方便，先假设 M=1.我们使用一个 D 维的向量$\mathrm{u_1}$来定义这个空间的方向。为了考虑方便，我们假设这是一个单位向量，$u_1^T u_1 = 1$.那么每个数据点$\mathrm{x_n}$就会被映射到新的数值$\mathrm{u_1^T} x_n$.那么数据集的均值为$\mathrm{u_1^T}\bar{x}$，其中$\bar{x}$的定义如下：

$\bar{x} = \frac{1}{\mathrm{N}}\sum_{n=1}^{N} x_n$　而方差的定义如下：

$$\frac{1}{\mathrm{N}}\sum_{n=1}^{N}\{u_1^T x_n - u_1^T\bar{x}\}^2 = u_1^T S u_1$$

其中 S 为协方差矩阵，定义为：

$$\mathrm{S} = \frac{1}{\mathrm{N}}\sum_{n=1}^{N}(x_n - \bar{x})(x_n - \bar{x})^T$$

现在考虑最大化方差，同时还有限制条件，即$u_1^T u_1 = 1$，所以引入拉格朗日算子，即最大化以下这个方程：

$$u_1^T S u_1 + \lambda_1(1 - u_1^T u_1)$$

那么对$\mathrm{u_1}$求导得到极点值。得到条件为：

$$\mathrm{S}\mathrm{u_1} = \lambda_1 u_1$$

所以由此可得，$u_1$为 S 的特征向量，而$\lambda_1$为对应的特征值。

$$u_1^T S u_1 = \lambda_1$$

所以最大化的条件就是选择$u_1$为 S 的最大特征值所对应的特征向量。

接下来考虑 M 更大的情况。即可以在现有的 1 维的情况下再加一维，那么易推得，我们最终所选择的$u_1, u_2, \ldots, u_M$一定是协方差矩阵 S 的排序前 M 的特征值所对应的特征向量。

## 2° Minimum-Error Formulation

现在考虑 PCA 的另一个形式。即最小化信息损失。

先给出一个 D 维的完全标准正交基$\{u_i\}$且满足：

$$u_i^T u_j = \delta_{ij}$$

因为这是一个完全的正交基，所以任意的一个数据点都可以被表示为基向量的线性组合：

$$x_n = \sum_{i=1}^{D} a_{ni} u_i$$

即原来的坐标系被“旋转”至新的坐标系$\{u_i\}$，且根据标准正交基的特点，可以不失一般性地得到：

$$x_n = \sum_{i=1}^{D} (x_n^T u_i) u_i$$

那么现在 PCA 的目标是在最小化信息损失的情况下，实行降维。那么我们可以使用正交基的前 M 个向量来不失一般性地表示，所以每个数据点都可以表示为：

$$\tilde{x}_n = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i$$

其中$\{z_{ni}\}$是取决于具体的数据点的，而$\{b_i\}$是常数，且对每个数据点都是一致的。所以现在我们要决定的包括$\{u_i\}, \{z_{ni}\}, \{b_i\}$来最小化降维所带来的扭曲影响。即我们需要最小化这个方程：

$$J = \frac{1}{N} \sum_{n=1}^{N} \|x_n - \tilde{x}_n\|^2$$

首先，考虑对于$\{z_{ni}\}$最小化，得到以下条件：

$$z_{nj} = x_n^T u_j$$

其次，对于$\{\mathrm{b_i}\}$最小化方程，得到以下条件：

$$\mathrm{b_j} = \bar{x}^{\mathrm{T}} u_j$$

将以上条件代入方程得到如下形式：

$$x_n - \tilde{x}_n = \sum_{i=M+1}^{D} \{(x_n - \bar{x})^T u_i\} u_i$$

以此得到一个 J 的纯粹使用$\{\mathrm{u_i}\}$表示的形式：

$$\mathrm{J} = \frac{1}{\mathrm{N}} \sum_{n=1}^{N} \sum_{M+1}^{D} (x_n^T u_i - \bar{x}^{\mathrm{T}} u_i)^2 = \sum_{i=M+1}^{D} u_i^T S u_i$$

至此需要决定的三组参数中唯一没被考虑的就是$\{\mathrm{u_i}\}$

那么至此，问题已经和第一种角度考虑的形式很接近了，同理通过代入限制条件引入拉格朗日算子得到导数条件：

$$u_2^T S u_2 = \lambda_2$$

为了最小化 J，只需要取较小 D-M 个特征值所对应的特征向量即可。那么就对应了第一种角度考虑下的选取前 M 个较大的特征值所对应的特征向量。

即这两种角度考虑 PCA 是等价的。