Figuring Out the Principal Component Coefficients and Why it Matters

The PCs are an orthonormal transformation of the original *X*s:

$$\left\{\begin{array}{lllll} \xi_1 & = & \text{PRIN1} & = & \beta_{11}X_1+\beta_{12}X_2+\cdots+\beta_{1p}X_p \\ \xi_2 & = & \text{PRIN2} & = & \beta_{21}X_1+\beta_{22}X_2+\cdots+\beta_{2p}X_p \\ \xi_3 & = & \text{PRIN3} & = & \beta_{31}X_1+\beta_{32}X_2+\cdots+\beta_{3p}X_p \\ \vdots & & \vdots & & \vdots \\ \xi_p & = & \text{PRINp} & = & \beta_{p1}X_1+\beta_{p2}X_2+\cdots+\beta_{pp}X_p \end{array}\right\} \quad (*)$$

We seek an understanding of the coefficients of (*) in order to understand how to interpret the meaning of the PCs. We gain that insight by figuring out the formulas for the coefficients in (*). The formulas are most easily derived indirectly, through the inverse transformation:
The *X*s are the inverse orthonormal transformation of the PCs – the transpose of (*):

$$\left\{\begin{array}{lll} X_1 & = & \beta_{11}\xi_1+\beta_{21}\xi_2+\beta_{31}\xi_3+\cdots+\beta_{p1}\xi_p \\ X_2 & = & \beta_{12}\xi_1+\beta_{22}\xi_2+\beta_{32}\xi_3+\cdots+\beta_{p2}\xi_p \\ X_3 & = & \beta_{13}\xi_1+\beta_{23}\xi_2+\beta_{33}\xi_3+\cdots+\beta_{p3}\xi_p \\ \vdots & \vdots & \vdots \\ X_p & = & \beta_{1p}\xi_1+\beta_{2p}\xi_2+\beta_{3p}\xi_3+\cdots+\beta_{pp}\xi_p \end{array}\right\} \quad (**)$$

Since each equation in (**) is an exactly fitting linear combination of the PCs, each equation in (**) is the multiple linear regression of an *X* on the PCs. (There is no better fit to a response variable than a fit that has an RMSE = 0 and R-square = 1.00.) Therefore, the coefficients in (**) are the multiple linear regression coefficients for the regressions of the Xs on the PCs.[1] Formulas for the regression coefficients are easy to figure out if the predictor variables are all uncorrelated with each other. In that special case, the coefficients are the same as in a simple linear regression. The predictor variables in (**) are the PCs, which are uncorrelated with each other by the way that they are constructed.

The formula for the slope in a simple linear regression of A on B is $\beta = r_{A,B}\dfrac{s_A}{s_B}$, where $r_{A,B}$ = correlation coefficient between A and B, and s_A and s_B are the respective standard deviations. Using this in (**), and also using $Var(X_j)=1$, since X_j is standardized, and using $Var(\xi_i)=\lambda_i$ (the i^{th} eigenvalue of the correlation matrix of the *X*s), then the coefficient of ξ_i in row *j* of (**) is

$$\beta_{ij} = r_{X_j,\xi_i}\frac{s_{X_j}}{s_{\xi_i}} = r_{X_j,\xi_i}\frac{1}{\sqrt{\lambda_i}}$$

Now the coefficients of (*) are all found in (**) – just transposed. Thus row *i* in (*) becomes

$$\xi_i = \beta_{i1}X_1+\beta_{i2}X_2+\cdots+\beta_{ip}X_p = r_{X_1,\xi_i}\frac{1}{\sqrt{\lambda_i}}X_1 + r_{X_2,\xi_i}\frac{1}{\sqrt{\lambda_i}}X_2+\cdots+r_{X_p,\xi_i}\frac{1}{\sqrt{\lambda_i}}X_p$$

[1] (*) is the set of multiple linear regression equations for regressing the PCs on the *X*s. But the *X*s are correlated with each other. So the simple approach to deriving the coefficient formulas from (**) is not available from (*), although the same result will hold after all of the dust settles.

That is, the coefficients for each PC in (*) are all proportional to the correlations between the PC and the *X*s – with the same proportionality constant (namely, $\frac{1}{\sqrt{\lambda_i}}$) for all *X*s in the same equation.

This mathematical result tells us how to interpret the meaning of the PCs:

- If an *X* has a high correlation with a PC, then the *X* and the PC are highly related and the meaning of the *X* transfers to the meaning of the PC.
- If an *X* has a low correlation with a PC, then the *X* and the PC are not much related and the meaning of the *X* has little to do with the meaning of the PC.
- Find the *X*s that are most highly correlated with a PC by picking out the coefficients that are highest in the PC – the actual correlations are proportional to these coefficients.
- The meaning of the PC should be defined by a theoretical/practical construct that the most highly correlating *X*s jointly measure – if such a construct can be divined.