The matrices of PCA

- **CV** = **PV**
- **C**: covariance matrix
  - C_{kl} is the covariance between dimensions k and l.
  - Usually empirically estimated from data.
- **V**: matrix of column *eigenvectors*
  - V_{kl} is the
    - importance of the original coordinate k
    - in the transformed coordinate l.
    - known as the principal components
    - the columns are orthonormal vectors
- **P**: diagonal matrix of *eigenvalues*
  - $P_{kl} = \lambda_{kl}$ if k = l
  - $P_{kl} = 0$ otherwise
  - variances in transformed coordinate system
  - scaling of the eigenvectors
- All three matrices have the same size

PCA analysis of molecular simulation

- In a molecular simulation, it could be helpful to visualize
  - the principal components themselves
  - the variance and cumulated variance
  - the projections onto the principal components in one or more dimensions
    - time series
    - histograms

- See PCA.ipynb, which shows PCA analysis for a simulation of ubiquitin