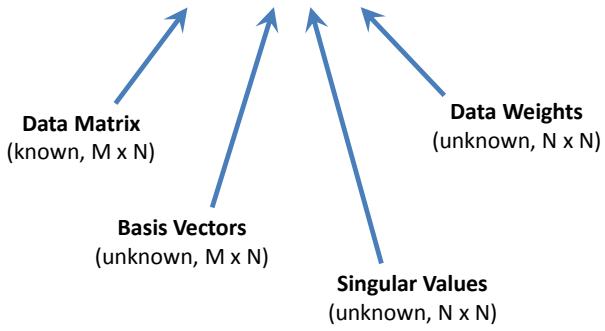
Singular Value Decomposition

- **Basis Spectrum:** A single spectrum that can be combined in linear combination with other basis spectra to create any observed spectrum
- **Singular Value Decomposition:** The mathematical process of determining the basis spectra for a given set of observed spectra

Singular Value Decomposition

$$\tilde{A} = \tilde{U}\,\tilde{S}\,\tilde{V}^T$$



- Unknown matrices are found via a computational algorithm

The Data Matrix

Each column is a separate spectrum, collected on a different protein

$$\begin{pmatrix} \theta_{1\lambda_1} & \theta_{2\lambda_1} & \cdots & \theta_{n\lambda_1} \\ \theta_{1\lambda_2} & & & \vdots \\ \vdots & & & \vdots \\ \theta_{1\lambda_m} & \cdots & \cdots & \theta_{n\lambda_m} \end{pmatrix}$$

Each row represents the CD signal at a given wavelength

The Basis Vector Matrix

$$\begin{pmatrix} \theta_{1\lambda_1} & \theta_{2\lambda_1} & \cdots & \theta_{n\lambda_1} \\ \theta_{1\lambda_2} & & & \vdots \\ \vdots & & & \vdots \\ \theta_{1\lambda_m} & \cdots & \cdots & \theta_{n\lambda_m} \end{pmatrix}$$

- Similar format to data matrix
- Each column is a vector representing a basis spectrum

The Singular Value Matrix

$$\tilde{S} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & 0 & s_n \end{pmatrix} \qquad \tilde{S}^{-1} = \begin{pmatrix} s_1^{-1} & 0 & 0 & \cdots & 0 \\ 0 & s_2^{-1} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & 0 & s_n^{-1} \end{pmatrix}$$

- Square, diagonal matrix. In general, $s_i > s_{i+1}$.
- s_i corresponds to relative weight (importance) of basis spectrum i.
- If $s_1 > s_2 > s_3 \gg s_4$, the majority of the data can be described with the first three basis vectors

The Data Weight Matrix

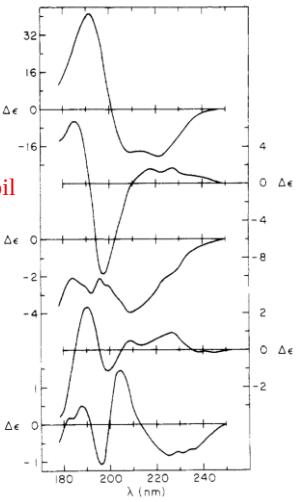
$$\tilde{V}^T = \begin{pmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1n} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2n} \\ \vdots & & & & \vdots \\ v_{n1} & \cdots & \cdots & v_{(n-1)n} & v_{nn} \end{pmatrix}$$

Each row corresponds to a basis vector

Each column corresponds to a protein

- v_{ij} = How much of basis spectrum i is needed to reconstruct protein spectrum j.

Five Most Significant Basis Spectra



Taken From:
Johnson, W. C., Jr. (1981) *Biochemistry*. **20** (5): 1085-1094

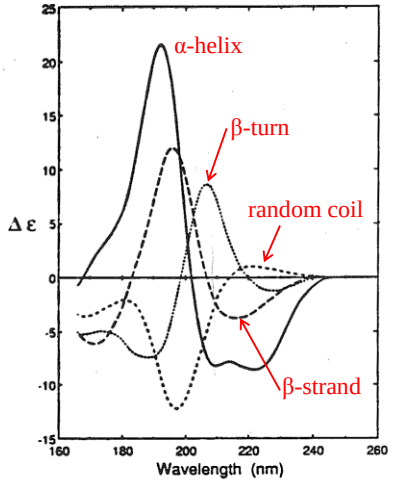


Fig. 2. The CD for various secondary structures: α-helix (——), antiparallel β-sheet (——), β-turn (····), and random coil (—), redrawn from Ref. 7.