Residuals in matrix form

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

$$\begin{bmatrix} 1 - 2\beta_1 \\ 3 - 3\beta_1 \\ 2 - 1\beta_1 \\ 7 - 4\beta_1 \\ 6 - 5\beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 2\beta_1 \\ 3\beta_1 \\ 1\beta_1 \\ 4\beta_1 \\ 5\beta_1 \end{bmatrix}$$

5×1 $\quad$ 5×1 $\quad$ 5×1

Sum of squared residuals

(Y - X β)' **(Y - X β)**

$$\begin{bmatrix} 1-2\beta_1 & 3-3\beta_1 & 2-1\beta_1 & 7-4\beta_1 & 6-5\beta_1 \end{bmatrix}$$

1 × 5

$$\begin{bmatrix} 1-2\beta_1 \\ 3-3\beta_1 \\ 2-1\beta_1 \\ 7-4\beta_1 \\ 6-5\beta_1 \end{bmatrix}$$

5 × 1

This matrix product yields a 1×1 matrix whose only element equals the sum of squared deviations. The prime symbol (which looks like an apostrophe) means that the matrix is transposed (i.e. the columns are changed to rows).

$$(\mathbf{Y}-\boldsymbol{\beta}\mathbf{X})'(\mathbf{Y}-\boldsymbol{\beta}\mathbf{X}) = \sum_{i=1}^{5}(y_i - \beta_1 x_i)^2$$

matrix version non-matrix version

Log-likelihood

$$\log L = -\frac{1}{2\sigma^2}\sum_{i=1}^{5}(y_i - \beta_1 x_i)^2$$

non-matrix

using matrices

$$\log L = -\frac{1}{2\sigma^2}\left(\mathbf{Y} - \boldsymbol{\beta}\mathbf{X}\right)'\left(\mathbf{Y} - \boldsymbol{\beta}\mathbf{X}\right)$$

Estimating the slope

$$\beta = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

non-matrix

using matrices

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{Y})$$

Adding the intercept

The model just presented was this:

$$Y = \beta_1 X$$

Ordinarily, regressions also include an intercept term, β_0 (which is the predicted value of Y when X = 0):

$$Y = \beta_0 + \beta_1 X$$

Matrix representation with intercept

$$\mathbf{Y} = \mathbf{X} \quad \boldsymbol{\beta} + \mathbf{e}$$

$$\begin{bmatrix} 1 \\ 3 \\ 2 \\ 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} = \begin{bmatrix} \beta_0 + 2\beta_1 + \epsilon_1 \\ \beta_0 + 3\beta_1 + \epsilon_2 \\ \beta_0 + 1\beta_1 + \epsilon_3 \\ \beta_0 + 4\beta_1 + \epsilon_4 \\ \beta_0 + 5\beta_1 + \epsilon_5 \end{bmatrix}$$

5 × 1   5 × 2   2 × 1   5 × 1   5 × 1

Variance-covariance matrix

$$\log L = -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

variance replaced by variance-covariance matrix

$$\log L = -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

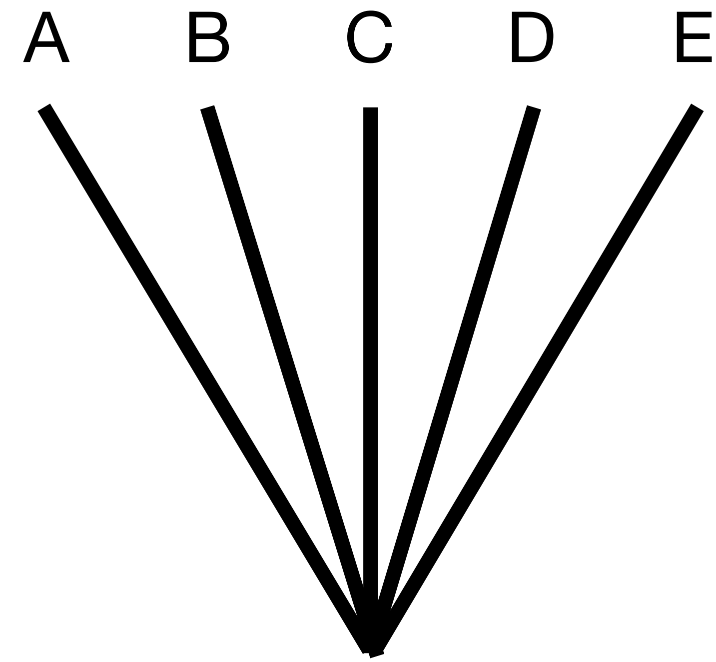
Variance matrix and its inverse

$$\mathbf{V} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

$$\mathbf{V}^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma^2} \end{bmatrix}$$

Variance-covariance matrix and phylogeny

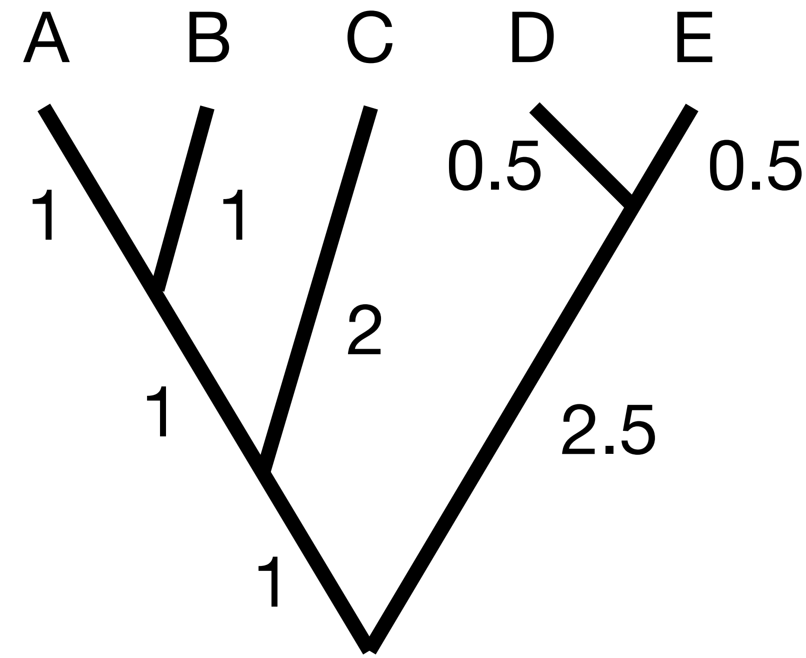
$$\mathbf{V} = \left[\begin{array}{ccccc} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{array}\right]$$



The matrix V corresponds to a star phylogeny

Variance-covariance matrix and Brownian motion on a phylogeny

$$\mathbf{V} = \sigma^2 \begin{bmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 0 \\ 1 & 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 2.5 \\ 0 & 0 & 0 & 2.5 & 3 \end{bmatrix}$$



Martins and Hansen 1997

Phylogenetic Generalized Least Squares (PGLS) Regression

$$\boldsymbol{\beta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\right)$$

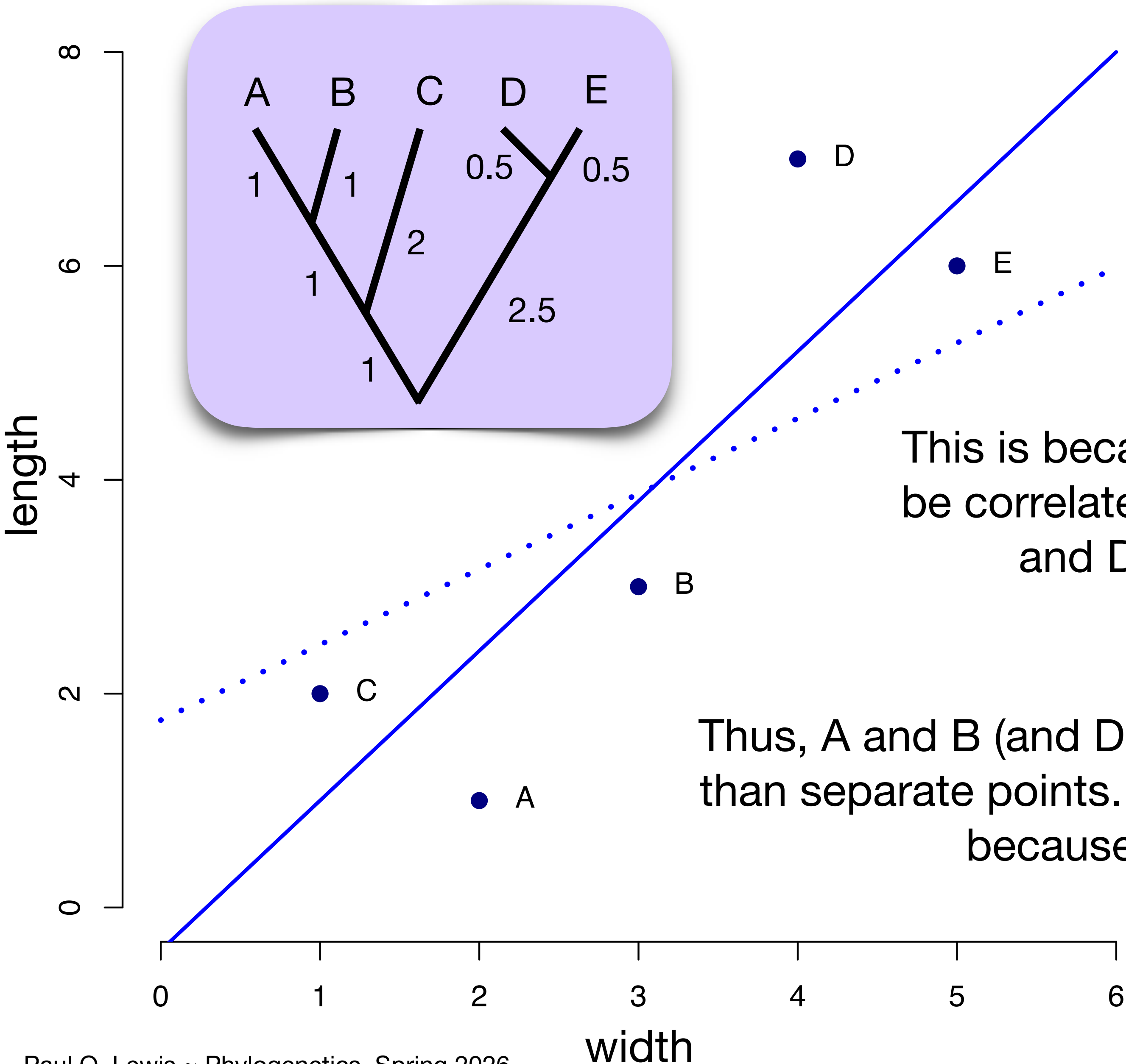
2×5 5×5 5×2 — 2×5 5×5 5×1

2×1 — 2×2 — 2×1

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \begin{matrix} \text{intercept} \\ \text{slope} \end{matrix} \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.7521 \\ 0.7055 \end{bmatrix}$$

(for this example)

PGLS Regression



The PGLS regression (dotted) is less influenced by A and D than the non-phylogenetic regression (solid).

This is because A and B are expected to be correlated due to their shared history, and D and E share an even greater fraction of their history.

Thus, A and B (and D,E) act more like single points than separate points. C now wields more influence because of its relative independence.