

Cubic spline:

$$f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \sum_{i=1}^K \beta_{i+4} (x - \xi_i)_+^3$$

We have for $x < \xi_1$:

$$f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3$$

$$f'(x) = \beta_2 + 2\beta_3 x + 3\beta_4 x^2$$

$$f''(x) = 2\beta_3 + 6\beta_4 x$$

$$f^{(3)}(x) = 6\beta_4$$

$$\rightarrow f''(\xi_1) = 2\beta_3 + 6\beta_4 \xi_1 = 0, \quad f^{(3)}(\xi_1) = 6\beta_4 = 0 \rightarrow \beta_4 = 0$$

$$\rightarrow 2\beta_3 + 6 \cdot 0 \cdot \xi_1 = 0 \rightarrow \beta_3 = 0$$

We get:

$$f(x) = \beta_1 + \beta_2 x + \sum_{i=1}^K \beta_{i+4} (x - \xi_i)_+^3$$

For $x \geq \xi_K$, we have:

$$f(x) = \beta_1 + \beta_2 x + \sum_{i=1}^K \beta_{i+4} (x - \xi_i)^3$$

$$\rightarrow f'(x) = \beta_2 + \sum_{i=1}^K 3\beta_{i+4} (x - \xi_i)^2$$

$$\rightarrow f''(x) = \sum_{i=1}^K 6\beta_{i+4} (x - \xi_i)$$

$$\rightarrow f^{(3)}(x) = \sum_{i=1}^K 6\beta_{i+4}$$

$$\rightarrow f''(\xi_K) = \sum_{i=1}^K 6\beta_{i+4} (\xi_K - \xi_i) = 0$$

$$\rightarrow \sum_{i=1}^{K-1} 6\beta_{i+4} (\xi_K - \xi_i) + 6\beta_{K+4} (\xi_K - \xi_K) = 0$$

$$\rightarrow \sum_{i=1}^{K-2} \cancel{\beta_{i+4} (\xi_K - \xi_i)} + \cancel{\beta_{K+3} (\xi_K - \xi_{K-1})} = 0$$

$$\rightarrow \beta_{K+3} = -\frac{1}{\xi_K - \xi_{K-1}} \sum_{i=1}^{K-2} \beta_{i+4} (\xi_K - \xi_i) = -\sum_{i=1}^{K-2} \beta_{i+4} \frac{\xi_K - \xi_i}{\xi_K - \xi_{K-1}}$$

$$f^{(3)}(\xi_K) = \sum_{i=1}^K 6\beta_{i+4} = \sum_{i=1}^{K-2} \cancel{\beta_{i+4}} + \cancel{\beta_{K+3}} + \cancel{\beta_{K+4}} = 0$$

$$\begin{aligned} \rightarrow \beta_{K+4} &= -\beta_{K+3} - \sum_{i=1}^{K-2} \beta_{i+4} \\ &= \sum_{i=1}^{K-2} \beta_{i+4} \frac{\xi_K - \xi_i}{\xi_K - \xi_{K-1}} - \sum_{i=1}^{K-2} \beta_{i+4} \\ &= \sum_{i=1}^{K-2} \beta_{i+4} \left(\frac{\xi_K - \xi_i}{\xi_K - \xi_{K-1}} - 1 \right) \\ &= \sum_{i=1}^{K-2} \beta_{i+4} \left(\frac{\cancel{\xi_K} - \xi_i - \cancel{\xi_K} + \xi_{K-1}}{\xi_K - \xi_{K-1}} \right) \\ &= \sum_{i=1}^{K-2} \beta_{i+4} \frac{\xi_{K-1} - \xi_i}{\xi_K - \xi_{K-1}} \end{aligned}$$

We obtain:

$$\begin{aligned} f(x) &= \beta_1 + \beta_2 x + \sum_{i=1}^{K-2} \beta_{i+4} (x - \xi_i)_+^3 + \beta_{K+4} (x - \xi_{K-1})_+^3 + \beta_{K+4} (x - \xi_K)_+^3 \\ &= \beta_1 + \beta_2 x + \sum_{i=1}^{K-2} \beta_{i+4} (x - \xi_i)_+^3 - \sum_{i=1}^{K-2} \beta_{i+4} \frac{\xi_K - \xi_i}{\xi_K - \xi_{K-1}} (x - \xi_{K-1})_+^3 + \sum_{i=1}^{K-2} \beta_{i+4} \frac{\xi_{K-1} - \xi_i}{\xi_K - \xi_{K-1}} (x - \xi_K)_+^3 \\ &= \beta_1 + \beta_2 x + \sum_{i=1}^{K-2} \beta_{i+4} \left((x - \xi_i)_+^3 - (\xi_K - \xi_i) \frac{(x - \xi_{K-1})_+^3}{\xi_K - \xi_{K-1}} + (\xi_{K-1} - \xi_i) \frac{(x - \xi_K)_+^3}{\xi_K - \xi_{K-1}} \right) \\ &= \beta_1 + \beta_2 x + \sum_{i=1}^{K-2} \beta_{i+4} \left((x - \xi_i)_+^3 - (\xi_K - \xi_i) \frac{(x - \xi_{K-1})_+^3}{\xi_K - \xi_{K-1}} + (\xi_K - \xi_i) \frac{(x - \xi_K)_+^3}{\xi_K - \xi_{K-1}} - \cancel{(\xi_{K-1} - \xi_i) \frac{(x - \xi_K)_+^3}{\xi_K - \xi_{K-1}}} \right) \\ &= \beta_1 + \beta_2 x + \sum_{i=1}^{K-2} \beta_{i+4} (\xi_K - \xi_i) \left(\frac{(x - \xi_i)_+^3}{\xi_K - \xi_i} - \frac{(x - \xi_K)_+^3}{\xi_K - \xi_i} \right) - \frac{(x - \xi_{K-1})_+^3 - (x - \xi_K)_+^3}{\xi_K - \xi_{K-1}} \\ &\quad \underbrace{\beta_{i+4}}_{\tilde{\beta}_{i+2}} \underbrace{\left(\frac{(x - \xi_i)_+^3}{\xi_K - \xi_i} - \frac{(x - \xi_K)_+^3}{\xi_K - \xi_i} \right)}_{d_i(x)} - \underbrace{\frac{(x - \xi_{K-1})_+^3 - (x - \xi_K)_+^3}{\xi_K - \xi_{K-1}}}_{d_{K-1}(x)} \\ &= \beta_1 + \beta_2 x + \sum_{i=1}^{K-2} \tilde{\beta}_{i+2} \underbrace{(d_i(x) - d_{K-1}(x))}_{N_{i+2}(x)} \\ &= \beta_1 + \beta_2 x + \sum_{i=1}^{K-2} \tilde{\beta}_{i+2} N_{i+2}(x) \end{aligned}$$

$$RSS(f, \lambda) = \sum_{i=1}^N (y_i - f(x_i))^2 + \lambda \int_{-\infty}^{\infty} (f''(t))^2 dt$$

We have:

$$f(x_i) = \sum_{j=1}^N \theta_j N_j(x_i) = \underline{N}(x_i)^T \underline{\theta} \quad , \quad \underline{N}(x_i) = \begin{pmatrix} N_1(x_i) \\ \vdots \\ N_N(x_i) \end{pmatrix}$$

$$\rightarrow \sum_{i=1}^N (y_i - f(x_i))^2 = \sum_{i=1}^N (y_i - \underline{N}(x_i)^T \underline{\theta})^2$$

$$= (\underline{y} - \underline{N} \underline{\theta})^T (\underline{y} - \underline{N} \underline{\theta})$$

$$\underline{N} = \begin{pmatrix} \underline{N}_1(x_1)^T \\ \vdots \\ \underline{N}_N(x_N)^T \end{pmatrix}$$

Further,

$$f(t) = \sum_{j=1}^N \theta_j N_j(t)$$

$$\rightarrow f'(t) = \sum_{j=1}^N \theta_j N_j'(t)$$

$$\rightarrow f''(t) = \sum_{j=1}^N \theta_j N_j''(t) = (\underline{N}''(t))^T \underline{\theta} \quad \text{with} \quad \underline{N}''(t) = \begin{pmatrix} N_1''(t) \\ \vdots \\ N_N''(t) \end{pmatrix}$$

This gives

$$\int_{-\infty}^{\infty} (f''(t))^2 dt = \int_{-\infty}^{\infty} ((\underline{N}''(t))^T \underline{\theta})^2 dt$$

$$= \int_{-\infty}^{\infty} \underline{\theta}^T \underline{N}''(t) (\underline{N}''(t))^T \underline{\theta} dt$$

$$= \underline{\theta}^T \int_{-\infty}^{\infty} \underline{N}''(t) (\underline{N}''(t))^T dt \underline{\theta}$$

We have:

$$\underline{N}''(t) (\underline{N}''(t))^T = \begin{pmatrix} N_1''(t) \\ \vdots \\ N_N''(t) \end{pmatrix} (N_1''(t) \dots N_N''(t))$$

$$= \begin{pmatrix} (N_1''(t))^2 & N_1''(t) N_2''(t) & \dots & N_1''(t) N_N''(t) \\ N_2''(t) N_1''(t) & (N_2''(t))^2 & & \\ \vdots & & \ddots & \\ N_N''(t) N_1''(t) & & & (N_N''(t))^2 \end{pmatrix}$$

$$\rightarrow \int_{-\infty}^{\infty} \underline{N}''(t) (\underline{N}''(t))^T dt = \underbrace{\begin{pmatrix} \int_{-\infty}^{\infty} (N_1''(t))^2 dt & \dots & \int_{-\infty}^{\infty} N_1''(t) N_N''(t) dt \\ \vdots & \ddots & \vdots \\ \int_{-\infty}^{\infty} N_N''(t) N_1''(t) dt & \dots & \int_{-\infty}^{\infty} (N_N''(t))^2 dt \end{pmatrix}}_{\underline{\Omega}}$$

This gives

$$RSS(f, \lambda) = (\underline{y} - \underline{N} \underline{\theta})^T (\underline{y} - \underline{N} \underline{\theta}) + \lambda \underline{\theta}^T \underline{\Omega} \underline{\theta}$$

$$= \underline{y}^T \underline{y} - 2 \underline{y}^T \underline{N} \underline{\theta} + \underline{\theta}^T \underline{N}^T \underline{N} \underline{\theta} + \lambda \underline{\theta}^T \underline{\Omega} \underline{\theta}$$

$$\text{Recall that } \frac{\partial \underline{a}^T \underline{b}}{\partial \underline{b}} = \underline{a} \quad \text{and} \quad \frac{\partial \underline{b}^T \underline{A} \underline{b}}{\partial \underline{b}} = (\underline{A} + \underline{A}^T) \underline{b}$$

$$\rightarrow \frac{\partial \underline{y}^T \underline{N} \underline{\theta}}{\partial \underline{\theta}} = \underline{N}^T \underline{y}$$

$$\frac{\partial \underline{\theta}^T \underline{N}^T \underline{N} \underline{\theta}}{\partial \underline{\theta}} = (\underline{N}^T \underline{N} + \underbrace{(\underline{N}^T \underline{N})^T}_{= \underline{N}^T \underline{N}}) \underline{\theta} = 2 \underline{N}^T \underline{N} \underline{\theta}$$

$$\frac{\partial \underline{\theta}^T \underline{\Omega} \underline{\theta}}{\partial \underline{\theta}} = (\underline{\Omega} + \underbrace{\underline{\Omega}^T}_{= \underline{\Omega}}) \underline{\theta} = 2 \underline{\Omega} \underline{\theta}$$

$$\rightarrow \frac{\partial RSS(f, \lambda)}{\partial \underline{\theta}} = -2 \underline{N}^T \underline{y} + 2 \underline{N}^T \underline{N} \underline{\theta} + 2 \underline{\Omega} \underline{\theta}$$

$$= \cancel{-2} (-\underline{N}^T \underline{y} + (\underline{N}^T \underline{N} + \lambda \underline{\Omega}) \underline{\theta}) = 0$$

$$\rightarrow (\underline{N}^T \underline{N} + \lambda \underline{\Omega}) \underline{\theta} = \underline{N}^T \underline{y}$$

$$\rightarrow \underline{\theta} = (\underline{N}^T \underline{N} + \lambda \underline{\Omega})^{-1} \underline{N}^T \underline{y}$$