# Notes on Smoothing Spline

Max Leung

January 25, 2022

## Definition 1

Cubic Spline is $f(x) = \sum_{j=0}^{3} \beta_j x^j + \sum_{k=1}^{K} \theta_k (x - \kappa_k)_+^3$

## Definition 2

Natural Cubic Spline is $f(x) = \sum_{j=0}^{3} \beta_j x^j + \sum_{k=1}^{K} \theta_k (x - \kappa_k)_+^3$ with linear constraint $\beta_2 = \beta_3 = 0$ and $\sum_{k=1}^{K} \theta_k \kappa_k = \sum_{k=1}^{K} \theta_k = 0$

The linear constraint can make $f(x)$ to be linear when $x < \kappa_1$ and $x > \kappa_K$. It can be seen

For $x < \kappa_1 < \cdots < \kappa_K$,

$$\begin{aligned}
f(x) &= \sum_{j=0}^{3} \beta_j x^j + \sum_{k=1}^{K} \theta_k (x - \kappa_k)_+^3 \\
&= \sum_{j=0}^{3} \beta_j x^j \\
&= \beta_0 + \beta_1 x && \beta_2 = \beta_3 = 0 \text{ by constraint}
\end{aligned}$$

which is linear in $x$; For $\kappa_1 < \cdots < \kappa_K < x$,

$$\begin{aligned}
f(x) &= \sum_{j=0}^{3} \beta_j x^j + \sum_{k=1}^{K} \theta_k (x - \kappa_k)_+^3 \\
&= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^{K} \theta_k (x^3 - 3\kappa_k x^2 + 3\kappa_k^2 x - \kappa_k^3) \\
&= (\beta_0 - \sum_{k=1}^{K} \theta_k \kappa_k^3) + (\beta_1 + 3\sum_{k=1}^{K} \theta_k \kappa_k^2) x + (\beta_2 - 3\sum_{k=1}^{K} \theta_k \kappa_k) x^2 + (\beta_3 + \sum_{k=1}^{K} \theta_k) x^3 \\
&= (\beta_0 - \sum_{k=1}^{K} \theta_k \kappa_k^3) + (\beta_1 + 3\sum_{k=1}^{K} \theta_k \kappa_k^2) x && \text{due to constraint}
\end{aligned}$$

which is also linear in $x$

## Theorem 1

Natural Cubic Spline with K number of knots can be represented by K number of basis functions.
Proof:

$$
\begin{aligned}
\sum_{k=1}^{K}\theta_k(x-\kappa_k)_+^3 &= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3+\theta_{K-1}(x-\kappa_{K-1})_+^3+\theta_K(x-\kappa_K)_+^3\\
&= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3-\theta_{K-1}(x-\kappa_K)_+^3+\theta_{K-1}(x-\kappa_{K-1})_+^3+\theta_K(x-\kappa_K)_+^3+\theta_{K-1}(x-\kappa_K)_+^3\\
&= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3+\theta_{K-1}[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]+(\theta_K+\theta_{K-1})(x-\kappa_K)_+^3\\
&= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3-(-\theta_{K-1}\frac{\kappa_K-\kappa_{K-1}}{\kappa_K-\kappa_{K-1}})[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]+(\theta_K+\theta_{K-1})(x-\kappa_K)_+^3\\
&= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3-(\frac{-\theta_{K-1}\kappa_K+\theta_{K-1}\kappa_{K-1}}{\kappa_K-\kappa_{K-1}})[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]+(\theta_K+\theta_{K-1})(x-\kappa_K)_+^3\\
&= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3-(\frac{-\theta_{K-1}\kappa_K-\theta_K\kappa_K+\theta_K\kappa_K+\theta_{K-1}\kappa_{K-1}}{\kappa_K-\kappa_{K-1}})[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]\\
&\quad+(\theta_K+\theta_{K-1})(x-\kappa_K)_+^3\\
&= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3-(\frac{(-\theta_{K-1}-\theta_K)\kappa_K-(-\theta_K\kappa_K-\theta_{K-1}\kappa_{K-1})}{\kappa_K-\kappa_{K-1}})[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]\\
&\quad-(-\theta_K-\theta_{K-1})(x-\kappa_K)_+^3
\end{aligned}
$$

As $\sum_{k=1}^{K}\theta_k=0 \implies \sum_{k=1}^{K-2}\theta_k=-\theta_{K-1}-\theta_K$ and $\sum_{k=1}^{K}\theta_k\kappa_k=0 \implies \sum_{k=1}^{K-2}\theta_k\kappa_k=-\theta_{K-1}\kappa_{K-1}-\theta_K\kappa_K$

$$
\begin{aligned}
&= \sum_{k=1}^{K-2}\theta_k(x-\kappa_k)_+^3-(\frac{\sum_{k=1}^{K-2}\theta_k\kappa_K-\sum_{k=1}^{K-2}\theta_k\kappa_k}{\kappa_K-\kappa_{K-1}})[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]-\sum_{k=1}^{K-2}\theta_k(x-\kappa_K)_+^3\\
&= \sum_{k=1}^{K-2}\theta_k[(x-\kappa_k)_+^3-(x-\kappa_K)_+^3]-\frac{\sum_{k=1}^{K-2}\theta_k(\kappa_K-\kappa_k)}{\kappa_K-\kappa_{K-1}}[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]\\
&= \sum_{k=1}^{K-2}\frac{\theta_k(\kappa_K-\kappa_k)}{\kappa_K-\kappa_k}[(x-\kappa_k)_+^3-(x-\kappa_K)_+^3]-\frac{\sum_{k=1}^{K-2}\theta_k(\kappa_K-\kappa_k)}{\kappa_K-\kappa_{K-1}}[(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3]\\
&= \sum_{k=1}^{K-2}\theta_k(\kappa_K-\kappa_k)\{\frac{(x-\kappa_k)_+^3-(x-\kappa_K)_+^3}{\kappa_K-\kappa_k}-\frac{(x-\kappa_{K-1})_+^3-(x-\kappa_K)_+^3}{\kappa_K-\kappa_{K-1}}\}\\
&= \sum_{k=1}^{K-2}\theta_k(\kappa_K-\kappa_k)\{d_k(x)-d_{K-1}(x)\}
\end{aligned}
$$

$$
\begin{aligned}
f(x) &= \sum_{j=0}^{3}\beta_j x^j+\sum_{k=1}^{K}\theta_k(x-\kappa_k)_+^3\\
&= \beta_0+\beta_1 x+\sum_{k=1}^{K}\theta_k(x-\kappa_k)_+^3 && \text{due to constraint}\\
&= \beta_0\cdot 1+\beta_1 x+\sum_{k=1}^{K-2}\theta_k(\kappa_K-\kappa_k)\{d_k(x)-d_{K-1}(x)\}\\
&= \theta_1' N_1(x)+\theta_2' N_2(x)+\sum_{k=1}^{K-2}\theta_{2+k}'\{N_{2+k}(x)\}\\
&= \sum_{k=1}^{K}\theta_k' N_k(x) && \text{Q.E.D.}
\end{aligned}
$$

## Definition 3

Smoothing Spline is the minimizer function of the problem $min_f\{\sum_{i=1}^N(y_i - f(x_i))^2 + \lambda\int_a^b[f''(t)]^2dt\}$

## Theorem 2

Smoothing Spline is a Natural Cubic Spline with knots at $x_i$ for $i = 1, \cdots, N$.

Proof:
Let $\tilde{g}$ be minimizer function of the problem $min_f\{\sum_{i=1}^N(y_i - f(x_i))^2 + \lambda\int_a^b[f''(t)]^2dt\}$ i.e., $\tilde{g}$ is the Smoothing Spline. Thus, $RSS_{min} = \sum_{i=1}^N(y_i - \tilde{g}(x_i))^2 + \lambda\int_a^b[\tilde{g}''(t)]^2dt$. Let $g$ be a Natural Cubic Spline with knots at $x_i$ for $i = 1, \cdots, N$ such that $g(x_i) = \tilde{g}(x_i)$ for $\forall i$. Define $RSS_g := \sum_{i=1}^N(y_i - g(x_i))^2 + \lambda\int_a^b[g''(t)]^2dt$. The difference between $RSS_{min}$ and $RSS_g$ is only due to the difference between $\int_a^b[\tilde{g}''(t)]^2dt$ and $\int_a^b[g''(t)]^2dt$.

$$\begin{aligned} RSS_{min} &\leq RSS_g && \text{As } \tilde{g} \text{ is the minimizer} \\ \sum_{i=1}^N(y_i - \tilde{g}(x_i))^2 + \lambda\int_a^b[\tilde{g}''(t)]^2dt &\leq \sum_{i=1}^N(y_i - g(x_i))^2 + \lambda\int_a^b[g''(t)]^2dt \\ \int_a^b[\tilde{g}''(t)]^2dt &\leq \int_a^b[g''(t)]^2dt \end{aligned}$$

Define $h(x) := \tilde{g}(x) - g(x)$,

$$\begin{aligned} \int_a^b[\tilde{g}''(t)]^2dt &= \int_a^b[(h+g)''(t)]^2dt \\ &= \int_a^b[h''(t) + g''(t)]^2dt \\ &= \int_a^b[h''(t)]^2 + 2h''(t)g''(t) + [g''(t)]^2dt \\ &= \int_a^b[h''(t)]^2dt + 2\int_a^b h''(t)g''(t)dt + \int_a^b[g''(t)]^2dt \\ &= \int_a^b[h''(t)]^2dt + \int_a^b[g''(t)]^2dt && \text{As } \int_a^b h''(t)g''(t)dt = 0\text{, which will be shown below} \\ &\geq \int_a^b[g''(t)]^2dt && \text{As } \int_a^b[h''(t)]^2dt \geq 0 \end{aligned}$$

Combine the result, we have $\int_a^b[\tilde{g}''(t)]^2dt = \int_a^b[g''(t)]^2dt$. Thus $\tilde{g} = g$.

$$\begin{aligned} \int_a^b g''(t)h''(t)dt &= g''(t)h'(t)|_a^b - \int_a^b g'''(t)h'(t)dt && \text{Integration by part} \\ &= \underbrace{g''(b)}_{0}h'(b) - \underbrace{g''(a)}_{0}h'(a) - \int_a^b g'''(t)h'(t)dt && g \text{ is linear at a and b} \\ &= -\int_a^b g'''(t)h'(t)dt \\ &= -\sum_{j=1}^{N-1}\int_j^{j+1} g'''(x_i)h'(x_i)di \\ &= -\sum_{j=1}^{N-1}[g'''(x_i)h(x_i)|_j^{j+1} - \int_j^{j+1} g''''(x_i)h(x_i)di] && \text{Integration by part} \\ &= 0 && \text{as } h(x_i) = 0 \text{ by definition. } g'''' = 0 \text{ as } g \text{ is cubic. Q.E.D.} \end{aligned}$$

## Theorem 3

Smoothing Spline $f(x)$ can be represented by N number of basis functions i.e., $f(x) = \sum_{j=1}^N N_j(x)\theta_j$

Proof:
By Theorem 2, Smoothing Spline is a Natural Cubic Spline with knots at $x_i$ for $i = 1, \cdots, N$ i.e., with N number of knots. By Theorem 1, Smoothing Spline can be represented by N number of basis functions i.e.,

$$f(x) = \sum_{j=1}^{N} N_j(x)\theta_j \qquad \text{Q.E.D.}$$

# Theorem 4

Smoothing Spline $f(x)$ is a Generalized Ridge Estimator times $\boldsymbol{N}(\boldsymbol{x})$.

Proof:

$$\begin{aligned}
RSS_{min} &= min_a\{\sum_{i=1}^{N}(y_i - a(x_i))^2 + \lambda \int_a^b [a''(t)]^2 dt\} \\
&= \sum_{i=1}^{N}(y_i - f(x_i))^2 + \lambda \int_a^b [f''(t)]^2 dt && \text{by Definition 3} \\
&= \sum_{i=1}^{N}(y_i - \sum_{j=1}^{N} N_j(x_i)\theta_j)^2 + \lambda \int_a^b [\sum_{j=1}^{N} N_j''(t)\theta_j]^2 dt && \text{by Theorem 3} \\
&= (\boldsymbol{y} - \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta}) + \lambda\boldsymbol{\theta}'\boldsymbol{\Omega}_N\boldsymbol{\theta}
\end{aligned}$$

Since

$$\begin{aligned}
\sum_{j=1}^{N} N_j''(t)\theta_j &= \boldsymbol{n}''(t)'\boldsymbol{\theta} \\
[\sum_{j=1}^{N} N_j''(t)\theta_j]^2 &= [\boldsymbol{n}''(t)'\boldsymbol{\theta}]^2 \\
&= [\boldsymbol{n}''(t)'\boldsymbol{\theta}][\boldsymbol{n}''(t)'\boldsymbol{\theta}] \\
&= [\boldsymbol{\theta}'\boldsymbol{n}''(t)][\boldsymbol{n}''(t)'\boldsymbol{\theta}] \\
&= \boldsymbol{\theta}'\boldsymbol{n}''(t)\boldsymbol{n}''(t)'\boldsymbol{\theta} \\
&= \boldsymbol{\theta}' \begin{pmatrix} N_1''(t) \\ \vdots \\ N_N''(t) \end{pmatrix} \begin{pmatrix} N_1''(t) \cdots N_N''(t) \end{pmatrix} \boldsymbol{\theta} \\
&= \boldsymbol{\theta}' \begin{pmatrix} N_1''(t)^2 & \cdots & N_1''(t)N_N''(t) \\ \vdots & \vdots & \vdots \\ N_N''(t)N_1''(t) & \cdots & N_N''(t)^2 \end{pmatrix} \boldsymbol{\theta} \\
\int_a^b [\sum_{j=1}^{N} N_j''(t)\theta_j]^2 dt &= \int_a^b \boldsymbol{\theta}' \begin{pmatrix} N_1''(t)^2 & \cdots & N_1''(t)N_N''(t) \\ \vdots & \vdots & \vdots \\ N_N''(t)N_1''(t) & \cdots & N_N''(t)^2 \end{pmatrix} \boldsymbol{\theta} dt \\
&= \boldsymbol{\theta}' \underbrace{\begin{pmatrix} \int_a^b N_1''(t)^2 dt & \cdots & \int_a^b N_1''(t)N_N''(t) dt \\ \vdots & \vdots & \vdots \\ \int_a^b N_N''(t)N_1''(t) dt & \cdots & \int_a^b N_N''(t)^2 dt \end{pmatrix}}_{\boldsymbol{\Omega}_N} \boldsymbol{\theta}
\end{aligned}$$

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{N} N_j(x_1)\theta_j \\ \vdots \\ \sum_{j=1}^{N} N_j(x_N)\theta_j \end{pmatrix} = \begin{pmatrix} \boldsymbol{n}(x_1)'\boldsymbol{\theta} \\ \vdots \\ \boldsymbol{n}(x_N)'\boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{n}(x_1)' \\ \vdots \\ \boldsymbol{n}(x_N)' \end{pmatrix} \boldsymbol{\theta} = \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta}$$

Gradient Vector

$$\begin{aligned}
\frac{\partial RSS_{min}}{\partial \boldsymbol{\theta}} &= \frac{\partial (\boldsymbol{y} - \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta}) + \lambda \boldsymbol{\theta}' \boldsymbol{\Omega}_N \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} \\
&= \frac{\partial (\boldsymbol{y} - \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} (\boldsymbol{I} + \boldsymbol{I}')(\boldsymbol{y} - \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta}) + \lambda \frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{\theta}} (\boldsymbol{\Omega}_N + \boldsymbol{\Omega}_N')\boldsymbol{\theta} \\
&= -2\boldsymbol{N}(\boldsymbol{x})'(\boldsymbol{y} - \boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta}) + 2\lambda \boldsymbol{\Omega}_N \boldsymbol{\theta} \\
&= -2\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{y} + 2\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{N}(\boldsymbol{x})\boldsymbol{\theta} + 2\lambda \boldsymbol{\Omega}_N \boldsymbol{\theta} \\
&= -2\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{y} + 2(\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{N}(\boldsymbol{x}) + \lambda \boldsymbol{\Omega}_N)\boldsymbol{\theta}
\end{aligned}$$

FOC

$$\begin{aligned}
-2\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{y} + 2(\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{N}(\boldsymbol{x}) + \lambda \boldsymbol{\Omega}_N)\widehat{\boldsymbol{\theta}} &= \boldsymbol{0} \\
\widehat{\boldsymbol{\theta}} &= [\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{N}(\boldsymbol{x}) + \lambda \boldsymbol{\Omega}_N]^{-1}\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{y}
\end{aligned}$$

$$\begin{aligned}
\widehat{\boldsymbol{f}}(\boldsymbol{x}) &= \boldsymbol{N}(\boldsymbol{x})\widehat{\boldsymbol{\theta}} \\
&= \underbrace{\boldsymbol{N}(\boldsymbol{x})[\boldsymbol{N}(\boldsymbol{x})'\boldsymbol{N}(\boldsymbol{x}) + \lambda \boldsymbol{\Omega}_N]^{-1}\boldsymbol{N}(\boldsymbol{x})'}_{\boldsymbol{S}(\boldsymbol{x},\lambda)}\boldsymbol{y} \qquad \text{Q.E.D.}
\end{aligned}$$

$\boldsymbol{S}(\boldsymbol{x}, \lambda)$ is called smoother matrix