Inference and Characterization of Planar Trajectories

Jerome

a thesis submitted for the degree of

Doctor of Philosophy

at the University of Otago, Dunedin,

New Zealand.

December 13, 2015

Contents

List of Tables

List of Figures

1.1 Introduction

Denote by $h_m(t) : \mathbb{R}^p \mapsto \mathbb{R}$ the mth transformation of $\mathbf{T}$, $m = 1, \cdots, M$. **[what is T?]** We then model

$$f(t) = \sum_{m=1}^{M} \beta_m h_m(t).$$

[best to number all your equations; someone will want to reference them] a linear basis expansion in $\mathbf{T}$, where $h_m(t)$ are called basis functions and β_m are coefficients.

Among all functions $f(t)$ with two continuous derivatives fitting these observed data, there is only one unique $f(t)$, **[proof???]** which can be found by minimizing the following penalized residual sum of squares, returning the smallest errors,

$$\mathrm{MSE}(f, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (f(t_i) - y_i)^2 + \lambda \int_0^1 (f''(t))^2 dt \tag{1.1}$$

where λ is a fixed smoothing parameter, (t_i, y_i), $i = 1, \cdots, n$ are observed data and $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$. In equation (1.1), the smoothing parameter λ controls the smoothness of function $f(t)$,

$$\begin{cases} \lambda = 0: & f \text{ can be any function that interpolates the data,} \\ \lambda = \infty: & \text{the simple least squares line fit since no second derivative can be tolerated.} \end{cases}$$

In our case, the velocity data set v_i with some independent Gaussian distributed errors $\varepsilon_i \sim N(0, \frac{\sigma_n^2}{\gamma})$ are used to estimate $f(t)$ as well. The velocity information is incorporated into MSE equation (1.1) by the addition of velocity term $(f'(t_i) - v_i)^2$. Then it becomes

$$\mathrm{MSE}(f, \lambda, \gamma) = \frac{1}{n} \sum_{i=1}^{n} (f(t_i) - y_i)^2 + \frac{\gamma}{n} \sum_{i=1}^{n} (f'(t_i) - v_i)^2 + \lambda \int_0^1 (f''(t))^2 dt. \tag{1.2}$$

In the model $y = f(t) + \varepsilon$, it is reasonable to assume that the observed data y_i is Gaussian distribution with mean $f(t_i)$ and variance σ_n^2. In a similar way, the velocity is estimated as $v = f'(t) + \frac{\varepsilon}{\gamma}$, where v_i is Gaussian distribution with mean $f'(t_i)$ and variance $\frac{\sigma_n^2}{\gamma}$. Then the joint distribution of $\mathbf{Y}, \mathbf{V}, f(t)$ and $f'(t)$ is normal with zero mean and a covariance matrix, which can be estimated through Gaussian Process Regression.

1.2 Gaussian Process Regression

A Gaussian Process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution, Rasmussen and Williams (2006).

A GP is fully defined by its mean $m(t)$ and covariance $K(s,t)$ functions as

$$\begin{aligned} m(t) &= \mathbb{E}[f(t)] \\ K(s,t) &= \mathbb{E}[(f(s)-m(s))(f(t)-m(t))], \end{aligned}$$

where s and t are two variables, and a function f distributed as such is denoted in form of

$$f \sim GP(m(t), K(s,t)). \tag{1.3}$$

Usually the mean function is assumed to be zero everywhere.

Given a set of input variables $\mathbf{T}$ for function $f(t)$ and the output $\mathbf{Y} = f(\mathbf{T}) + \varepsilon$ with independent identically distributed Gaussian noise ε with variance σ_n^2, we can use the above definition to predict the value of the function $f_* = f(t_*)$ at a particular input t_*. As the noisy observations becoming

$$\operatorname{cov}(y_p, y_q) = K(t_p, t_q) + \sigma_n^2 \delta_{pq}$$

where δ_{pq} is a Kronecker delta which is one iff $p = q$ and zero otherwise, the joint distribution of the observed outputs $\mathbf{Y}$ and the estimated output f_* according to prior is

$$\begin{bmatrix} \mathbf{Y} \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\mathbf{T},\mathbf{T}) + \sigma_n^2 I & K(\mathbf{T}, t_*) \\ K(t_*, \mathbf{T}) & K(t_*, t_*) \end{bmatrix}\right). \tag{1.4}$$

The posterior distribution over the predicted value is obtained by conditioning on the observed data **[cov goes to var here???]**

$$f_* | \mathbf{Y}, \mathbf{T}, t_* \sim N(\bar{f}_*, \operatorname{cov}(f_*)) \tag{1.5}$$

where

$$\begin{aligned} \bar{f}_* &= \mathbb{E}[f_* | \mathbf{Y}, \mathbf{T}, t_*] = K(t_*, \mathbf{T})[K(\mathbf{T},\mathbf{T}) + \sigma_n^2 I]^{-1} \mathbf{Y}, \\ \operatorname{cov}(f_*) &= K(t_*, t_*) - K(t_*, \mathbf{T})[K(\mathbf{T},\mathbf{T}) + \sigma_n^2 I]^{-1} K(\mathbf{T}, t_*). \end{aligned}$$

We now add velocity information $\mathbf{V}$. It is assumed that the velocity $\mathbf{V} = f'(\mathbf{T}) + \frac{\varepsilon}{\sqrt{\gamma}}$, where ε is as the same **[you mean same in distribution. explain more clearly. also no correlation]** as that in $\mathbf{Y}$.

It is expected that a position point y_i and velocity point v_i are all effected by other points $\mathbf{Y}$ and $\mathbf{V}$. So the covariance matrix for $\mathbf{Y}$ and $\mathbf{V}$ is

$$\Sigma(\mathbf{Y},\mathbf{V}) = \begin{bmatrix} \mathrm{cov}(\mathbf{Y},\mathbf{Y}) & \mathrm{cov}(\mathbf{Y},\mathbf{V}) \\ \mathrm{cov}(\mathbf{V},\mathbf{Y}) & \mathrm{cov}(\mathbf{V},\mathbf{V}) \end{bmatrix}, \tag{1.6}$$

where obviously $\mathrm{cov}(\mathbf{Y},\mathbf{V}) = \mathrm{cov}(\mathbf{V},\mathbf{Y})$. Then the joint distribution is

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{V} \end{bmatrix} \sim N(\mu_{y,v}, \Sigma_{y,v}).$$

Define f_* and f'_* the estimated position and velocity values at point t_*. From equation (1.6) and using similar idea, it is easily to get the covariance matrices

$$\Sigma(f_*,\mathbf{V}) = \begin{bmatrix} \mathrm{cov}(f_*,f_*) & \mathrm{cov}(f_*,\mathbf{V}) \\ \mathrm{cov}(\mathbf{V},f_*) & \mathrm{cov}(\mathbf{V},\mathbf{V}) \end{bmatrix}, \Sigma(\mathbf{Y},f'_*) = \begin{bmatrix} \mathrm{cov}(\mathbf{Y},\mathbf{Y}) & \mathrm{cov}(\mathbf{Y},f'_*) \\ \mathrm{cov}(f'_*,\mathbf{Y}) & \mathrm{cov}(f'_*,f'_*) \end{bmatrix},$$

$$\Sigma(f_*,f'_*) = \begin{bmatrix} \mathrm{cov}(f_*,f_*) & \mathrm{cov}(f_*,f'_*) \\ \mathrm{cov}(f'_*,f_*) & \mathrm{cov}(f'_*,f'_*) \end{bmatrix},$$

[will need to give the form of these covariances at some point. in an appendix? i think you need discussion of how f' is related to f for a GP]

1.3 A Reproducing Kernel in Space $\mathbb{H}$

$N_1(t), \cdots, N_n(t)$ denote n basis functions in space $\mathbb{H}$. For any continuous function $f \in \mathbb{H}$, it is a combination of these basis functions

$$f(t) = \sum_{i=1}^{n} \alpha_i N_i(t),$$

where $\alpha_i (i = 1, \cdots, n)$ are coefficients. With an inner product **[use \langle for $<$, \rangle for $>$]**

$$< f, g >=< \sum_{i=1}^{n} \alpha_i N_i(t), \sum_{i=1}^{n} \beta_i N_i(t) >= \sum_{i=1}^{n} \alpha_i \beta_i, \tag{1.7}$$

it can be shown that the representer of evaluation $[s](\cdot)$ is

$$R_s(t) = \sum_{i=1}^{n} N_i(s) N_i(t), \tag{1.8}$$

Then we can prove that the space $\mathbb{H}$ is a Reproducing Kernel Hilbert Space. In fact,

$$< f(t), R(s,t) >=< \sum_{i=1}^{n} \alpha_i N_i(t), \sum_{i=1}^{n} N_i(s)N_i(t) >= \sum_{i=1}^{n} \alpha_i N_i(s) = f(t).$$

The term $R(s,t) = R_s(t)$ is called the reproducing kernel function.

We now introduce a new notation $\dot{R}(s,t)$ in the following and use it to find the covariance matrix Σ of the joint distribution of $\mathbf{Y}, \mathbf{V}, f$ and f'.

Define $\dot{R}(s,t)$ and $R'(s,t)$ are the first partial derivative of $R(s,t)$ with respect to the first and second argument respectively

$$\dot{R}(s,t) = \frac{\partial R(s,t)}{\partial s} = \sum_{i=1}^{n} \frac{dN_i(s)}{ds} N_i(t) = \sum_{i=1}^{n} N_i'(s)N_i(t), \tag{1.9}$$

$$R'(s,t) = \frac{\partial R(s,t)}{\partial t} = \sum_{i=1}^{n} N_i(s) \frac{dN_i(t)}{dt} = \sum_{i=1}^{n} N_i(s)N_i'(t), \tag{1.10}$$

Then $\dot{R}'(s,t)$ is the second partial derivative of $R(s,t)$ with respect to both arguments

$$\dot{R}'(s,t) = \frac{\partial^2 R(s,t)}{\partial s \partial t} = \sum_{i=1}^{n} \frac{dN_i(s)}{ds} \frac{dN_i(t)}{dt} = \sum_{i=1}^{n} N_i'(s)N_i'(t).$$

It is easily to prove that $\dot{R}(s,t) = R'(t,s)$ and

$$\begin{aligned}
< f(t), R'(s,t) > &=< \sum \alpha_i N_i(t), \sum N_i(s)N_i'(t) >= \sum \alpha_i N_i(s) = f(t), \\
< f'(t), \dot{R}(s,t) > &=< \sum \alpha_i N_i(t), \sum N_i'(s)N_i(t) >= \sum \alpha_i N_i'(s) = f'(t), \\
< f'(t), \dot{R}'(s,t) > &=< \sum \alpha_i N_i(t), \sum N_i'(s)N_i'(t) >= \sum \alpha_i N_i'(s) = f'(t), \\
< \dot{R}(s,t), f(t) > &=< \sum N_i'(s)N_i(t), \sum \alpha_i N_i(t), >= \sum \alpha_i N_i'(s) = f'(t), \\
< R'(s,t), f'(t) > &=< \sum N_i(s)N_i'(t), \sum \alpha_i N_i'(t) >= \sum \alpha_i N_i(s) = f(t).
\end{aligned}$$

Given the sample points $t_i, i = 1, \cdots, n$ and noting that the space

$$\mathbb{A} = \{f : f = \sum_{i=1}^{n} \alpha_i R(t_i, \cdot)\}$$

is one linear subspace of $\mathbb{H}$. Then $f \in \mathbb{H}$ can be written as

$$f(t) = \sum_{i=1}^{n} c_i R(t_i, t) + \rho(t) \tag{1.11}$$

where c_i are coefficients, $\rho(t) \in \mathbb{H} \ominus \mathbb{A}$, and

$$f'(t) = \sum_{i=1}^{n} c_i R'(t_i, t) + \rho'(t). \tag{1.12}$$

The equation (1.2) can be written as

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{n} c_j R(t_j, t_i) - \rho(t_i))^2 + \frac{\gamma}{n}\sum_{i=1}^{n}(v_i - \sum_{j=1}^{n} c_j R'(t_j, t_i) - \rho'(t_i))^2$$
$$+\lambda \int_{t_1}^{t_n} (\sum_{j=1}^{n} c_j R''(t_j, t) + \rho''(t))^2 dt$$

As $\dot{R}(t_i, \cdot) = \sum_{j=1}^{n} N_j'(t_i) N_j(t) \in \mathbb{A}$, then by orthogonality and property of reproducing kernel functions, $\rho(t_i) =< R(t_i, \cdot), \rho >= 0$, and $\rho'(t_i) =< \rho, R'(t_i, \cdot) >=< \dot{R}(t_i, \cdot), \rho >= 0$, where $i = 1, \cdots, n$.

Denoting by Q the $n \times n$ matrix with the (i, j)th entry $R(t_i, t_j)$, by P the $n \times n$ matrix with the (i, j)th entry $\dot{R}'(t_i, t_j)$ the equation (1.2) can be written as

$$(\mathbf{Y} - Q\mathbf{c})^T(\mathbf{Y} - Q\mathbf{c}) + \gamma(\mathbf{V} - P\mathbf{c})^T(\mathbf{V} - P\mathbf{c}) + n\lambda\Omega + \lambda(\rho, \rho). \tag{1.13}$$

Note that ρ only appears in the third term in (1.22), which is minimized at $\rho = 0$. Hence, a polynomial smoothing spline resides in the space $\mathbb{A}$ of finite dimension. Then the solution could be computed via minimization of term in (1.22) with respect to $\mathbf{c}$, Gu (2013).

1.4 Covariance Matrix and Posterior Mean

Consider f and f' in $\mathbb{H}$, having Gaussian priors with zero mean and covariance functions

$$\begin{aligned}
\text{cov}(f(s), f(t)) &= \tau^2 R(s, t) + \sigma_n^2 \Lambda \\
\text{cov}(f(s), f'(t)) &= \tau^2 R'(s, t) + \frac{\sigma_n^2}{\sqrt{\gamma}}\Lambda \\
\text{cov}(f'(s), f(t)) &= \tau^2 \dot{R}(s, t) + \frac{\sigma_n^2}{\sqrt{\gamma}}\Lambda \\
\text{cov}(f'(s), f'(t)) &= \tau^2 \dot{R}'(s, t) + \frac{\sigma_n^2}{\gamma}\Lambda
\end{aligned}$$

where Λ is the correlation matrix **[where does Λ come from?]**. Observing $y_i \sim N(f(t_i), \sigma_n^2)$ and $v_i \sim N(f'(t_i), \frac{\sigma_n^2}{\gamma})$, the joint distribution of $\mathbf{Y}, \mathbf{V}, f(t)$ and $f'(t)$ is normal with zero mean and covariance matrix

$$
\begin{aligned}
\operatorname{cov}(\mathbf{Y}, \mathbf{V}, f, f') &= \begin{bmatrix} \tau^2 R(t_i,t_j) + \sigma_n^2 \Lambda & \tau^2 R'(t_i,t_j) + \frac{\sigma_n^2}{\sqrt{\gamma}}\Lambda & \tau^2 R(t_i,t) & \tau^2 R'(t_i,t) \\ \tau^2 \dot{R}(t_i,t_j) + \frac{\sigma_n^2}{\sqrt{\gamma}}\Lambda & \tau^2 \dot{R}'(t_i,t_j) + \frac{\sigma_n^2}{\gamma}\Lambda & \tau^2 \dot{R}(t_i,t) & \tau^2 \dot{R}'(t_i,t) \\ \tau^2 R^T(t_i,t) & \tau^2 \dot{R}^T(t_i,t) & \tau^2 R(t,t) & \tau^2 \dot{R}(t,t) \\ \tau^2 R'^T(t_i,t) & \tau^2 \dot{R}'^T(t_i,t) & \tau^2 R'(t,t) & \tau^2 \dot{R}'(t,t) \end{bmatrix} \\
&= \begin{bmatrix} \tau^2 Q + \sigma_n^2 \Lambda & \tau^2 O + \frac{\sigma_n^2}{\sqrt{\gamma}}\Lambda & \tau^2 \xi & \tau^2 \xi' \\ \tau^2 O + \frac{\sigma_n^2}{\sqrt{\gamma}}\Lambda & \tau^2 P + \frac{\sigma_n^2}{\gamma}\Lambda & \tau^2 \dot{\xi} & \tau^2 \dot{\xi}' \\ \tau^2 \xi^T & \tau^2 \dot{\xi}^T & \tau^2 R(t,t) & \tau^2 \dot{R}(t,t) \\ \tau^2 \xi'^T & \tau^2 \dot{\xi}'^T & \tau^2 R'(t,t) & \tau^2 \dot{R}'(t,t) \end{bmatrix}
\end{aligned}
$$

where $\{Q\}_{ij}$ is the matrix with elements $R(t_i, t_j)$, $\{O\}_{ij}$ is the matrix with elements $\dot{R}(t_i, t_j) = R'(t_j, t_i)$, $\{P\}_{ij}$ is the matrix with elements $\dot{R}'(t_i, t_j)$, ξ is a $n \times 1$ matrix with ith elements $R(x_i, x)$, and $\dot{\xi}$ is a $n \times 1$ matrix with ith elements $\dot{R}(x_i, x)$. Then

$$
\begin{aligned}
E\begin{bmatrix} f & \mathbf{Y} \\ f' & \mathbf{V} \end{bmatrix} &= \begin{bmatrix} \xi^T & \dot{\xi}^T \\ \xi'^T & \dot{\xi}'^T \end{bmatrix} \begin{bmatrix} Q + n\lambda\Lambda & O + \frac{n\lambda}{\sqrt{\gamma}}\Lambda \\ O + \frac{n\lambda}{\sqrt{\gamma}}\Lambda & P + \frac{n\lambda}{\gamma}\Lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y} \\ \gamma \mathbf{V} \end{bmatrix} \\
&\triangleq \begin{bmatrix} \xi^T & \dot{\xi}^T \\ \xi'^T & \dot{\xi}'^T \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{Y} \\ \gamma \mathbf{V} \end{bmatrix} \\
&= \begin{bmatrix} \xi^T (A\mathbf{Y} + B\gamma\mathbf{V}) + \dot{\xi}^T (C\mathbf{Y} + D\gamma\mathbf{V}) \\ \xi'^T (A\mathbf{Y} + B\gamma\mathbf{V}) + \dot{\xi}'^T (C\mathbf{Y} + D\gamma\mathbf{V}) \end{bmatrix}
\end{aligned}
$$

where $n\lambda = \sigma_n^2/\tau^2$. The posterior mean $E(f|\mathbf{Y}, \mathbf{V})$ is of the form $\xi^T \mathbf{c} + \dot{\xi}^T \mathbf{d}$ and $E(f'|\mathbf{Y}, \mathbf{V})$ is of the form $\xi'^T \mathbf{c} + \dot{\xi}'^T \mathbf{d}$, with the same coefficients given by

$$
\begin{aligned}
\mathbf{c} &= A\mathbf{Y} + B\gamma\mathbf{V} \\
\mathbf{d} &= C\mathbf{Y} + D\gamma\mathbf{V}
\end{aligned}
$$

1.5 A 1-D Gaussian Process Spline Construction

Trajectories are represented by a series of 2D position points (x_t, y_t) and velocity points (u_t, v_t) corresponding to measurements taken at discrete time steps t, where x_t and u_t represented longitude, y_t and v_t represented latitude position and velocity respectively Ellis *et al.* (2009). For now, we just focus on the problem of fitting trajectories in 1 Dimension situation.

For any $t \in [t_1, t_n]$, we wish to estimate the latitude position $y(t)$ and velocity $v(t)$ with model

$$\begin{aligned} y(t) &= f(t) + \varepsilon, \\ v(t) &= f'(t) + \frac{\varepsilon}{\gamma}, \end{aligned}$$

where ε is zero-mean Gaussian noise. A Gaussian process prior over $f \sim GP(m(t), K(s,t))$ leading to the approximate estimation model

$$p(y_t, v_t | \mathbf{Y}, \mathbf{V}) \sim N(GP_\mu(\mathbf{Y}, \mathbf{V}), GP_\Sigma(\mathbf{Y}, \mathbf{V})). \tag{1.14}$$

1.5.1 Tractor Spline

Suppose we have observed data set $t_1, \cdots, t_n$ satisfying $t_1 < t_2 < \cdots < t_n$. The function $f(t)$ defined on this interval $[t_1, t_n]$ is called tractor spline, if on each interval (t_i, t_{i+1}), $i = 2, \cdots, n-2$, $f(t)$ is a cubic polynomial, but on interval (t_1, t_2) and (t_{n-1}, t_n) can be a linear function; $f(t)$ fits together at each point t_i in such a way that $f(t)$ itself and its first and second derivatives are continuous at each t_i.

Define some basis functions on interval $[t_1, t_n]$. On an arbitrary interval $[t_i, t_{i+1}]$, we have Hermite Spline basis functions as following

$$h_{00}^{(i)}(t) = \begin{cases} 2(\frac{t-t_i}{t_{i+1}-t_i})^3 - 3(\frac{t-t_i}{t_{i+1}-t_i})^2 + 1, & t_i \le t \le t_{i+1} \\ 0 & otherwise \end{cases},$$

$$h_{10}^{(i)}(t) = \begin{cases} \frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} - 2\frac{(t-t_i)^2}{t_{i+1}-t_i} + (t-t_i), & t_i \le t \le t_{i+1} \\ 0 & otherwise \end{cases},$$

$$h_{01}^{(i)}(t) = \begin{cases} -2(\frac{t-t_i}{t_{i+1}-t_i})^3 + 3(\frac{t-t_i}{t_{i+1}-t_i})^2, & t_i \le t \le t_{i+1} \\ 0 & otherwise \end{cases},$$

$$h_{11}^{(i)}(t) = \begin{cases} \frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} - \frac{(t-t_i)^2}{t_{i+1}-t_i}, & t_i \le t \le t_{i+1} \\ 0 & otherwise \end{cases}.$$

Define $N_1 = h_{00}^{(1)}$, $N_2 = h_{10}^{(1)}$, $N_{2n-1} = h_{01}^{(n)}$, $N_{2n} = h_{11}^{(n)}$. For all $k = 1, 2, \ldots, n-2$ define N_{2k+1} by

$$N_{2k+1}(t) = \begin{cases} h_{01}^{(k)} + h_{00}^{(k+1)} & t \ne t_{k+1} \\ 1 & t = t_{k+1}. \end{cases}$$

and $N_{2k+2} = h_{11}^{(k)} + h_{10}^{(k+1)}$.

We now prove that $N_1, N_2, \cdots, N_{2n}$ are linear independent.

Lemma 1. *Peng (1983) Functions $x_1(t), x_2(t), \cdots, x_n(t)$ on interval $[a, b]$, if they are linear dependent, the necessary and sufficient condition is for any $c_1, c_2, \cdots, c_n \in [a, b]$, the determinant $D(c_1, c_2, \cdots, c_n) = 0$; if they are linear independent, the necessary and sufficient condition is that there exist $c_1, c_2, \cdots, c_n \in [a, b]$, so that the determinant $D(c_1, c_2, \cdots, c_n) \neq 0$, where*

$$D(c_1, c_2, \cdots, c_n) = \begin{vmatrix} x_1(c_1) & x_1(c_2) & \cdots & x_1(c_n) \\ x_2(c_1) & x_2(c_2) & \cdots & x_2(c_n) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(c_1) & x_n(c_2) & \cdots & x_n(c_n) \end{vmatrix}$$

Theorem 1. *The functions $N_1, \ldots, N_{2n}$ provide a basis for the set of functions on $[t_1, t_n]$ which are continuous, have continuous first derivatives and which are cubic on each open interval (t_i, t_{i+1}).*

Proof. It is obviously that every basis functions are continuous on subinterval $[t_k, t_{k+1}]$. We firstly prove that these basis functions are independent.

We have $2n$ basis functions and n knots. Then choose t_1, $\frac{t_1+t_2}{2}$, t_2, $\frac{t_2+t_3}{2}$, $\cdots$, t_{n-1}, $\frac{t_{n-1}+t_n}{3}$, $\frac{2(t_{n-1}+t_n)}{3}$, t_n as new $2n$ knots, and denoted by $c_1, c_2, \cdots, c_{2n}$. Then the determinant is

$$D(c_1, c_2, \cdots, c_{2n}) = \begin{vmatrix} N_1(c_1) & N_1(c_2) & \cdots & N_1(c_{2n}) \\ N_2(c_1) & N_2(c_2) & \cdots & N_2(c_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ N_{2n}(c_1) & N_{2n}(c_2) & \cdots & N_{2n}(c_{2n}) \end{vmatrix} = \begin{vmatrix} 1 & a_{12} & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_{22} & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_{32} & 1 & a_{34} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a_{2n-1,2n-1} & 1 \\ 0 & 0 & 0 & 0 & \cdots & a_{2n,2n-1} & 0 \end{vmatrix}, \tag{1.15}$$

where $D(c_1, c_2, \cdots, c_{2n}) = \begin{cases} N_1(t_1) = 1 \\ N_1(\frac{t_1+t_2}{2}) = a_{12} \\ N_2(t_1) = 0 \\ N_2(\frac{t_1+t_2}{2}) = a_{22} \\ N_{2k+1}(\frac{t_k+t_{k+1}}{2}) = a_{2k+1,2k} & k = 1, 2, \cdots, 2n \\ N_{2k+1}(t_{k+1}) = 1 & k = 1, 2, \cdots, 2n \\ N_{2k+1}(\frac{t_{k+1}+t_{k+2}}{2}) = a_{2k+1,2k+2} & k = 1, 2, \cdots, 2n \\ N_{2k+2}(\frac{t_k+t_{k+1}}{2}) = a_{2k+2,2k} & k = 1, 2, \cdots, 2n \\ N_{2k+2}(\frac{t_{k+1}+t_{k+2}}{2}) = a_{2k+2,2k+2} & k = 1, 2, \cdots, 2n \\ N_{2n-1}(t_{2n-1}) = 0 \\ N_{2n-1}(\frac{t_{2n-1}+t_{2n}}{3}) = a_{2n-1,2n-2} \\ N_{2n-1}(\frac{2(t_{2n-1}+t_{2n})}{3}) = a_{2n-1,2n-1} \\ N_{2n-1}(t_{2n}) = 1 \\ N_{2n}(\frac{t_{2n-1}+t_{2n}}{3}) = a_{2n,2n-2} \\ N_{2n}(\frac{2(t_{2n-1}+t_{2n})}{3}) = a_{2n,2n-1} \\ N_{2n}(t_{2n}) = 0 \\ 0 & otherwise \end{cases}$, and $a_{ij} \neq$ 0.

After decomposing determinant D in equation (24), gives

$$\det D = \begin{vmatrix} a_{22} & 0 & 0 & \cdots & 0 & 0 \\ a_{32} & 1 & a_{34} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{2n-1,2n-1} & 1 \\ 0 & 0 & 0 & \cdots & a_{2n,2n-1} & 0 \end{vmatrix} = a_{22} \begin{vmatrix} 1 & a_{34} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{2n-1,2n-1} & 1 \\ 0 & 0 & \cdots & a_{2n,2n-1} & 0 \end{vmatrix}$$

$$= \cdots = a_{22}a_{44} \cdots a_{2n-4,2n-4} \begin{vmatrix} a_{2n-2,2n-2} & a_{2n-2,2n-1} & 0 \\ a_{2n-1,2n-2} & a_{2n-1,2n-1} & 1 \\ a_{2n,2n-2} & a_{2n,2n-1} & 0 \end{vmatrix}$$

$$= a_{22}a_{44} \cdots a_{2n-4,2n-4}(a_{2n-2,2n-1}a_{2n,2n-2} - a_{2n,2n-1}a_{2n-2,2n-2}) \neq 0$$

With the conclusion in Lemma 1, $N_1(t), \ldots, N_{2n}(t)$ are linearly independent on interval $[t_1, t_n]$.

Secondly, we prove that basis functions represent any cubic function on each interval $[t_k, t_{k+1}]$. Due to the definition of cubic spline, on interval $[t_k, t_{k+1}]$, a cubic spline $g(t)$ can be written in the form of

$$g(t) = d_k(t-t_k)^3 + c_k(t-t_k)^2 + b_k(t-t_k) + a_k, \text{for } t_k \le t \le t_{k+1} \tag{1.16}$$

For any $f(t)$ on $[t_1, t_n]$, it can be represented as $f(t) = \sum_{k=1}^{2n} \theta_k N_k(t)$. Then for $\forall t \in [t_k, t_{k+1}]$, we have

$$f(t) = \begin{cases} \theta_{2k-1}N_{2k-1}(t) + \theta_{2k}N_{2k}(t) + \theta_{2k+1}N_{2k+1}(t) + \theta_{2k+2}N_{2k+3}(t), & t_k \le t \le t_{k+1} \\ 0, & \text{otherwise} \end{cases},$$

thus

$$\begin{aligned} f(t) =& \theta_{2k-1}\{2(\frac{t-t_k}{t_{k+1}-t_k})^3 - 3(\frac{t-t_k}{t_{k+1}-t_k})^2 + 1\} + \theta_{2k}\{\frac{(t-t_k)^3}{(t_{k+1}-t_k)^2} - 2\frac{(t-t_k)^2}{t_{k+1}-t_k} + (t-t_k)\} \\ &+\theta_{2k+1}\{-2(\frac{t-t_k}{t_{k+1}-t_k})^3 + 3(\frac{t-t_k}{t_{k+1}-t_k})^2\} + \theta_{2k+2}\{\frac{(t-t_k)^3}{(t_{k+1}-t_k)^2} - \frac{(t-t_k)^2}{t_{k+1}-t_k}\}. \end{aligned}$$

After rearranging, we have

$$\begin{aligned} f(t) =& \{\frac{2\theta_{2k-1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k+1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k+2}}{(t_{k+1}-t_k)^2}\}(t-t_k)^3 \\ &+\{-\frac{3\theta_{2k-1}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k}}{(t_{k+1}-t_k)} + \frac{3\theta_{2k+1}}{(t_{k+1}-t_k)^2} - \frac{\theta_{2k+2}}{(t_{k+1}-t_k)}\}(t-t_k)^2 \\ &+\theta_{2k}(t-t_k) + \theta_{2k-1} \end{aligned}$$

where coefficients are

$$\begin{cases} \theta_{2k-1} = a_k \\ \theta_{2k} = b_k \\ -\frac{3\theta_{2k-1}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k}}{(t_{k+1}-t_k)} + \frac{3\theta_{2k+1}}{(t_{k+1}-t_k)^2} - \frac{\theta_{2k+2}}{(t_{k+1}-t_k)} = c_k \\ \frac{2\theta_{2k-1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k+1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k+2}}{(t_{k+1}-t_k)^2} = d_k \end{cases}$$

the resulting can always be solved for $\theta_{2k-1}, \theta_{2k}, \theta_{2k+1}, \theta_{2k+2}$ in terms of a_k, b_k, c_k, d_k on interval $[t_k, t_{k+1}]$. So cubic spline on each interval can be represented by basis functions.

Finally, we will prove basis functions are continuous on $[t_1, t_n]$. For any knot t_k, where $t_1 < t_k < t_n$, it is known that $f(t_k) = \theta_{2k-1}$. Moreover,

$$\lim_{t\to t_k+} f(t) = \lim_{t\to t_k+} (\theta_{2k-1}N_{2k-1}(t) + \theta_{2k}N_{2k}(t) + \theta_{2k+1}N_{2k+1}(t) + \theta_{2k+2}N_{2k+3}(t)) = \theta_{2k-1},$$

$$\lim_{t\to t_k-} f(t) = \lim_{t\to t_k-} (\theta_{2k-1}N_{2k-1}(t) + \theta_{2k}N_{2k}(t) + \theta_{2k+1}N_{2k+1}(t) + \theta_{2k+2}N_{2k+3}(t)) = \theta_{2k-1}.$$

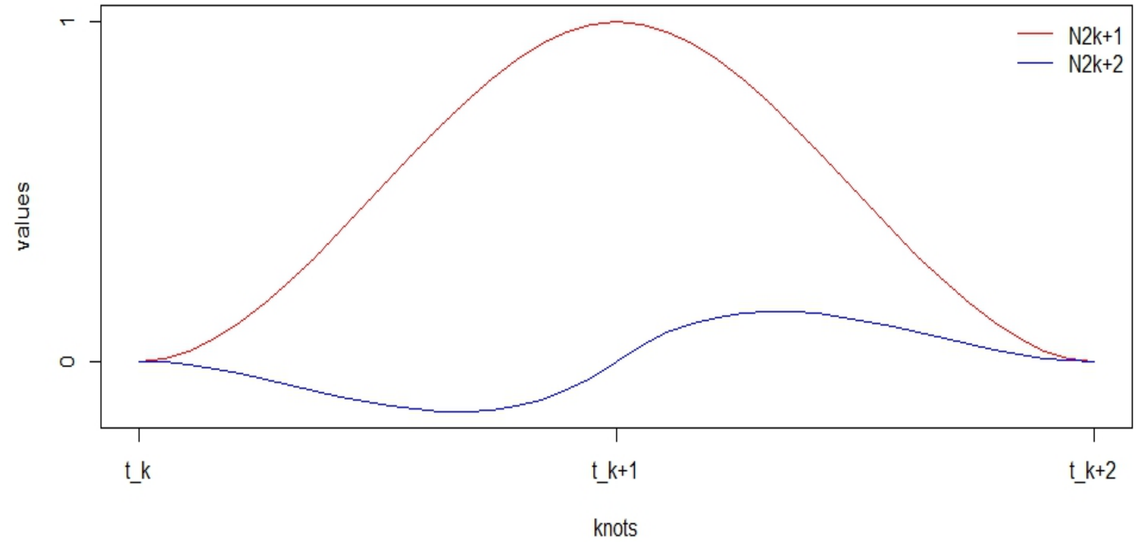

Figure 1.1: The two basis functions N_{2k+1} and N_{2k+2} on interval $[t_k, t_{k+2}]$. It is apparently that these basis functions are continuous on this interval and have continuous first and second derivatives.

So

$$\lim_{t \to t_k+} f(t) = \lim_{t \to t_k-} f(t) = f(t),$$

$f(t)$ is continuous at knots, and then continuous on whole interval $[t_1, t_n]$.

$f(t)$ is a continuous cubic spline on interval $[t_1, t_n]$, then $f(t)$ has continuous first and second derivatives. $\square$

As independent basis functions, $N_1(t), \cdots, N_{2n}(t)$ span a $2n$ dimensional space $\mathbb{H}$. For any $f \in \mathbb{H}$, f is represented in the form of

$$f = \sum_{i=1}^{2n} \theta_i N_i(t).$$

Suppose that we have observations $y_1, \cdots, y_n$ and $v_1, \cdots, v_n$. $f(t)$ can be found by minimizing equation (1.2), which reduces to

$$\mathrm{MSE}(\theta, \lambda, \gamma) = (\mathbf{Y} - \mathbf{B}\theta)^T(\mathbf{Y} - \mathbf{B}\theta) + \gamma(\mathbf{V} - \mathbf{C}\theta)^T(\mathbf{V} - \mathbf{C}\theta) + n\lambda\theta^T\Omega\theta \tag{1.17}$$

where $\{\mathbf{B}\}_{ij} = N_j(t_i)$, $\{\mathbf{C}\}_{ij} = N_j'(t_i)$ and $\{\Omega_{2n}\}_{jk} = \int N_j''(t)N_k''(t)dt$. The solution is easily seen to be

$$\hat{\theta} = (\mathbf{B}^T\mathbf{B} + \gamma\mathbf{C}^T\mathbf{C} + n\lambda\Omega)^{-1}(\mathbf{B}^T\mathbf{y} + \gamma\mathbf{C}^T\mathbf{V}) \tag{1.18}$$

a generalized ridge regression. The fitted smoothing spline is given by

$$\hat{f}(t) = \sum_{i=1}^{2n} N_i(t)\hat{\theta}_i \tag{1.19}$$

A smoothing spline with parameters λ and γ is an example of a linear smoother Trevor Hastie (2009). This is because the estimated parameters in (1.18) are a linear combination of the y_i and v_i. Denote by $\hat{\mathbf{f}}$ the $2n$ vector of fitted values $\hat{f}(t_i)$ and $\hat{\mathbf{f}}'$ the $2n$ vector of fitted values $\hat{f}'(t_i)$ at the training points t_i. Then

$$\begin{aligned}
\hat{\mathbf{f}} &= \mathbf{B}(\mathbf{B}^T\mathbf{B} + \gamma\mathbf{C}^T\mathbf{C} + n\lambda\Omega)^{-1}(\mathbf{B}^T\mathbf{y} + \gamma\mathbf{C}^T\mathbf{V}) \\
&= \mathbf{S}_{\lambda,\gamma}\mathbf{y} + \mathbf{T}_{\lambda,\gamma}\mathbf{v} \\
\hat{\mathbf{f}}' &= \mathbf{C}(\mathbf{B}^T\mathbf{B} + \gamma\mathbf{C}^T\mathbf{C} + n\lambda\Omega)^{-1}(\mathbf{B}^T\mathbf{y} + \gamma\mathbf{C}^T\mathbf{V}) \\
&= \mathbf{U}_{\lambda,\gamma}\mathbf{y} + \mathbf{V}_{\lambda,\gamma}\mathbf{v}
\end{aligned}$$

The fitted $\hat{\mathbf{f}}$ and $\hat{\mathbf{f}}'$ are linear in $\mathbf{Y}$ and $\mathbf{V}$, and the finite linear operators $\mathbf{S}_{\lambda,\gamma}, \mathbf{T}_{\lambda,\gamma}, \mathbf{U}_{\lambda,\gamma}$ and $\mathbf{V}_{\lambda,\gamma}$ are known as the smoother matrices. One consequence of this linearity is that the recipe for producing $\hat{\mathbf{f}}$ and $\hat{\mathbf{f}}'$ from $\mathbf{Y}$ and $\mathbf{V}$, do not depend on $\mathbf{Y}$ and $\mathbf{V}$ themselves; $\mathbf{S}_{\lambda,\gamma}, \mathbf{T}_{\lambda,\gamma}, \mathbf{U}_{\lambda,\gamma}$ and $\mathbf{V}_{\lambda,\gamma}$ depend only on the t_i, λ and γ.

Suppose in a traditional least squares fitting, $\mathbf{B}_\xi$ is $N \times M$ matrix of M cubic-spline basis functions evaluated at the N training points x_i, with knot sequence ξ and $M \ll N$. Then the vector of fitted spline values is given by

$$\begin{aligned}
\hat{\mathbf{f}} &= \mathbf{B}_\xi(\mathbf{B}_\xi^T\mathbf{B}_\xi)^{-1}\mathbf{B}_\xi\mathbf{Y} \\
&= \mathbf{H}_\xi\mathbf{Y}
\end{aligned}$$

Here the linear operator $\mathbf{H}_\xi$ is a symmetric, positive semidefinite matrices, and $\mathbf{H}_\xi\mathbf{H}_\xi = \mathbf{H}_\xi$ (idempotent). In our case, it is easily seen that $\mathbf{S}_{\lambda,\gamma}, \mathbf{T}_{\lambda,\gamma}, \mathbf{U}_{\lambda,\gamma}$ and $\mathbf{V}_{\lambda,\gamma}$ are symmetric, positive semidefinite matrices as well. However, only when $\lambda = \gamma = 0$, the matrix $\mathbf{S}_{\lambda=0,\gamma=0}$ is idempotent.

1.5.2 Tractor Spline Estimated by GP

With tractor spline basis functions $N_1(t), \cdots, N_{2n}(t)$, given the sample points $t_i, i = 1, \cdots, n$ in space $\mathbb{H}$ and noting that the space

$$\mathbb{A} = \{f : f = \sum_{i=1}^{n} \alpha_i R(t_i, \cdot)\}$$

is one linear subspace of $\mathbb{H}$,

$$\mathbb{B} = \{f : f = \sum_{i=1}^{n} \beta_i \dot{R}(t_i, \cdot)\}$$

is another linear subspace of $\mathbb{H}$, where $\dot{R}$ is defined in equation (1.9). After calculating, it is proved that $\mathbb{A} \cap \mathbb{B} = \emptyset$. Then $f \in \mathbb{H}$ can be written as

$$f(t) = \sum_{i=1}^{n} c_i R(t_i, t) + \sum_{i=1}^{n} d_i \dot{R}(t_i, t) + \rho(t) \tag{1.20}$$

where c_i, d_i are coefficients, $\rho(t) \in \mathbb{H} \ominus (\mathbb{A} \oplus \mathbb{B})$, and

$$f'(t) = \sum_{i=1}^{n} c_i R'(t_i, t) + \sum_{i=1}^{n} d_i \dot{R}'(t_i, t) + \rho'(t). \tag{1.21}$$

The equation (1.2) can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{n} c_j R(t_j, t_i) - \sum_{j=1}^{n} d_j \dot{R}(t_j, t_i) - \rho(t_i))^2 \\
+&\frac{\gamma}{n}\sum_{i=1}^{n}(v_i - \sum_{j=1}^{n} c_j R'(t_j, t_i) - \sum_{j=1}^{n} d_j \dot{R}'(t_j, t_i) - \rho'(t_i))^2 \\
+&\lambda \int_{t_1}^{t_n} (\sum_{j=1}^{n} c_j R''(t_j, t) + \sum_{j=1}^{n} d_j \dot{R}''(t_j, t) + \rho''(t))^2 dt
\end{aligned}$$

By orthogonality, $\rho(t_i) =< R(t_i, \cdot), \rho >= 0$, and $\rho'(t_i) =< \dot{R}(t_i, \cdot), \rho >= 0$, where $i = 1, \cdots, n$.

Denoting by Q the $n \times n$ matrix with the (i, j)th entry $R(t_i, t_j)$, by P the $n \times n$ matrix with the (i, j)th entry $\dot{R}(t_i, t_j)$ the equation (1.2) can be written as

$$(\mathbf{Y} - Q\mathbf{c} - P\mathbf{d})^T(\mathbf{Y} - Q\mathbf{c} - P\mathbf{d}) + \gamma(\mathbf{V} - \frac{\partial Q}{\partial t}\mathbf{c} - \frac{\partial P}{\partial t}\mathbf{d})^T(\mathbf{V} - \frac{\partial Q}{\partial t}\mathbf{c} - \frac{\partial P}{\partial t}\mathbf{d}) + n\lambda\Omega + \lambda(\rho, \rho). \tag{1.22}$$

Note that ρ only appears in the third term in (1.22), which is minimized at $\rho = 0$. Hence, a polynomial smoothing spline resides in the space $\mathbb{A} \oplus \mathbb{B}$ of finite dimension. Then the solution could be computed via minimization of term in (1.22) with respect to **c** and **d**, Gu (2013).

1.6 Cross Validation

Probably the simplest and most widely used method for estimating prediction error is cross-validation, Trevor Hastie (2009). Assuming that the random error has zero mean, the true regression curve f has the property that, if an observation y is taken at a point t, the value $f(t)$ is the best predictor of y in terms of mean square error, Green and Silverman (1993).

Now we focus on an observation y_i at point t_i as being a new observation by omitting it from the set of data, which are used to estimate $\hat{f}$. Denote by $\hat{f}^{(-i)}(t,\lambda)$ the estimated function from the remaining data, where λ is the smoothing parameter. Then $\hat{f}^{(-i)}(t,\lambda)$ minimizes

$$\frac{1}{n}\sum_{j\neq i}(y_j - f(t_j))^2 + \lambda \int f''^2 dt$$

and λ can be quantified by cross-validation score function

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{f}^{(-i)}(t_i,\lambda)\}^2.$$

The basis idea of cross-validation is to choose the value of λ that minimizes $CV(\lambda)$, Green and Silverman (1993).

Generalized Cross-Validation (GCV) is a modified form for choosing smoothing parameters, and the value λ can be carried out by minimizing the function

$$GCV(\lambda) = \frac{1}{n}\frac{\sum_{i=1}^{n}\{y_i - \hat{f}(t_i,\lambda)\}^2}{\{1-\frac{1}{n}tr(A(\lambda))\}^2}.$$

For weighted smoothing splines, the residual sum of squares is

$$\sum_{i=1}^{n} w_i(y_i - f(t_i))^2$$

and the cross-validation score function is

$$CV(\lambda) = \sum_{i=1}^{n} w_i\{\frac{y_i - \hat{f}(t_i,\lambda)}{(I-A(\lambda))_{ii}}\}^2.$$

Then we assume that the cross-validation score function for tractor spline (TCV) is in the form of

$$TCV(\lambda,\gamma) = \frac{1}{n}\frac{\sum|y_i - f(t_i)|^2 + \sum\gamma|v_i - f'(t_i)|^2}{|1-\frac{1}{n}tr(\hat{\theta}(\lambda,\gamma))|^2}$$

1.6.1 Leave-one-out Cross Validation

As the parameter $\hat{\theta} = (B^TB + \gamma C^TC + n\Omega_\lambda)^{-1}(B^Ty + \gamma C^Tv)$, then

$$\hat{f} = B\hat{\theta} = B(B^TB + \gamma C^TC + n\Omega_\lambda)^{-1}B^Tx + B(B^TB + \gamma C^TC + n\Omega_\lambda)^{-1}C^Tv = Sy + Tv$$

$$\hat{f}' = C\hat{\theta} = C(B^TB + \gamma C^TC + n\Omega_\lambda)^{-1}B^Tx + C(B^TB + \gamma C^TC + n\Omega_\lambda)^{-1}C^Tv = Uy + Vv$$

Then the cross validation formula could be represented in the form of

$$CV = \sum \frac{\hat{f}(t_i) - y_i + \gamma \frac{T_{ii}}{1-\gamma V_{ii}} (\hat{f}'(t_i) - v_i)}{1 - S_{ii} - \gamma \frac{T_{ii}}{1-\gamma V_{ii}} U_{ii}} \tag{1.23}$$

1.6.2 K-Fold Cross Validation

Based on the procedure given by Wahba and Wold (1975), we follow the improved steps to calculate a K-fold cross validation.

Step 1. Remove the first data t_1 and last date t_n from the dataset.

Step 2. Divide dataset into k groups:

$$\begin{aligned}
&\text{Group 1} : t_2, t_{2+k}, \cdots \\
&\text{Group 2} : t_3, t_{3+k}, \cdots \\
&\vdots \\
&\text{Group k} : t_{k+1}, t_{2k+1}, \cdots
\end{aligned}$$

Step 3. Guess values of $\lambda_{down}, \lambda_{up}$ and γ.

Step 4. Delete the first group of data. Fit a smoothing spline to the first data, the rest groups of dataset and the last data, with $\lambda_{down}, \lambda_{up}$ and γ in step 3. Compute the sum of squared deviations of this smoothing spline from the deleted data points.

Step 5. Delete instead the second group of data. Fit a smoothing spline to the remaining data with $\lambda_{down}, \lambda_{up}$ and γ. Compute the sum of squared deviations of the spline from deleted data points.

Step 6. Repeat Step 5 for the 3rd, 4th, $\cdots$, kth group of data.

Step 7. Add the sums of squared deviations from steps 4 to 6 and divide by k. This is the cross validation score of three parameters $\lambda_{down}, \lambda_{up}$ and γ.

Step 8. Vary $\lambda_{down}, \lambda_{up}$ and γ systematically and repeat steps 4-7 until CV shows a minimum.

References

Ellis, D., Sommerlade, E., and Reid, I. (2009). Modelling pedestrian trajectory patterns with gaussian processes. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 1229–1234. IEEE.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach.* CRC Press.

Gu, C. (2013). *Smoothing Spline ANOVA Models Second Edition.* Springer New York Heidelberg Dordrecht London.

Peng, F. (1983). A Necessary and Sufficient Condition to Judg Function Linearly Independent. *JOURNAL OF JISHOU UNIVERSITY(NATURAL SCIENCE EDITION)*, (0), 25–28.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* The MIT Press.

Trevor Hastie, Robert Tibshirani, J. F. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition.* Springer-Verlag.

Wahba, G. and Wold, S. (1975). A completely automatic French curve: Fitting spline functions by cross validation. *Communications in Statistics-Theory and Methods*, *4*(1), 1–17.

Penalty Matrix in (??)

The k-th $\Omega^{(k)}$ is a $2n \times 2n$ matrix in the form of

$$\begin{aligned}
\Omega^{(k)}_{2k-1,2k-1} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2} \frac{d^2 h^{(k)}_{00}(t)}{dt^2} dt = \frac{12}{\Delta_k^3} \\
\Omega^{(k)}_{2k-1,2k} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} dt = \frac{6}{\Delta_k^2} \\
\Omega^{(k)}_{2k-1,2k+1} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} dt = \frac{-12}{\Delta_k^3} \\
\Omega^{(k)}_{2k-1,2k+2} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{6}{\Delta_k^2} \\
\Omega^{(k)}_{2k,2k} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} dt = \frac{4}{\Delta_k^2} \\
\Omega^{(k)}_{2k,2k+1} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} dt = \frac{-6}{\Delta_k^2} \\
\Omega^{(k)}_{2k,2k+2} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{2}{\Delta_k} \\
\Omega^{(k)}_{2k+1,2k+1} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} dt = \frac{12}{\Delta_k^3} \\
\Omega^{(k)}_{2k+1,2k+2} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{-6}{\Delta_k^2} \\
\Omega^{(k)}_{2k+2,2k+2} &= \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{4}{\Delta_k}
\end{aligned}$$

$k = 1, 2, \cdots, n-1$. It's a bandwidth four matrix. Then

$$\mathbf{\Omega} = \sum_{k=1}^{n-1} \Omega^{(k)}$$

Penalty Matrix in (1.22)

The penalty matrix Ω in (1.22) is a combination of three sub matrix Ω_1, Ω_2 and Ω_3, which are in the following form

$$\Omega^{(1)}_{k,k} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2}\frac{d^2 h^{(k)}_{00}(t)}{dt^2} dt = \frac{12}{\Delta_k^3}$$
$$\Omega^{(1)}_{k,k+1} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2}\frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} dt = \frac{-12}{\Delta_k^3}$$
$$\Omega^{(1)}_{k+1,k+1} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k+1)}_{01}(t)}{dt^2}\frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} dt = \frac{12}{\Delta_k^3}$$
$$\Omega^{(2)}_{k,k} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2}\frac{d^2 h^{(k)}_{10}(t)}{dt^2} dt = \frac{4}{\Delta_k^2}$$
$$\Omega^{(2)}_{k,k+1} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2}\frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{2}{\Delta_k}$$
$$\Omega^{(2)}_{k+1,k+1} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2}\frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{4}{\Delta_k}$$
$$\Omega^{(3)}_{k,k} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2}\frac{d^2 h^{(k)}_{10}(t)}{dt^2} dt = \frac{6}{\Delta_k^2}$$
$$\Omega^{(3)}_{k,k+1} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{00}(t)}{dt^2}\frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{6}{\Delta_k^2}$$
$$\Omega^{(3)}_{k+1,k} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2}\frac{d^2 h^{(k+1)}_{01}(t)}{dt^2} dt = \frac{-6}{\Delta_k^2}$$
$$\Omega^{(3)}_{k+1,k+1} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k+1)}_{01}(t)}{dt^2}\frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{-6}{\Delta_k^2}$$

Proof of Theorem ??

Proof. It is obviously that every basis functions are continuous on subinterval $[t_k, t_{k+1}]$. We firstly prove that these basis functions are independent.

We have $2n$ basis functions and n knots. Then choose t_1, $\frac{t_1+t_2}{2}$, t_2, $\frac{t_2+t_3}{2}$, $\cdots$, t_{n-1}, $\frac{t_{n-1}+t_n}{3}$, $\frac{2(t_{n-1}+t_n)}{3}$, t_n as new $2n$ knots, and denoted by $c_1, c_2, \cdots, c_{2n}$. Then the determinant is

$$D(c_1, c_2, \cdots, c_{2n}) = \begin{vmatrix} N_1(c_1) & N_1(c_2) & \cdots & N_1(c_{2n}) \\ N_2(c_1) & N_2(c_2) & \cdots & N_2(c_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ N_{2n}(c_1) & N_{2n}(c_2) & \cdots & N_{2n}(c_{2n}) \end{vmatrix} = \begin{vmatrix} 1 & a_{12} & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_{22} & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_{32} & 1 & a_{34} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a_{2n-1,2n-1} & 1 \\ 0 & 0 & 0 & 0 & \cdots & a_{2n,2n-1} & 0 \end{vmatrix}, \tag{24}$$

$$
\text{where } D(c_1, c_2, \cdots, c_{2n}) = \begin{cases}
N_1(t_1) = 1 & \\
N_1(\frac{t_1+t_2}{2}) = a_{12} & \\
N_2(t_1) = 0 & \\
N_2(\frac{t_1+t_2}{2}) = a_{22} & \\
N_{2k+1}(\frac{t_k+t_{k+1}}{2}) = a_{2k+1,2k} & k = 1, 2, \cdots, 2n \\
N_{2k+1}(t_{k+1}) = 1 & k = 1, 2, \cdots, 2n \\
N_{2k+1}(\frac{t_{k+1}+t_{k+2}}{2}) = a_{2k+1,2k+2} & k = 1, 2, \cdots, 2n \\
N_{2k+2}(\frac{t_k+t_{k+1}}{2}) = a_{2k+2,2k} & k = 1, 2, \cdots, 2n \\
N_{2k+2}(\frac{t_{k+1}+t_{k+2}}{2}) = a_{2k+2,2k+2} & k = 1, 2, \cdots, 2n, \\
N_{2n-1}(t_{2n-1}) = 0 & \\
N_{2n-1}(\frac{t_{2n-1}+t_{2n}}{3}) = a_{2n-1,2n-2} & \\
N_{2n-1}(\frac{2(t_{2n-1}+t_{2n})}{3}) = a_{2n-1,2n-1} & \\
N_{2n-1}(t_{2n}) = 1 & \\
N_{2n}(\frac{t_{2n-1}+t_{2n}}{3}) = a_{2n,2n-2} & \\
N_{2n}(\frac{2(t_{2n-1}+t_{2n})}{3}) = a_{2n,2n-1} & \\
N_{2n}(t_{2n}) = 0 & \\
0 & otherwise
\end{cases}
$$

and $a_{ij} \neq 0$. After decomposing determinant D in equation (24), gives

$$
\begin{aligned}
\det D = &\begin{vmatrix}
a_{22} & 0 & 0 & \cdots & 0 & 0 \\
a_{32} & 1 & a_{34} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & a_{2n-1,2n-1} & 1 \\
0 & 0 & 0 & \cdots & a_{2n,2n-1} & 0
\end{vmatrix} = a_{22} \begin{vmatrix}
1 & a_{34} & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & a_{2n-1,2n-1} & 1 \\
0 & 0 & \cdots & a_{2n,2n-1} & 0
\end{vmatrix} \\
&= \cdots = a_{22}a_{44} \cdots a_{2n-4,2n-4} \begin{vmatrix}
a_{2n-2,2n-2} & a_{2n-2,2n-1} & 0 \\
a_{2n-1,2n-2} & a_{2n-1,2n-1} & 1 \\
a_{2n,2n-2} & a_{2n,2n-1} & 0
\end{vmatrix} \\
&= a_{22}a_{44} \cdots a_{2n-4,2n-4}(a_{2n-2,2n-1}a_{2n,2n-2} - a_{2n,2n-1}a_{2n-2,2n-2}) \neq 0.
\end{aligned}
$$

With the conclusion in Lemma 1, $N_1(t), \ldots, N_{2n}(t)$ are linearly independent on interval $[t_1, t_n]$.

Secondly, we prove that basis functions represent any cubic function on each interval $[t_k, t_{k+1}]$. Due to the definition of cubic spline, on interval $[t_k, t_{k+1}]$, a cubic spline $g(t)$

can be written in the form of

$$g(t) = d_k(t-t_k)^3 + c_k(t-t_k)^2 + b_k(t-t_k) + a_k, \text{for } t_k \le t \le t_{k+1} \tag{25}$$

For any $f(t)$ on $[t_1, t_n]$, it can be represented as $f(t) = \sum_{k=1}^{2n} \theta_k N_k(t)$. Then for $\forall t \in [t_k, t_{k+1}]$, we have

$$f(t) = \begin{cases} \theta_{2k-1}N_{2k-1}(t) + \theta_{2k}N_{2k}(t) + \theta_{2k+1}N_{2k+1}(t) + \theta_{2k+2}N_{2k+3}(t), & t_k \le t \le t_{k+1} \\ 0, & \text{otherwise} \end{cases},$$

thus

$$\begin{aligned} f(t) =& \theta_{2k-1}\{2(\frac{t-t_k}{t_{k+1}-t_k})^3 - 3(\frac{t-t_k}{t_{k+1}-t_k})^2 + 1\} + \theta_{2k}\{\frac{(t-t_k)^3}{(t_{k+1}-t_k)^2} - 2\frac{(t-t_k)^2}{t_{k+1}-t_k} + (t-t_k)\} \\ &+ \theta_{2k+1}\{-2(\frac{t-t_k}{t_{k+1}-t_k})^3 + 3(\frac{t-t_k}{t_{k+1}-t_k})^2\} + \theta_{2k+2}\{\frac{(t-t_k)^3}{(t_{k+1}-t_k)^2} - \frac{(t-t_k)^2}{t_{k+1}-t_k}\}. \end{aligned}$$

After rearranging, we have

$$\begin{aligned} f(t) =& \{\frac{2\theta_{2k-1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k+1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k+2}}{(t_{k+1}-t_k)^2}\}(t-t_k)^3 \\ &+ \{-\frac{3\theta_{2k-1}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k}}{(t_{k+1}-t_k)} + \frac{3\theta_{2k+1}}{(t_{k+1}-t_k)^2} - \frac{\theta_{2k+2}}{(t_{k+1}-t_k)}\}(t-t_k)^2 \\ &+ \theta_{2k}(t-t_k) + \theta_{2k-1} \end{aligned}$$

where coefficients are

$$\begin{cases} \theta_{2k-1} = a_k \\ \theta_{2k} = b_k \\ -\frac{3\theta_{2k-1}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k}}{(t_{k+1}-t_k)} + \frac{3\theta_{2k+1}}{(t_{k+1}-t_k)^2} - \frac{\theta_{2k+2}}{(t_{k+1}-t_k)} = c_k \\ \frac{2\theta_{2k-1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k}}{(t_{k+1}-t_k)^2} - \frac{2\theta_{2k+1}}{(t_{k+1}-t_k)^3} + \frac{\theta_{2k+2}}{(t_{k+1}-t_k)^2} = d_k \end{cases}$$

the resulting can always be solved for $\theta_{2k-1}, \theta_{2k}, \theta_{2k+1}, \theta_{2k+2}$ in terms of a_k, b_k, c_k, d_k on interval $[t_k, t_{k+1}]$. So cubic spline on each interval can be represented by basis functions.

Finally, we will prove basis functions are continuous on $[t_1, t_n]$. For any knot t_k, where $t_1 < t_k < t_n$, it is known that $f(t_k) = \theta_{2k-1}$. Moreover,

$$\lim_{t\to t_k+} f(t) = \lim_{t\to t_k+}(\theta_{2k-1}N_{2k-1}(t) + \theta_{2k}N_{2k}(t) + \theta_{2k+1}N_{2k+1}(t) + \theta_{2k+2}N_{2k+3}(t)) = \theta_{2k-1},$$

$$\lim_{t\to t_k-} f(t) = \lim_{t\to t_k-}(\theta_{2k-1}N_{2k-1}(t) + \theta_{2k}N_{2k}(t) + \theta_{2k+1}N_{2k+1}(t) + \theta_{2k+2}N_{2k+3}(t)) = \theta_{2k-1}.$$

So

$$\lim_{t\to t_k+} f(t) = \lim_{t\to t_k-} f(t) = f(t),$$

$f(t)$ is continuous at knots, and then continuous on whole interval $[t_1, t_n]$.

$f(t)$ is a continuous cubic spline on interval $[t_1, t_n]$, then $f(t)$ has continuous first and second derivatives. $\square$