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.

March 9, 2016

Contents

1.1 Introduction 1
1.2 Gaussian Process Regression 2
1.3 A Reproducing Kernel in Space $\mathbb{H}$ 4
1.4 Covariance Matrix and Posterior Mean 6
1.5 A 1-D Gaussian Process Spline Construction 7
1.5.1 Tractor Spline 7
1.5.2 Tractor Spline Estimated by GP 11
1.6 Cross Validation 13
1.6.1 K-Fold Cross Validation 15

References **17**

List of Tables

List of Figures

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. 9

1.1 Introduction

In regression problem, linear regression, linear discriminant analysis, logistic regression and separating hyperplanes all rely on a linear model. With the good property of linear model, easy to be interpreted and first order Taylor approximation to $f(t)$, it is more convenient to represent $f(t)$ by linear model. However, the true function $f(t)$ is unlikely to be an actual linear function in space $\mathbb{R}$. Researchers found some methods for moving beyond linearity. One of them is replacing the vector of inputs $\mathbf{T}$ with its transformations as new variables, and then use linear models in this new space of derived input features.

Denote by $h_m(t) : \mathbb{R} \mapsto \mathbb{R}$ the mth transformation of t, $m = 1, \cdots, M$. We then model

$$f(t) = \sum_{m=1}^{M} \beta_m h_m(t). \tag{1.1}$$

a linear basis expansion of $\mathbf{t}$ in $\mathbb{R}$, where $h_m(t)$ are named basis functions, β_m are coefficients. Once the basis functions h_m have been determined, the models are linear in these new variables, and the fitting proceeds as before.

Suppose we are given observed data $t_1, t_2, \cdots, t_n$ on interval $[0, 1]$, satisfying $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$. A piecewise polynomial function $f(t)$ can be obtained by dividing the interval into contiguous intervals $(t_1, t_2), \cdots, (t_{n-1}, t_n)$, and representing f by a separate polynomial in each interval. The points t_i are called knots. For example,

$$f(t) = d_i(t - t_i)^3 + c_i(t - t_i)^2 + b_i(t - t_i) + a_i, \tag{1.2}$$

for given coefficients d_i, c_i, b_i and a_i, where $t_i \leq t \leq t_{i+1}$, $i = 1, 2, \cdots, n$. f is a cubic spline on $[0, 1]$ if (1) on each intervals f is a polynomial; (2) the polynomial pieces fit together at knots t_i in such a way that f itself and its first and second derivatives are continuous at each t_i. If the second and third derivatives of f are zero at 0 and 1, f is said to be a natural cubic spline. These conditions are called natural boundary conditions.

Over all spline functions $f(t)$ with two continuous derivatives fitting these observed data, the curve estimate $\hat{f}(t)$ will be defined to be the minimizer the following penalized residual sum of squares, **[edited expressions]**

$$\text{MSE}(f, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (f(t_i) - y_i)^2 + \lambda \int_0^1 (f''(t))^2 dt \tag{1.3}$$

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 (1.3), the smoothing parameter λ controls the trade-off between over-fitting and bias,

$$\begin{cases} \lambda = 0: & f \text{ can be any function that interpolates the data,} \\ \lambda = \infty: & \text{the simple least squares line fit since no second derivative can be tolerated.} \end{cases} \tag{1.4}$$

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 $f(t)$ simultaneously. f is a linear combination of basis functions, as shown in equation (1.1), in the meantime, f' is a linear combination of the first derivative of these basis functions

$$f'(t) = \sum_{m=1}^{M} \alpha_m h'_m(t). \tag{1.5}$$

The velocity information is incorporated into MSE equation (1.3) by the addition of velocity term $(f'(t_i) - v_i)^2$. Then it becomes

$$\mathrm{MSE}(f, \lambda, \gamma) = \frac{1}{n} \sum_{i=1}^{n} (f(t_i) - y_i)^2 + \frac{\gamma}{n} \sum_{i=1}^{n} (f'(t_i) - v_i)^2 + \lambda \int_0^1 (f''(t))^2 dt, \tag{1.6}$$

and $\hat{f}$ is the minimizer of the MSE equation (1.6).

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 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

$$m(t) = \mathbb{E}[f(t)] \tag{1.7}$$

$$K(s, t) = \mathbb{E}[(f(s) - m(s))(f(t) - m(t))], \tag{1.8}$$

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

$$f \sim GP(m(t), K(s, t)). \tag{1.9}$$

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

$$\mathrm{cov}(y_p, y_q) = K(t_p, t_q) + \sigma_n^2 \delta_{pq} \tag{1.10}$$

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

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\mathbf{T}, \mathbf{T}) + \sigma_n^2 I & K(\mathbf{T}, t_*) \\ K(t_*, \mathbf{T}) & K(t_*, t_*) \end{bmatrix}\right). \tag{1.11}$$

The posterior distribution over the predicted value is obtained by conditioning on the observed data

$$f_* | \mathbf{y}, \mathbf{T}, t_* \sim N(\bar{f}_*, \mathrm{cov}(f_*)) \tag{1.12}$$

where

$$\bar{f}_* = \mathbb{E}[f_* | \mathbf{y}, \mathbf{T}, t_*] = K(t_*, \mathbf{T})[K(\mathbf{T}, \mathbf{T}) + \sigma_n^2 I]^{-1} \mathbf{y}, \tag{1.13}$$

$$\mathrm{cov}(f_*) = K(t_*, t_*) - K(t_*, \mathbf{T})[K(\mathbf{T}, \mathbf{T}) + \sigma_n^2 I]^{-1} K(\mathbf{T}, t_*). \tag{1.14}$$

We now add velocity information $\mathbf{v} = f'(\mathbf{T}) + \varepsilon'$, where ε' is independent distributed Gaussian noise with variance $\frac{\sigma_n^2}{\gamma}$.

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

$$\Sigma(\mathbf{y}, \mathbf{v}) = \begin{bmatrix} \mathrm{cov}(\mathbf{y}, \mathbf{y}) & \mathrm{cov}(\mathbf{y}, \mathbf{v}) \\ \mathrm{cov}(\mathbf{v}, \mathbf{y}) & \mathrm{cov}(\mathbf{v}, \mathbf{v}) \end{bmatrix}, \tag{1.15}$$

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

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{v} \end{bmatrix} \sim N(\mu_{y,v}, \Sigma_{y,v}). \tag{1.16}$$

Define f_* and f'_* the estimated position and velocity values at point t_*. From equation

(1.15) and using similar idea, it is easily to get the covariance matrices

$$\Sigma(f_*, \mathbf{v}) = \begin{bmatrix} \mathrm{cov}(f_*, f_*) & \mathrm{cov}(f_*, \mathbf{v}) \\ \mathrm{cov}(\mathbf{v}, f_*) & \mathrm{cov}(\mathbf{v}, \mathbf{v}) \end{bmatrix},$$
$$\Sigma(\mathbf{y}, f'_*) = \begin{bmatrix} \mathrm{cov}(\mathbf{y}, \mathbf{y}) & \mathrm{cov}(\mathbf{y}, f'_*) \\ \mathrm{cov}(f'_*, \mathbf{y}) & \mathrm{cov}(f'_*, f'_*) \end{bmatrix}, \tag{1.17}$$
$$\Sigma(f_*, f'_*) = \begin{bmatrix} \mathrm{cov}(f_*, f_*) & \mathrm{cov}(f_*, f'_*) \\ \mathrm{cov}(f'_*, f_*) & \mathrm{cov}(f'_*, f'_*) \end{bmatrix},$$

[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 function having first derivative in space $\mathbb{H}$. For any continuous function $f \in \mathbb{H}$, it is a combination of these basis functions

$$f(t) = \sum_{i=1}^{n} \alpha_i N_i(t), \tag{1.18}$$

where $\alpha_i (i = 1, \cdots, n)$ are coefficients. With an inner product

$$\langle f, g \rangle = \langle \sum_{i=1}^{n} \alpha_i N_i(t), \sum_{i=1}^{n} \beta_i N_i(t) \rangle = \sum_{i=1}^{n} \alpha_i \beta_i, \tag{1.19}$$

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

$$R_s(t) = \sum_{i=1}^{n} N_i(s) N_i(t), \tag{1.20}$$

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

$$\langle f(t), R(s,t) \rangle = \langle \sum_{i=1}^{n} \alpha_i N_i(t), \sum_{i=1}^{n} N_i(s) N_i(t) \rangle = \sum_{i=1}^{n} \alpha_i N_i(s) = f(s). \tag{1.21}$$

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

$$\dot{R}(s,t) = \frac{\partial R(s,t)}{\partial s} = \sum_{i=1}^{n} \frac{dN_i(s)}{ds} N_i(t) = \sum_{i=1}^{n} N'_i(s) N_i(t), \tag{1.22}$$
$$R'(s,t) = \frac{\partial R(s,t)}{\partial t} = \sum_{i=1}^{n} N_i(s) \frac{dN_i(t)}{dt} = \sum_{i=1}^{n} N_i(s) N'_i(t), \tag{1.23}$$

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

$$\dot{R}'(s,t) = \frac{\partial^2 R(s,t)}{\partial s \partial t} = \sum_{i=1}^{n} \frac{dN_i(s)}{ds}\frac{dN_i(t)}{dt} = \sum_{i=1}^{n} N_i'(s)N_i'(t). \tag{1.24}$$

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

$$\begin{aligned} \langle f(t), R'(s,t)\rangle &= \langle \sum_i \alpha_i N_i(t), \sum_i N_i(s)N_i'(t)\rangle = \sum_i \alpha_i N_i(t) = f(t), \\ \langle f(t), \dot{R}(s,t)\rangle &= \langle \sum_i \alpha_i N_i(t), \sum_i N_i'(s)N_i(t)\rangle = \sum_i \alpha_i N_i'(s) = f'(s), \end{aligned} \tag{1.25}$$

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

$$\mathbb{A} = \{f : f = \sum_{i=1}^{n} \alpha_i R(t_i, \cdot)\} \tag{1.26}$$

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

$$f(t) = \sum_{i=1}^{n} c_i R(t_i, t) + \rho(t) \tag{1.27}$$

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

$$f'(t) = \sum_{i=1}^{n} c_i R'(t_i, t) + \rho'(t). \tag{1.28}$$

The equation (1.6) can be written inform of

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{n} c_j R(t_j, t_i) - \rho(t_i))^2 + \frac{\gamma}{n}\sum_{i=1}^{n}(v_i - \sum_{j=1}^{n} c_j R'(t_j, t_i) - \rho'(t_i))^2 \\ + \lambda \int_0^1 (\sum_{j=1}^{n} c_j R''(t_j, t) + \rho''(t))^2 dt \end{aligned} \tag{1.29}$$

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, $\rho(t_i) = \langle R(t_i, \cdot), \rho\rangle = 0$, and $\rho'(t_i) = \langle \rho, \dot{R}(t_i, \cdot)\rangle = 0$, 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 (1.6) can be written as

$$(\mathbf{y} - Q\mathbf{c})^\top(\mathbf{y} - Q\mathbf{c}) + \gamma(\mathbf{v} - P\mathbf{c})^\top(\mathbf{v} - P\mathbf{c}) + n\lambda\Omega + \lambda(\rho, \rho). \tag{1.30}$$

Note that ρ only appears in the third term in (1.64), which is minimized at $\rho = 0$. Hence, a polynomial smoothing spline resides in the space $\mathbb{A}$ of finite dimension. Then, following the method given from Gu (2013), the solution can be computed via minimization of term in (1.30) with respect to $\mathbf{c}$

1.4 Covariance Matrix and Posterior Mean

Consider f and f' in $\mathbb{H}$, having Gaussian priors with zero mean. By equation (1.10), their covariance functions are

$$\begin{aligned}
\mathrm{cov}(f(s), f(t)) &= \tau^2 R(s,t) + \sigma_n^2 I \\
\mathrm{cov}(f(s), f'(t)) &= \tau^2 R'(s,t) + \frac{\sigma_n^2}{\sqrt{\gamma}} I \\
\mathrm{cov}(f'(s), f(t)) &= \tau^2 \dot{R}(s,t) + \frac{\sigma_n^2}{\sqrt{\gamma}} I \\
\mathrm{cov}(f'(s), f'(t)) &= \tau^2 \dot{R}'(s,t) + \frac{\sigma_n^2}{\gamma} I
\end{aligned} \tag{1.31}$$

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

$$\begin{aligned}
\mathrm{cov}(\mathbf{y}, \mathbf{v}, f, f') &= \begin{bmatrix} \tau^2 R(t_i,t_j) + \sigma_n^2 I & \tau^2 R'(t_i,t_j) + \frac{\sigma_n^2}{\sqrt{\gamma}} I & \tau^2 R(t_i,t) & \tau^2 R'(t_i,t) \\ \tau^2 \dot{R}(t_i,t_j) + \frac{\sigma_n^2}{\sqrt{\gamma}} I & \tau^2 \dot{R}'(t_i,t_j) + \frac{\sigma_n^2}{\gamma} I & \tau^2 \dot{R}(t_i,t) & \tau^2 \dot{R}'(t_i,t) \\ \tau^2 R^\top(t_i,t) & \tau^2 \dot{R}^\top(t_i,t) & \tau^2 R(t,t) & \tau^2 R'(t,t) \\ \tau^2 R'^\top(t_i,t) & \tau^2 \dot{R}'^\top(t_i,t) & \tau^2 \dot{R}(t,t) & \tau^2 \dot{R}'(t,t) \end{bmatrix} \\
&= \begin{bmatrix} \tau^2 Q + \sigma_n^2 I & \tau^2 O + \frac{\sigma_n^2}{\sqrt{\gamma}} I & \tau^2 \xi & \tau^2 \xi' \\ \tau^2 O + \frac{\sigma_n^2}{\sqrt{\gamma}} I & \tau^2 P + \frac{\sigma_n^2}{\gamma} I & \tau^2 \dot{\xi} & \tau^2 \dot{\xi}' \\ \tau^2 \xi^\top & \tau^2 \dot{\xi}^\top & \tau^2 R(t,t) & \tau^2 R'(t,t) \\ \tau^2 \xi'^\top & \tau^2 \dot{\xi}'^\top & \tau^2 \dot{R}(t,t) & \tau^2 \dot{R}'(t,t) \end{bmatrix}
\end{aligned} \tag{1.32}$$

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

$$\begin{aligned}
E \begin{bmatrix} f \\ f' \end{bmatrix} \Big| \begin{matrix} \mathbf{y} \\ \mathbf{v} \end{matrix} &= \begin{bmatrix} \xi^\top & \dot{\xi}^\top \\ \xi'^\top & \dot{\xi}'^\top \end{bmatrix} \begin{bmatrix} Q + n\lambda I & O + \frac{n\lambda}{\sqrt{\gamma}} I \\ O + \frac{n\lambda}{\sqrt{\gamma}} I & P + \frac{n\lambda}{\gamma} I \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ \gamma \mathbf{v} \end{bmatrix} \\
&\triangleq \begin{bmatrix} \xi^\top & \dot{\xi}^\top \\ \xi'^\top & \dot{\xi}'^\top \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \gamma \mathbf{v} \end{bmatrix} \\
&= \begin{bmatrix} \xi^\top (A\mathbf{y} + B\gamma\mathbf{v}) + \dot{\xi}^\top (C\mathbf{y} + D\gamma\mathbf{v}) \\ \xi'^\top (A\mathbf{y} + B\gamma\mathbf{v}) + \dot{\xi}'^\top (C\mathbf{y} + D\gamma\mathbf{v}) \end{bmatrix}
\end{aligned} \tag{1.33}$$

where $n\lambda = \sigma_n^2/\tau^2$. The posterior mean $E(f|\mathbf{y}, \mathbf{v})$ is a linear combination of basis functions $N_i(t)$, and both ξ and $\dot{\xi}$ contain $N_i(t)$, thus the posterior mean is of the form

$\xi^\top \mathbf{c} + \dot{\xi}^\top \mathbf{d}$. Similarly, $E(f'|\mathbf{y}, \mathbf{v})$ is of the form $\xi'^\top \mathbf{c} + \dot{\xi}'^\top \mathbf{d}$, with the same coefficients given by

$$\mathbf{c} = A\mathbf{y} + B\gamma\mathbf{v} \tag{1.34}$$

$$\mathbf{d} = C\mathbf{y} + D\gamma\mathbf{v} \tag{1.35}$$

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

$$y(t) = f(t) + \varepsilon, \tag{1.36}$$

$$v(t) = f'(t) + \frac{\varepsilon}{\gamma}, \tag{1.37}$$

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

$$p(y_t, v_t|\mathbf{y}, \mathbf{v}) \sim N(GP_\mu(\mathbf{y}, \mathbf{v}), GP_\Sigma(\mathbf{y}, \mathbf{v})). \tag{1.38}$$

1.5.1 Tractor Spline

Suppose we have observed dataset $t_1 < t_2 < \cdots < t_n$. 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, $i = 2, \cdots, n-2$.

On an arbitrary interval $[t_i, t_{i+1}]$, we have Hermite Spline basis functions as following

$$h_{00}^{(i)}(t) = \begin{cases} 2(\frac{t-t_i}{t_{i+1}-t_i})^3 - 3(\frac{t-t_i}{t_{i+1}-t_i})^2 + 1, & t_i \leq t \leq t_{i+1} \\ 0 & otherwise \end{cases}, \tag{1.39}$$

$$h_{10}^{(i)}(t) = \begin{cases} \frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} - 2\frac{(t-t_i)^2}{t_{i+1}-t_i} + (t - t_i), & t_i \leq t \leq t_{i+1} \\ 0 & otherwise \end{cases}, \tag{1.40}$$

$$h_{01}^{(i)}(t) = \begin{cases} -2(\frac{t-t_i}{t_{i+1}-t_i})^3 + 3(\frac{t-t_i}{t_{i+1}-t_i})^2, & t_i \leq t \leq t_{i+1} \\ 0 & otherwise \end{cases}, \tag{1.41}$$

$$h_{11}^{(i)}(t) = \begin{cases} \frac{(t-t_i)^3}{(t_{i+1}-t_i)^2} - \frac{(t-t_i)^2}{t_{i+1}-t_i}, & t_i \leq t \leq t_{i+1} \\ 0 & otherwise \end{cases}. \tag{1.42}$$

Construct new basis functions on entire interval $[t_1, t_n]$ in such way, that $N_1 = h_{00}^{(1)}$, $N_2 = h_{10}^{(1)}$, $N_{2n-1} = h_{01}^{(n)}$, $N_{2n} = h_{11}^{(n)}$. For all $k = 1, 2, \ldots, n-2$ define N_{2k+1} by

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

and $N_{2k+2} = h_{11}^{(k)} + h_{10}^{(k+1)}$. Then $N_1(t), \cdots, N_{2n}(t)$ are the new basis functions on $[t_1, t_n]$.

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*

$$D(c_1, c_2, \cdots, c_n) = \begin{vmatrix} x_1(c_1) & x_1(c_2) & \cdots & x_1(c_n) \\ x_2(c_1) & x_2(c_2) & \cdots & x_2(c_n) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(c_1) & x_n(c_2) & \cdots & x_n(c_n) \end{vmatrix} \tag{1.43}$$

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}).*

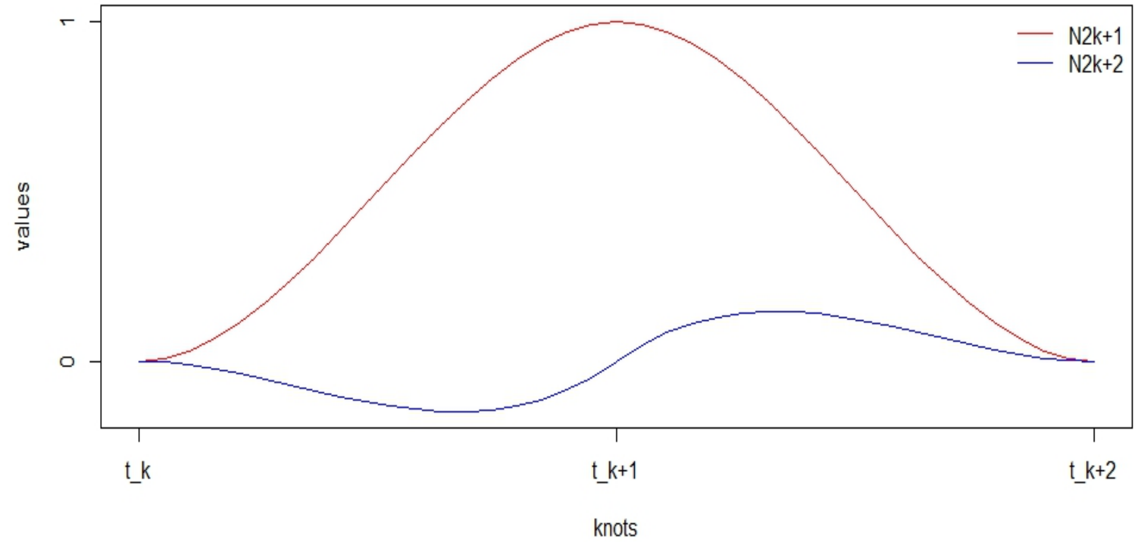

Figure 1.1: The two basis functions N_{2k+1} and N_{2k+2} on interval $[t_k, t_{k+2}]$. It is apparently that these basis functions are continuous on this interval and have continuous first and second derivatives.

The proof of theorem 1 is in appendices.

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

$$f = \sum_{i=1}^{2n} \theta_i N_i(t). \tag{1.44}$$

Suppose that we have observations $y_1, \cdots, y_n$ and $v_1, \cdots, v_n$. $f(t)$ can be found by minimizing equation (1.6), which reduces to

$$\mathrm{MSE}(\theta, \lambda, \gamma) = (\mathbf{y} - \mathbf{B}\theta)^\top (\mathbf{y} - \mathbf{B}\theta) + \gamma(\mathbf{v} - \mathbf{C}\theta)^\top (\mathbf{v} - \mathbf{C}\theta) + n\lambda\theta^\top \Omega\theta \tag{1.45}$$

where $\{\mathbf{B}\}_{ij} = N_j(t_i)$, $\{\mathbf{C}\}_{ij} = N_j'(t_i)$ and $\{\Omega_{2n}\}_{jk} = \int N_j''(t) N_k''(t) dt$. After substituting the series observation $t_1, \cdots, t_n$ into basis functions, we get $N_1(t_1) = 1, N_1(t_2) = 0, \cdots, N_{2k-1}(t_k) = 1, N_{2k}(t_k) = 0, \cdots, N_{2n-1}(t_n) = 1, N_{2n}(t_n) = 0$; and into first derivative of basis functions, we get $N_1'(t_1) = 0, N_1'(t_2) = 1, \cdots, N_{2k-1}'(t_k) = 0, N_{2k}'(t_k) = 1, \cdots, N_{2n-1}'(t_n) = 0, N_{2n}'(t_n) = 1$. That means the matrices $\mathbf{B}$ and $\mathbf{C}$ in MSE equation

(1.45) are $n \times 2n$ dimensional and the elements are

$$\mathbf{B} = \{B\}_{ij} = \begin{cases} 1, & j = 2i - 1 \\ 0, & otherwise \end{cases} \tag{1.46}$$

$$\mathbf{C} = \{C\}_{ij} = \begin{cases} 1, & j = 2i \\ 0, & otherwise \end{cases} \tag{1.47}$$

where $i = 1, \cdots, n$. Elements of penalty matrix $\{\Omega_{2n}\}_{jk}$ is given in appendices.

The solution to (1.45) is easily seen to be

$$\hat{\theta} = (\mathbf{B}^\top \mathbf{B} + \gamma \mathbf{C}^\top \mathbf{C} + n\lambda\Omega)^{-1}(\mathbf{B}^\top \mathbf{y} + \gamma \mathbf{C}^\top \mathbf{v}) \tag{1.48}$$

a generalized ridge regression. Then the fitted smoothing spline is given by

$$\hat{f}(t) = \sum_{i=1}^{2n} N_i(t)\hat{\theta}_i \tag{1.49}$$

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 (1.48) are a linear combination of 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

$$\begin{aligned} \hat{\mathbf{f}} &= \mathbf{B}(\mathbf{B}^\top \mathbf{B} + \gamma \mathbf{C}^\top \mathbf{C} + n\lambda\Omega)^{-1}(\mathbf{B}^\top \mathbf{y} + \gamma \mathbf{C}^\top \mathbf{v}) \\ &\triangleq \mathbf{S}_{\lambda,\gamma}\mathbf{y} + \gamma \mathbf{T}_{\lambda,\gamma}\mathbf{v} \end{aligned} \tag{1.50}$$

$$\begin{aligned} \hat{\mathbf{f}}' &= \mathbf{C}(\mathbf{B}^\top \mathbf{B} + \gamma \mathbf{C}^\top \mathbf{C} + n\lambda\Omega)^{-1}(\mathbf{B}^\top \mathbf{y} + \gamma \mathbf{C}^\top \mathbf{v}) \\ &\triangleq \mathbf{U}_{\lambda,\gamma}\mathbf{y} + \gamma \mathbf{V}_{\lambda,\gamma}\mathbf{v} \end{aligned} \tag{1.51}$$

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

$$\hat{\mathbf{f}} = \mathbf{B}_\xi(\mathbf{B}_\xi^\top \mathbf{B}_\xi)^{-1}\mathbf{B}_\xi \mathbf{y} = \mathbf{H}_\xi \mathbf{y} \tag{1.52}$$

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

A tractor spline on interval $[t_1, t_n]$ has $2n$ basis functions $N_1(t), \cdots, N_{2n}(t)$, which are linear independent. So the space $\mathbb{H}$, spanned by these basis functions, is a $2n$ dimensional space. Following the definition in section 1.3, it can be proved that the space $\mathbb{H}$ is a Reproducing Kernel Hilbert Space with inner product given in (1.19), and kernel function $R(s,t)$ defined in (1.20).

Noticing the definition of Hermite Spline from equation (1.39) to (1.42), prior status of y_i and v_i will only affect the status y_{i+1} and v_{i+1} in the following time period t_i. Define a covariance matrix Λ_i as

$$\Lambda_i = \text{cov}(\begin{bmatrix} y_i \\ v_i \end{bmatrix}, \begin{bmatrix} y_{i+1} \\ v_{i+1} \end{bmatrix}) \tag{1.53}$$

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

$$\begin{aligned} \text{cov}(f(s), f(t)) &= \tau^2 R(s,t) + \sigma_n^2 \Lambda \\ \text{cov}(f(s), f'(t)) &= \tau^2 R'(s,t) + \frac{\sigma_n^2}{\sqrt{\gamma}} \Lambda \\ \text{cov}(f'(s), f(t)) &= \tau^2 \dot{R}(s,t) + \frac{\sigma_n^2}{\sqrt{\gamma}} \Lambda \\ \text{cov}(f'(s), f'(t)) &= \tau^2 \dot{R}'(s,t) + \frac{\sigma_n^2}{\gamma} \Lambda \end{aligned} \tag{1.54}$$

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

$$\text{cov}(\mathbf{y}, \mathbf{v}, f, f') = \begin{bmatrix} \tau^2 Q + \sigma_n^2 \Lambda & \tau^2 O + \frac{\sigma_n^2}{\sqrt{\gamma}} \Lambda & \tau^2 \xi & \tau^2 \xi' \\ \tau^2 O + \frac{\sigma_n^2}{\sqrt{\gamma}} \Lambda & \tau^2 P + \frac{\sigma_n^2}{\gamma} \Lambda & \tau^2 \dot{\xi} & \tau^2 \dot{\xi}' \\ \tau^2 \xi^\top & \tau^2 \dot{\xi}^\top & \tau^2 R(t,t) & \tau^2 R'(t,t) \\ \tau^2 \xi'^\top & \tau^2 \dot{\xi}'^\top & \tau^2 \dot{R}(t,t) & \tau^2 \dot{R}'(t,t) \end{bmatrix} \tag{1.55}$$

where $\{Q\}_{ij}$, $\{O\}_{ij}$, $\{P\}_{ij}$, ξ and $\dot{\xi}$ are the same as that in (1.32). Then

$$\begin{aligned} E\begin{bmatrix} f & \mathbf{y} \\ f' & \mathbf{v} \end{bmatrix} &= \begin{bmatrix} \xi^\top & \dot{\xi}^\top \\ \xi'^\top & \dot{\xi}'^\top \end{bmatrix} \begin{bmatrix} Q + n\lambda\Lambda & O + \frac{n\lambda}{\sqrt{\gamma}}\Lambda \\ O + \frac{n\lambda}{\sqrt{\gamma}}\Lambda & P + \frac{n\lambda}{\gamma}\Lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ \gamma \mathbf{v} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \xi^\top & \dot{\xi}^\top \\ \xi'^\top & \dot{\xi}'^\top \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \gamma \mathbf{v} \end{bmatrix} \\ &= \begin{bmatrix} \xi^\top (A\mathbf{y} + B\gamma\mathbf{v}) + \dot{\xi}^\top (C\mathbf{y} + D\gamma\mathbf{v}) \\ \xi'^\top (A\mathbf{y} + B\gamma\mathbf{v}) + \dot{\xi}'^\top (C\mathbf{y} + D\gamma\mathbf{v}) \end{bmatrix} \end{aligned} \tag{1.56}$$

where $n\lambda = \sigma_n^2/\tau^2$. The posterior mean is of the form $\xi^\top \mathbf{c} + \dot{\xi}^\top \mathbf{d}$, and $E(f'|\mathbf{y}, \mathbf{v})$ is of the form $\xi'^\top \mathbf{c} + \dot{\xi}'^\top \mathbf{d}$, with the same coefficients given by

$$\mathbf{c} = A\mathbf{y} + B\gamma\mathbf{v} \tag{1.57}$$

$$\mathbf{d} = C\mathbf{y} + D\gamma\mathbf{v} \tag{1.58}$$

Following the procedure in section 1.3 and 1.4, we use observations $t_i, i = 1, \cdots, n$ to construct a subspace $\mathbb{A} \subset \mathbb{H}$, which is a linear combination of kernel functions $R(s,t)$, as

$$\mathbb{A} = \{f : f = \sum_{i=1}^{n} \alpha_i R(t_i, \cdot)\}. \tag{1.59}$$

The covariance matrix and posterior mean all given above, and the solution can be computed via finding $\mathbf{c}$ and $\mathbf{d}$.

In fact, for a tractor spline, the space $\mathbb{A}$ only contains terms of odd basis functions N_{2k-1}, which can be seen from matrix $\mathbf{B}$ in (1.45). So we construct another subspace

$$\mathbb{B} = \{f : f = \sum_{i=1}^{n} \beta_i \dot{R}(t_i, \cdot) = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{2n} N_j'(t_i) N_j(\cdot)\} \tag{1.60}$$

where $\dot{R}$ is defined in equation (1.22). This subspace contains even terms basis functions N_{2k}. So $\mathbb{A} \cap \mathbb{B} = \emptyset$.

Thus $f \in \mathbb{H}$ can be written as

$$f(t) = \sum_{i=1}^{n} c_i R(t_i, t) + \sum_{i=1}^{n} d_i \dot{R}(t_i, t) + \rho(t) \tag{1.61}$$

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

$$f'(t) = \sum_{i=1}^{n} c_i R'(t_i, t) + \sum_{i=1}^{n} d_i \dot{R}'(t_i, t) + \rho'(t). \tag{1.62}$$

The equation (1.6) can be written as

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{n} c_j R(t_j, t_i) - \sum_{j=1}^{n} d_j \dot{R}(t_j, t_i) - \rho(t_i))^2 \\ &+\frac{\gamma}{n}\sum_{i=1}^{n}(v_i - \sum_{j=1}^{n} c_j R'(t_j, t_i) - \sum_{j=1}^{n} d_j \dot{R}'(t_j, t_i) - \rho'(t_i))^2 \\ &+\lambda \int_{t_1}^{t_n} (\sum_{j=1}^{n} c_j R''(t_j, t) + \sum_{j=1}^{n} d_j \dot{R}''(t_j, t) + \rho''(t))^2 dt \end{aligned} \tag{1.63}$$

By orthogonality, $\rho(t_i) = \langle R(t_i,\cdot),\rho\rangle = 0$, and $\rho'(t_i) = \langle \dot{R}(t_i,\cdot),\rho\rangle = 0$, 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 (1.45) can be written as

$$(\mathbf{y}-Q\mathbf{c}-P\mathbf{d})^\top(\mathbf{y}-Q\mathbf{c}-P\mathbf{d})+\gamma(\mathbf{v}-\frac{\partial Q}{\partial t}\mathbf{c}-\frac{\partial P}{\partial t}\mathbf{d})^\top(\mathbf{v}-\frac{\partial Q}{\partial t}\mathbf{c}-\frac{\partial P}{\partial t}\mathbf{d})+n\lambda\Omega+\lambda(\rho,\rho). \tag{1.64}$$

The elements of penalty matrix Ω is in appendices. Note that ρ only appears in the third term in (1.64), 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 (1.64) with respect to $\mathbf{c}$ and $\mathbf{d}$.

1.6 Cross Validation

The coefficients can be calculated by minimizing MSE function. While another problem is how to choose smoothing parameter. There are two different philosophical approaches to the question of choosing the smoothing parameter. The first approach is to regard the free choice of smoothing parameter as an advantageous feature of the procedure. The other is a need for an automatic method whereby the smoothing parameter values is chose by the data, Green and Silverman (1993).

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 returning a small value of $(y-f(t))^2$.

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

$$\frac{1}{n}\sum_{j\neq i}(y_j-f(t_j))^2+\lambda\int f''^2 dt \tag{1.65}$$

and λ can be quantified by cross-validation score function

$$CV(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\{y_i-\hat{f}^{(-i)}(t_i,\lambda)\}^2. \tag{1.66}$$

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

An efficient way to calculate cross validation score is given by Green and Silverman (1993). Through the equation (1.52), we know that the value of the smoothing spline

$\hat{f}$ depend linearly on the data y_i. Define the matrix $A(\lambda)$, which is a map vector of observed values y_i to predicted values $\hat{f}(t_i)$. Then we have

$$\mathbf{f} = A(\lambda)\mathbf{y} \tag{1.67}$$

and the following lemma.

Lemma 2. *The cross validation score satisfies*

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}(t_i)}{1 - A_{ii}(\lambda)}\right)^2 \tag{1.68}$$

where $\hat{f}$ is the spline smoother calculated from the full data set $\{(t_i, y_i)\}$ with smoothing paramter λ.

For a tractor spline and its MSE function, there are two parameters need to be estimated λ and γ. Thus the objective function becomes

$$\frac{1}{n}\sum_{j\neq i}(y_j - f(t_j))^2 + \frac{\gamma}{n}\sum_{j\neq i}(v_j - f'(t_j))^2 + \lambda\int f''^2 dt, \tag{1.69}$$

and the cross-validation score function is

$$CV(\lambda, \gamma) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{f}^{(-i)}(t_i, \lambda, \gamma)\}^2. \tag{1.70}$$

For a tractor spline, the parameter $\hat{\theta} = (B^\top B + \gamma C^\top C + n\Omega_\lambda)^{-1}(B^\top\mathbf{y} + \gamma C^\top\mathbf{v})$, then

$$\begin{aligned}\hat{f} = B\hat{\theta} &= B(B^\top B + \gamma C^\top C + n\Omega_\lambda)^{-1}B^\top\mathbf{y} + B(B^\top B + \gamma C^\top C + n\Omega_\lambda)^{-1}C^\top\mathbf{v}\\ &= S\mathbf{y} + \gamma T\mathbf{v},\end{aligned} \tag{1.71}$$

$$\begin{aligned}\hat{f}' = C\hat{\theta} &= C(B^\top B + \gamma C^\top C + n\Omega_\lambda)^{-1}B^\top\mathbf{y} + C(B^\top B + \gamma C^\top C + n\Omega_\lambda)^{-1}C^\top\mathbf{v}\\ &= U\mathbf{y} + \gamma V\mathbf{v}.\end{aligned} \tag{1.72}$$

Then

Theorem 2. *The cross validation score of a tractor spline satisfies*

$$CV(\lambda, \gamma) = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{f}(t_i) - y_i + \gamma\frac{T_{ii}}{1-\gamma V_{ii}}(\hat{f}'(t_i) - v_i)}{1 - S_{ii} - \gamma\frac{T_{ii}}{1-\gamma V_{ii}}U_{ii}} \tag{1.73}$$

where $\hat{f}$ is the tractor spline smoother calculated from the full data set $\{(t_i, y_i, v_i)\}$ with smoothing paramter λ and γ.

The proof of Theorem 2 follows immediately from a lemma, and gives an expression for the deleted residuals $y_i - \hat{f}^{(-i)}(t_i)$ and $v_i - \hat{f}'^{(-i)}(t_i)$ in terms of $y_i - \hat{f}(t_i)$ and $v_i - \hat{f}'(t_i)$ respectively.

Lemma 3. *For fixed λ, γ and i, denote $\mathbf{f}^{(-i)}$ by the vector with components $f_j^{(-i)} = \hat{f}^{(-i)}(t_j, \lambda, \gamma)$, $\mathbf{f}'^{(-i)}$ by the vector with components $f_j'^{(-i)} = \hat{f}'^{(-i)}(t_j, \lambda, \gamma)$, and define vectors $\mathbf{y}^*$ and $\mathbf{v}^*$ by*

$$\begin{cases} y_j^* = y_j & j \neq i \\ y_i^* = \hat{f}^{(-i)}(t_i) & otherwise \end{cases}, \tag{1.74}$$

$$\begin{cases} v_j^* = v_j & j \neq i \\ v_i^* = \hat{f}'^{(-i)}(t_i) & otherwise \end{cases}. \tag{1.75}$$

Then

$$\hat{\mathbf{f}}^{(-i)} = S\mathbf{y}^* + \gamma T \mathbf{v}^* \tag{1.76}$$

$$\hat{\mathbf{f}}'^{(-i)} = U\mathbf{y}^* + \gamma V \mathbf{v}^* \tag{1.77}$$

1.6.1 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:

Group 1 : $t_2, t_{2+k}, \cdots$

Group 2 : $t_3, t_{3+k}, \cdots$

$\vdots$

Group k : $t_{k+1}, t_{2k+1}, \cdots$

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 (1.45)

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

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

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

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

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

$$\Omega^{(k)}_{2k,2k} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} dt = \frac{4}{\Delta_k}$$

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

$$\Omega^{(k)}_{2k,2k+2} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k)}_{10}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{2}{\Delta_k}$$

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

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

$$\Omega^{(k)}_{2k+2,2k+2} = \int_{t_k}^{t_{k+1}} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} \frac{d^2 h^{(k+1)}_{11}(t)}{dt^2} dt = \frac{4}{\Delta_k}$$

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

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

Penalty Matrix in (1.64)

The penalty matrix Ω in (1.64) is a combination of three sub matrix Ω_1, Ω_2 and Ω_3, which are in the following form

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

Proof of Theorem 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

$$D(c_1, c_2, \cdots, c_{2n}) = \begin{vmatrix} N_1(c_1) & N_1(c_2) & \cdots & N_1(c_{2n}) \\ N_2(c_1) & N_2(c_2) & \cdots & N_2(c_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ N_{2n}(c_1) & N_{2n}(c_2) & \cdots & N_{2n}(c_{2n}) \end{vmatrix} = \begin{vmatrix} 1 & a_{12} & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_{22} & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_{32} & 1 & a_{34} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a_{2n-1,2n-1} & 1 \\ 0 & 0 & 0 & 0 & \cdots & a_{2n,2n-1} & 0 \end{vmatrix}, \quad (78)$$

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

and $a_{ij} \neq 0$. After decomposing determinant D in equation (78), 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

$$g(t) = d_k(t-t_k)^3 + c_k(t-t_k)^2 + b_k(t-t_k) + a_k, \text{for } t_k \le t \le t_{k+1} \tag{79}$$

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. $\square$

Proof of Lemma 3

Proof. For any spline f,

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}(y_j^* - f(t_j))^2 + \frac{\gamma}{n}\sum_{j=1}^{n}(v_j^* - f'(t_j)) + \lambda\int f''^2 \\
\geq&\frac{1}{n}\sum_{j\neq i}^{n}(y_j^* - f(t_j))^2 + \frac{\gamma}{n}\sum_{j\neq i}^{n}(v_j^* - f'(t_j)) + \lambda\int f''^2 \\
\geq&\frac{1}{n}\sum_{j\neq i}^{n}(y_j^* - \hat{f}^{(-i)}(t_j))^2 + \frac{\gamma}{n}\sum_{j\neq i}^{n}(v_j^* - \hat{f}'^{(-i)}(t_j)) + \lambda\int \hat{f}^{(-i)''2} \\
=&\frac{1}{n}\sum_{j=1}^{n}(y_j^* - \hat{f}^{(-i)}(t_j))^2 + \frac{\gamma}{n}\sum_{j=1}^{n}(v_j^* - \hat{f}'^{(-i)}(t_j)) + \lambda\int \hat{f}^{(-i)''2}
\end{aligned} \tag{80}$$

by the definition of $\hat{f}^{(-i)}$, $\hat{f}'^{(-i)}$ and $y_i^* = \hat{f}^{(-i)}(t_i)$, $v_i^* = \hat{f}'^{(-i)}(t_i)$. It follows that $\hat{f}^{(-i)}$ is the minimizer of the MSE function (1.6), so that

$$\hat{\mathbf{f}}^{(-i)} = S\mathbf{y}^* + \gamma T\mathbf{v}^* \tag{81}$$

$$\hat{\mathbf{f}}'^{(-i)} = U\mathbf{y}^* + \gamma V\mathbf{v}^* \tag{82}$$

$\square$

Proof of Theorem 2

Proof.

$$\begin{aligned}
\hat{f}^{(-i)}(t_i) - y_i =&\sum_{j=1}^{n} S_{ij}y_j^* + \gamma\sum_{j=1}^{n} T_{ij}v_j^* - y_i^* \\
=&\sum_{j\neq i}^{n} S_{ij}y_j + \gamma\sum_{j\neq i}^{n} T_{ij}v_j + S_{ii}\hat{f}^{(-i)}(t_i) + \gamma T_{ii}\hat{f}'^{(-i)}(t_i) - y_i \\
=&\sum_{j=1}^{n} S_{ij}y_j + \gamma\sum_{j=1}^{n} T_{ij}v_j + S_{ii}(\hat{f}^{(-i)}(t_i) - y_i) + \gamma T_{ii}(\hat{f}'^{(-i)}(t_i) - v_i) - y_i \\
=&(\hat{f}(t_i) - y_i) + S_{ii}(\hat{f}^{(-i)}(t_i) - y_i) + \gamma T_{ii}(\hat{f}'^{(-i)}(t_i) - v_i).
\end{aligned} \tag{83}$$

Additionally,

$$\begin{aligned}
\hat{f}'^{(-i)}(t_i) - v_i =& \sum_{j=1}^{n} U_{ij} y_j^* + \gamma \sum_{j=1}^{n} V_{ij} v_j^* - v_i^* \\
=& \sum_{j \neq i}^{n} U_{ij} y_j + \gamma \sum_{j \neq i}^{n} V_{ij} v_j + U_{ii} \hat{f}^{(-i)}(t_i) + \gamma V_{ii} \hat{f}'^{(-i)}(t_i) - v_i \\
=& \sum_{j=1}^{n} U_{ij} y_j + \gamma \sum_{j=1}^{n} V_{ij} v_j + U_{ii} (\hat{f}^{(-i)}(t_i) - y_i) + \gamma V_{ii} (\hat{f}'^{(-i)}(t_i) - v_i) - v_i \\
=& (\hat{f}'(t_i) - v_i) + U_{ii} (\hat{f}^{(-i)}(t_i) - y_i) + \gamma V_{ii} (\hat{f}'^{(-i)}(t_i) - v_i).
\end{aligned} \tag{84}$$

Then

$$\hat{f}'^{(-i)}(t_i) - v_i = \frac{\hat{f}'(t_i) - v_i}{1 - \gamma V_{ii}} + \frac{U_{ii}(\hat{f}^{(-i)}(t_i) - y_i)}{1 - \gamma V_{ii}}. \tag{85}$$

After substituting equation (85) into (83), we get

$$\hat{f}^{(-i)}(t_i) - y_i = \frac{\hat{f}(t_i) - y_i + \gamma \frac{T_{ii}}{1-\gamma V_{ii}} (\hat{f}'(t_i) - v_i)}{1 - S_{ii} - \gamma \frac{T_{ii}}{1-\gamma V_{ii}} U_{ii}}. \tag{86}$$

Then

$$CV(\lambda, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{f}(t_i) - y_i + \gamma \frac{T_{ii}}{1-\gamma V_{ii}} (\hat{f}'(t_i) - v_i)}{1 - S_{ii} - \gamma \frac{T_{ii}}{1-\gamma V_{ii}} U_{ii}}. \tag{87}$$

□