NON-LINEAR FILTERING APPROACH TO AN ADJUSTMENT OF NON-UNIFORM SAMPLING LOCATIONS IN SPATIAL DATASETS

Masako Kamiyama

Railway Technical Research Inst. &
The Graduate Univ. for Advanced Studies
Tokyo, Japan
Email: kamiyama@rtri.or.jp

Tomoyuki Higuchi

The Institute of Statistical Mathematics
Department of Prediction and Control
Tokyo, Japan
Email: higuchi@ism.ac.jp

ABSTRACT

A procedure is described for adjusting sampling locations in one spatially discretized dataset to those in another when the value differences between these sets are mainly caused by the sampling intervals that locally lengthen and shorten. This adjustment is formulated into an optimization form that can be solved by dynamic programming. Unknown parameters involved in the form can be identified using the maximum likelihood procedure that employs non-linear filtering for a generalized state-space model. This procedure is based on the fact that the optimal solution in dynamic programming is equivalent to the "Maximum A Posteriori (MAP) estimation" in a Bayesian framework.

1. INTRODUCTION

To manage the rail geometry of a railway track, a special rolling stock called the "track inspection car" periodically measures the rail geometry because the rail geometry varies slightly under the load of passing trains. These geometry variations must be managed to keep trains running safely. While running on the rails, the track inspection car continuously measures various aspects of rail geometry. These geometry measurements are simultaneously discretized at fixed spatial intervals, and are recorded in digital datasets.

Although it is desirable that these locations be fixed in order to observe variations in the rail geometry, the set of the discretized locations on the rail changes slightly with each measurement, as illustrated in Fig. 1. These changes caused by the pulse used for selecting the sampling locations (called a "wheel-rotation pulse") being linked to the rotation of the car wheel as illustrated in Fig. 2. Thus, identical spatial discretization cannot be reproduced. Moreover, it is difficult to adjust these location gaps after the discretization, since some sampling intervals shorten or lengthen locally due to slipping or sliding of the car wheel (illustrated in Fig. 3). Unfortunately, the length and location of these locally irregular intervals cannot be detected.

An example of these irregular sampling-intervals is shown in Fig. 4. The upper two curves show the values of the track gauges measured in the same railway section and demonstrate similar behavior since track gauge does not usually vary under the load of passing trains. The difference sequences between them are also shown below. Note that a difference operation is performed on the different intervals of (A). The amplitude of both difference-sets changes abnormally in the second half although the two gauge waveforms appear similar. Hence, the wheel appears to slip or slide during one or the other of the measuring runs.

Thus, we have developed a procedure for adjusting the difference between sampling locations in two datasets. Figure 4 shows an example of adjustment of the locations with our proposed procedure. The abnormal amplitude, which is seen in the difference datasets mentioned above, is suppressed. Our procedure reveals that the wheel slid about 1 meter when the training dataset was measured.

Our procedure is based on the fact that the optimal solution of dynamic programming is equivalent to a Maximum A Posteriori(MAP) estimation in a Bayesian framework [1].

The procedure is outlined as follows:

1. Select a supervised dataset and a training dataset that satisfy the several criteria for adjustment purposes.
2. Model a mechanism to yield the non-uniform sampling, i.e., the wheel rotation including slip and slide.
3. Formulate this model in an optimization problem that can be solved by dynamic programming, which is a general method for solving non-linear discrete optimization problems. However, this form contains unknown parameters.
4. Represent this non-linear optimization problem with a generalized state-space representation and identify the unknown parameters (called hyperparameters in a Bayesian framework) using a non-linear filtering algorithm based on the maximum likelihood method.

0-7803-7997-7/03/$17.00 © 2003 IEEE

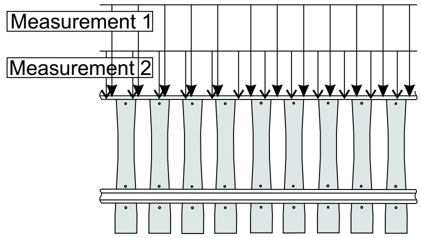

Fig. 1. Scheme for discretized-location gaps in a railway track with two measurements

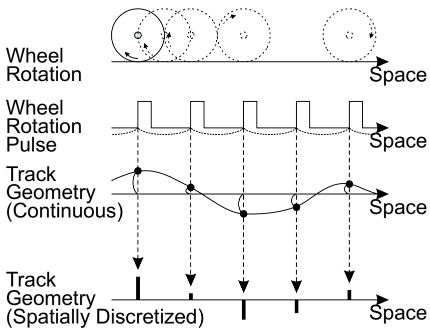

Fig. 2. Scheme for spatially discretized geometry of a railway track with wheel-rotation pluses

5. Adjust the sampling location differences by dynamic programming with the identified parameters.

2. FORMULATING ADJUSTMENT INTO AN OPTIMIZATION PROBLEM AND ITS INTERPRETATION

2.1. Transformation of an adjustment problem

We adjust the non-uniform sampling locations according to the following procedure:

1. Divide the original sampling interval of the training set by positive integer α, and approximate the values for the newly interpolated data (interpolation).

2. Select data points corresponding to each supervised data from the interpolated training set.

These steps are illustrated in Fig. 5.

In this procedure, the adjustment is transformed into the selection of data points from the interpolated dataset. Let n_t be the data point index of the training set, and select it corresponding to the t-th supervised data. This paper expresses $\{n_1, n_2, \cdots, n_{T-1}, n_T\}$ as $\mathbf{n}_{1:T}$. The other variables are similarly expressed. Thus, the adjustment is transformed

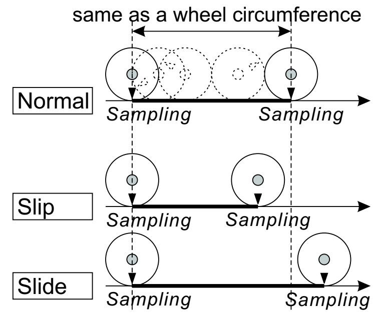

Fig. 3. Scheme for variations in spatial-sampling intervals with wheel-rotation states

into construction of optimal sequence $\mathbf{n}_{1:T}$, where T is the number of the supervised data.

The interpolated training dataset is called the "training dataset" in the rest of this paper.

2.2. Formulation into an optimization problem

Sequence $\mathbf{n}_{1:T}$ is the optimal solution that minimizes the following constrained non-linear target function:

$$\begin{aligned}
\mathbf{n}_{1:T} &= \arg\min \sum_{t=1}^{T} F(n'_t) \\
&= \arg\min_{n'_t} \Bigg[\sum_{t=1}^{T} (y_t - z_{n'_t})^2 \\
&+ \mu_1 \sum_{t=2}^{T} \xi(\Delta n'_t - \alpha, \alpha - 1) \\
&+ \mu_2 \sum_{t=3}^{T} \xi(\Delta^2 n'_t, 2\alpha - 2) \Bigg],
\end{aligned} \tag{1}$$

subject to $1 \le n'_1 \le 2\alpha - 1$, and $N - (2\alpha - 1) + 1 \le n'_T \le T$, where

· y_t is the value of the t-th supervised dataset.
· T is the number of the supervised data.
· $N \approx \alpha T$ is the number of the training data.
· $z_{n'_t}$ is the value of the n'_t-th training data.
· $\mu_1 \ge 0$ is the penalty for $\Delta n_t \stackrel{\text{def}}{=} n_t - n_{t-1} \ne \alpha$.
· $\mu_2 \ge 0$ is the penalty for $\Delta^2 n_t \stackrel{\text{def}}{=} n_t - 2n_{t-1} + n_{t-2} \ne 0$.
· ξ is the function defined by

$$\xi(n, \gamma) = \begin{cases} 0, & n = 0, \\ 1, & 0 < |n| \le \gamma, \\ \infty, & \gamma < |n|, \end{cases} .$$

To simplify this notation, we replace n'_t and n'_{t-1} with vector $\mathbf{x}'_t \stackrel{\text{def}}{=} [n'_t \; n'_{t-1}]^T$ (where a "^{T}" represents transpose) with $\mathbf{x}'_t \in \mathbf{S}_t \stackrel{\text{def}}{=} \{[n'_t \; n'_{t-1}]^T | n'_t \in S_t, n'_{t-1} \in$

$S_{t-1}\}$, where S_t is the search area given for n'_t. By substituting these equations for $F(n'_t)$, the value of $g(\mathbf{x}'_t) \stackrel{\text{def}}{=} \min \sum_{j=1}^{t} F(n'_j)$ depends on $g(\mathbf{x}'_{t-1})$, $\mathbf{x}'_t$, and $\mathbf{x}'_{t-1}$ alone, if the values of μ_1, μ_2, and α are all given. Therefore, this problem observes the principle of optimality, which underlies the "dynamic programming" technique. This techinique is effective for optimizing this type of problem.

However, the obtained solution depends on the value of weighting coefficients μ_1 and μ_2. We should carefully select these parameters. For this selection, consider an interpretation of solutions $\mathbf{n}_{1:T}$ and parameters in Section 2.3.

2.3. Dynamic programming and MAP estimation

Assume that $(y_t - z_{n_t})$ is a normal random variable with a zero mean and variance σ^2. By multiplying $\sum F(n'_t)$ in (1) by $-1/(2\sigma^2)$ and exponentiating it, we obtain

$$\begin{aligned}\mathbf{n}_{1:T} \quad = \quad & \arg\max_{n'_t} \left\{ \prod_{t=1}^{T} \exp\left[-\frac{1}{2\sigma^2}(y_t - z_{n'_t})^2 \right] \right\} \\ & \exp\left\{ -\frac{1}{2\sigma^2}\left[\mu_1 \sum_{t=2}^{T} \xi(\Delta n'_t - \alpha, \alpha - 1) \right.\right. \\ & \left.\left. +\mu_2 \sum_{t=3}^{T} \xi(\Delta n'_t, 2\alpha - 2) \right] \right\}. \qquad (2)\end{aligned}$$

Next, by normalizing the first term on the right side of (2), we find the function that can be interpreted as the conditional density $p(\mathbf{y}_{1:T}|\mathbf{n}_{1:T})$ assuming normality. Similarly, by using normalization factor, the second term of (2) can be interpreted as probability distribution $p(\mathbf{n}_{1:T})$. In this way, (2) yields the following interpretation within a Bayesian framework: $p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T}|\mathbf{n}_{1:T})p(\mathbf{n}_{1:T})$, where $p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T})$ is the posterior distribution of $\mathbf{n}_{1:T}$, $p(\mathbf{y}_{1:T}|\mathbf{n}_{1:T})$ is the data distribution conditional on $\mathbf{n}_{1:T}$, and $p(\mathbf{n}_{1:T})$ is the prior distribution.

Hence, minimizing $\sum F(n'_t)$ by dynamic programming is equivalent to constructing a sequence that maximizes the posterior distribution of $\mathbf{n}_{1:T}$ with a given $\mathbf{y}_{1:T}$; $\mathbf{n}_{1:T}$ is then called the MAP estimate.

3. IDENTIFYING THE HYPERPARAMETERS INCLUDED IN AN OPTIMIZATION FORM

3.1. Method of computing log-likelihood

According to the above interpretation, the log-likelihood of a model specified by μ_1, μ_2, and σ is obtained by

$$\begin{aligned}LL(\mu_1, \mu_2, \sigma) \quad & \stackrel{\text{def}}{=} \quad \log p(\mathbf{y}_{1:T}|\mu_1, \mu_2, \sigma) \\ & = \quad \sum_{t=1}^{T} \log p(y_t|\mathbf{y}_{1:t-1}, \mu_1, \mu_2, \sigma). \quad (3)\end{aligned}$$

Hence, the values of the hyperparameters (μ_1, μ_2, σ) can be evaluated using the value of LL [2]. Note that the concept of "hyperparameter" in a Bayesian framework is the same as that of "parameter" in a state-space model [3]. The most suitable hyperparameters $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma})$ are identified by maximizing LL, which is obtained as the by-product of the computation for $p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T})$.

3.2. Transformation into a generalized state-space representation

The aforementioned Bayesian interpretations can be transformed into the following generalized state-space representation:

(System model)

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) = \begin{bmatrix} n_{t-1} + \alpha + v_t(n_{t-1}, n_{t-2}) \\ n_{t-1} \end{bmatrix},$$

(Observation model)

$$y_t = h(\mathbf{x}_t) + w_t = z_{n_t} + w_t,$$

where $\mathbf{x}_t \stackrel{\text{def}}{=} [n_t \; n_{t-1}]^T$, $v_t \sim q(\cdot|\mathbf{x}_{t-1}, \mu_1, \mu_2, \sigma)$, $h(\mathbf{x}_t) \stackrel{\text{def}}{=} z_{n_t}$, and $w_t \sim N(0, \sigma^2)$. Distribution $q(\cdot)$ is defined as follows:

· **If** $n_{t-1} - n_{t-2} = \alpha$ **, then**

$$q(v_t|\mathbf{x}_{t-1}, \mu_1, \mu_2, \sigma) \stackrel{\text{def}}{=} \begin{cases} 1/\beta_1, \\ \qquad v_t = 0, \\ \exp\left(-\frac{\mu_1+\mu_2}{2\sigma^2} \right)/\beta_1, \\ \qquad 1 \le |v_t| \le \alpha - 1, \end{cases}$$

where $\beta_1 = 1 + 2(\alpha - 1)\exp\left(-\frac{\mu_1+\mu_2}{2\sigma^2} \right)$ for $\sum q(\cdot) = 1$.

· **If** $n_{t-1} - n_{t-2} \neq \alpha$**, then**

$$q(v_t|\mathbf{x}_{t-1}, \mu_1, \mu_2, \sigma) \stackrel{\text{def}}{=} \begin{cases} \exp\left(-\frac{\mu_2}{2\sigma^2} \right)/\beta_2, \\ \qquad v_t = 0, \\ \exp\left(-\frac{\mu_1}{2\sigma^2} \right)/\beta_2, \\ \qquad v_t = n_{t-1} - n_{t-2} - \alpha, \\ \exp\left(-\frac{\mu_1+\mu_2}{2\sigma^2} \right)/\beta_2, \\ \qquad \text{other than the above } v_t, \\ \qquad \text{and } 1 \le |v_t| \le \alpha - 1, \end{cases}$$

where $\beta_2 = \exp\left(-\frac{\mu_1}{2\sigma^2} \right) + \exp\left(-\frac{\mu_2}{2\sigma^2} \right) + [2(\alpha - 1) - 1]\exp\left(-\frac{\mu_1+\mu_2}{2\sigma^2} \right)$ for $\sum q(\cdot) = 1$.

3.3. Identification of hyperparameters with a maximum likelihood function

The generalized state-space representation has the advantage of computation of aforementioned likelihood $p(\mathbf{y}_{1:T}|\mu_1, \mu_2, \sigma)$ since the following recursive relations between state distributions are available [3].

(Prediction)

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \quad (4)$$

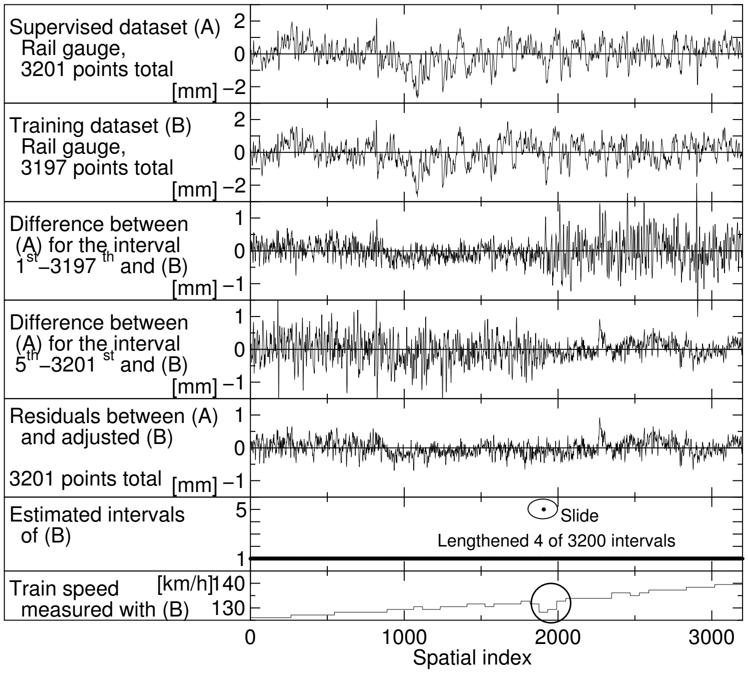

Fig. 4. Example of measured datasets from two measurements in the same railway section, their differences, and residuals after location adjustment

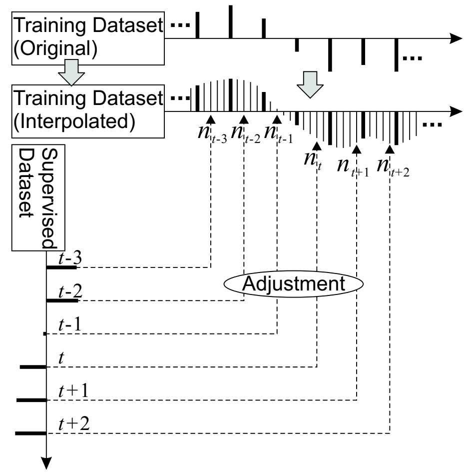

Fig. 5. Scheme for the basic idea for adjusting the discretized-location gaps

(Filtering)

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(y_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(y_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} = \frac{p(y_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(y_t|\mathbf{y}_{1:t-1})}. \quad (5)$$

Note that $p(y_t|\mathbf{y}_{1:t-1})$ appears as the denominator in (5). Thus, log-likelihood LL in (3) is obtained after these recursive computations. In addition, $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ can be computed arithmetically without complicated numerical integration such as [3]. This is because the values of $\mathbf{x}_t$ are discrete. We maximize LL using a grid search over a hyperparameter space.

4. APPLICATION

We adjusted the sampling locations of (B) to those of (A) in Fig. 4 where $\alpha = 5$. Datasets (A) and (B) are actually measured by a track inspection car in the same railway section. Table 1 is a summary of the estimation results (LL_{max} in Table 1 means the maximum $LL(\mu_1, \mu_2, \sigma)$). The most likely adjustment is shown in Fig. 4. In addition, these results also predict that the wheel has slid over a distance equivalent to 4 sampling intervals on measuring run (B).

In addition, Fig. 4 also displays the train speed obtained with (B); this speed data usually cannot be referred to. The fact that the speed decreased with no other apparent reason to around $t = 1900$ indicates that the wheel probably slid.

Table 1. Most likely estimated parameters with (A) and (B) in Fig. 4

LL_{max}	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\sigma}$
608	0.42	0.43	0.176

5. REFERENCES

[1] S. Godsill, A. Doucet and M. West, “Maximum a posteriori sequence estimation using Monte Carlo particle filters,” Ann. Inst. Statist. Math., Vol. 53, pp. 82-96, Mar. 2001.

[2] H. Akaike, “Likelihood and the Bayes procedure,” Bayesian Statistics, J. M. Bernardo, M. H. DeGroot, D.V. Lindley and A. F. M. Smith, eds., Valencia: University Press, 1980, pp. 143-166.

[3] G. Kitagawa, “Non-Gaussian state-space modeling of nonstationary time series,” Journal of the American Statistical Association, Vol. 82, pp. 1032-1063, Dec. 1987.