Ann. Inst. Statist. Math.
Vol. 45, No. 1, 21–34 (1993)

# ON USE OF THE KALMAN FILTER FOR SPATIAL SMOOTHING

NOBUHISA KASHIWAGI

*The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan*

(Received August 1, 1991; revised April 28, 1992)

**Abstract.** A state-space model to perform discrete thin plate smoothing for data on a two-dimensional rectangular lattice is proposed with the use of the Kalman filter. The use of the Kalman filter reduces computational difficulties in the maximum likelihood estimation of a smoothing parameter. A procedure to reduce computational difficulties in the estimation of trend is given also. Numerical illustration is provided using two sets of artificial data.

*Key words and phrases*: Discrete thin plate smoothing, image analysis, Kalman filter, likelihood, spatial statistics, state-space approach.

## 1. Introduction

Let $y_{st}$ be an observation at the $(s,t)$-th site on a two-dimensional rectangular lattice, where $s,t = 1,\ldots,n$. We consider to estimate the trend $\boldsymbol{x} = (x_{11},\ldots,x_{n1},x_{12},\ldots,x_{nn})'$ from the data $\boldsymbol{y} = (y_{11},\ldots,y_{n1},y_{12},\ldots,y_{nn})'$ by the discrete thin plate smoothing method (DTPSM). In the DTPSM, the trend is provided by minimizing the following cost with respect to $x_{st}$'s.

$$
\begin{aligned}
(1.1)\qquad &\sum_{s=1}^{n}\sum_{t=1}^{n}(y_{st}-x_{st})^2 \\
&\quad + \lambda\sum_{s=2}^{n-1}\sum_{t=2}^{n-1}(4x_{st}-x_{s-1,t}-x_{s+1,t}-x_{s,t-1}-x_{s,t+1})^2 \\
&\quad + b(x_{st}\text{'s})
\end{aligned}
$$

where $\lambda$ ($\geq 0$) is a smoothing parameter and $b(x_{st}\text{'s})$ is a function to impose additional constraints on $x_{st}$'s. The name of the method, DTPSM, is due to the fact that the form of the difference constraint in the second term of (1.1) is derived by discretizing the potential energy function in thin plate bending. Difference constraints have been employed for smoothing by many authors, especially in the analysis of time-series data (e.g., Whittaker (1923), Shiller (1973), Kitagawa (1987), Kashiwagi and Yanagimoto (1992)).

In the DTPSM, when $\lambda$ is known, our task is only to minimize the cost. However, $\lambda$ is generally unknown, and the estimated trend is greatly influenced by $\lambda$. The selection of $\lambda$ is an essential problem in the DTPSM. One solution to the problem is to use likelihood. The likelihood function for selecting $\lambda$ is derived by assuming a statistical model which corresponds to the cost function. To explain this, we specify $b(x_{st}\text{'s})$ as

$$b(x_{st}\text{'s}|\lambda) = \lambda \sum_{(s,t)\in\partial_B} (4x_{st} - x_{s-1,t} - x_{s+1,t} - x_{s,t-1} - x_{s,t+1})^2$$

where $\partial_B = \{(s,t)|1 \le s \le n,\ t = 1, n \text{ or } s = 1, n,\ 1 \le t \le n\}$ and it is assumed that the first difference along the external normal at a point on the boundary is equal to zero, that is, $x_{0t} = x_{1t}$, $x_{n+1,t} = x_{nt}$, $x_{s0} = x_{s1}$ and $x_{s,n+1} = x_{sn}$. In this case, the cost function can be written as

$$(1.2) \quad \sum_{s=1}^{n}\sum_{t=1}^{n}(y_{st} - x_{st})^2 + \lambda \sum_{s=1}^{n}\sum_{t=1}^{n}(4x_{st} - x_{s-1,t} - x_{s+1,t} - x_{s,t-1} - x_{s,t+1})^2.$$

The corresponding statistical model is given in the density form as

$$p(y|\boldsymbol{x}, \sigma^2) \propto \sigma^{-n^2} \exp\left\{-\frac{1}{2\sigma^2}(y - x)'(y - x)\right\},$$
$$p(\boldsymbol{x}|\lambda, \sigma^2) \propto \sigma^{-(n^2-1)}\lambda^{(n^2-1)/2} \exp\left\{-\frac{\lambda}{2\sigma^2}(Dx)'Dx\right\},$$

where $D$ is the $n^2 \times n^2$ matrix of rank $n^2 - 1$ such that the components of $D\boldsymbol{x}$ are the linear combinations appearing in the second term of (1.2). Actually, the cost function (1.2) can be derived from this model as the function to obtain the posterior mean of $\boldsymbol{x}$. The posterior density of $\boldsymbol{x}$ is given by

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{y}, \lambda, \sigma^2) &\propto p(\boldsymbol{y}|\boldsymbol{x}, \sigma^2)p(\boldsymbol{x}|\lambda, \sigma^2) \\ &\propto \sigma^{-(2n^2-1)}\lambda^{(n^2-1)/2} \\ &\quad \times \exp\left[-\frac{1}{2\sigma^2}\{(\boldsymbol{y} - \boldsymbol{x})'(\boldsymbol{y} - \boldsymbol{x}) + \lambda(D\boldsymbol{x})'D\boldsymbol{x}\}\right]. \end{aligned}$$

The likelihood function for selecting $\lambda$ is derived by using Akaike's integral form (Akaike (1980)) as

$$\begin{aligned} L_B(\lambda, \sigma^2|\boldsymbol{y}) &= \int_{R(x)} p(\boldsymbol{y}|\boldsymbol{x}, \sigma^2)p(\boldsymbol{x}|\lambda, \sigma^2)d\boldsymbol{x} \\ &\propto \sigma^{-(n^2-1)}\lambda^{(n^2-1)/2} \mid I_{n^2} + \lambda D'D \mid^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2\sigma^2}\{(\boldsymbol{y} - \hat{\boldsymbol{x}})'(\boldsymbol{y} - \hat{\boldsymbol{x}}) + \lambda(D\hat{\boldsymbol{x}})'D\hat{\boldsymbol{x}}\}\right] \end{aligned}$$

where $R(x)$ is the support of $\boldsymbol{x}$, $I_{n^2}$ is the $n^2 \times n^2$ identity matrix and $\hat{\boldsymbol{x}}$ is the posterior mean of $\boldsymbol{x}$ given $\lambda$. The maximum likelihood estimate of $\sigma^2$ can be written as $\hat{\sigma}^2(\lambda) = (n^2-1)^{-1}\{(\boldsymbol{y}-\hat{\boldsymbol{x}})'(\boldsymbol{y}-\hat{\boldsymbol{x}}) + \lambda(D\hat{\boldsymbol{x}})'D\hat{\boldsymbol{x}}\}$. Therefore, the maximum likelihood estimate of $\lambda$ can be obtained by maximizing $L_B(\lambda, \hat{\sigma}^2(\lambda)|\boldsymbol{y})$.

However, the estimation of $\lambda$ by $L_B(\lambda, \hat{\sigma}^2(\lambda)|\boldsymbol{y})$ is not always feasible because of the computational difficulties. For example, if we take $n = 1000$, the order of $E$ $(= I_{n^2} + \lambda D'D)$ becomes $10^6$ (in image engineering, $n = 1000$ is frequently employed, as recent ordinary graphic terminals have about $1000 \times 1000$ resolution). This suggests that the number of elements in $E$ is equal to $10^{12}$. Fortunately, $E$ is a band matrix, and therefore, the number of the elements which need to be stored is reduced to *const.* $\times\, 10^9$ $(=$ *const.* $\times\, n^3)$. However, even this size exceeds the limit of what can be solved at a reasonable cost.

The purpose of this paper is to propose a state-space model to perform discrete thin plate smoothing, and to show that, if we use the Kalman filter (Kalman (1960) and Schweppe (1965)), the maximum likelihood estimation procedure for $\lambda$ can be constructed with the computer memory size of $O(n^2)$. The reduction of the required memory size is also possible by using an appropriate iterative method to calculate $\hat{\boldsymbol{x}}$, by using an appropriate orthogonalization method to calculate $|E|$ and by considering a computer program. However, the number of operations in this method is greater than that in the Kalman filter.

Wahba (1979) discussed thin plate smoothing by using splines and suggested to use Generalized Cross-Validation (GCV) instead of likelihood to select the smoothing parameter. However, the likelihood approach enables us to solve various problems including non-Gaussian smoothing and even a test problem by natural extension of the method. In addition, when the size of data is large, the computational difficulties also arise in the calculation of GCV.

In Section 2, a state-space model for the DTPSM is proposed. In Section 3, the maximum likelihood estimation procedure for $\lambda$ is presented, and it is shown that the required computer memory size is given by $O(n^2)$. A procedure to estimate the trend is also presented. Finally, numerical examples are shown in Section 4.

## 2. Model

In this section, we propose a state-space model for the DTPSM, which is derived by regarding only one suffix among $(s,t)$ as a time-suffix. Here, $t$ is assumed as a time-suffix, and the additional trend components $x_{s,-1}$ and $x_{s,0}$ $(s = 1,\ldots,n)$ are introduced to use the Kalman filter. This derives an irregular $b(x_{st}\text{'s})$, which is shown in the last part of this section.

Let $\boldsymbol{y}_t = (y_{1t},\ldots,y_{nt})'$, $\boldsymbol{x}_t = (x_{1t},\ldots,x_{nt})'$ and $\boldsymbol{\mu}_t = (\boldsymbol{x}_t', \boldsymbol{x}_{t-1}')'$. The proposed model is

$$
\begin{aligned}
&\boldsymbol{y}_t = F\boldsymbol{\mu}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim N(\boldsymbol{0}_n, \sigma^2 I_n),\\
&\boldsymbol{\mu}_t = G\boldsymbol{\mu}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N\left(\boldsymbol{0}_{2n}, \frac{\sigma^2}{\lambda}H\right),\\
&\quad t = 1,\ldots,n,\\
&\boldsymbol{\mu}_0 \sim N(\boldsymbol{\alpha}, \beta I_{2n}),
\end{aligned}
$$

$$F = [I_n, \emptyset], \qquad G = \left[\begin{array}{ccccc|c} 3 & -1 & & & 0 & \\ -1 & 4 & -1 & & & \\ & & \ddots & & & -I_n \\ & & -1 & 4 & -1 & \\ 0 & & & -1 & 3 & \\ \hline & & I_n & & & \emptyset \end{array}\right], \qquad H = \begin{bmatrix} I_n & \emptyset \\ \emptyset & \emptyset \end{bmatrix},$$

where $\mathbf{0}_n$ is the $n \times 1$ zero vector, $\emptyset$ is the $n \times n$ zero matrix, $\boldsymbol{\alpha}$ is the $2n \times 1$ vector $(\alpha, \dots, \alpha)'$, and $\alpha$ and $\beta$ are unknown parameters. The selection of $\alpha$ and $\beta$ will be discussed in the last section.

The above model can be written in the density form as

$$p(\boldsymbol{y}_t | \boldsymbol{\mu}_t, \sigma^2) \propto \sigma^{-n} \exp\left\{ -\frac{1}{2\sigma^2} (\boldsymbol{y}_t - F\boldsymbol{\mu}_t)'(\boldsymbol{y}_t - F\boldsymbol{\mu}_t) \right\},$$

$$p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \lambda, \sigma^2) \propto \sigma^{-n} \lambda^{n/2} \exp\left\{ -\frac{\lambda}{2\sigma^2} (\boldsymbol{\mu}_t - G\boldsymbol{\mu}_{t-1})'(\boldsymbol{\mu}_t - G\boldsymbol{\mu}_{t-1}) \right\},$$

$$p(\boldsymbol{\mu}_0) \propto \beta^{-n} \exp\left\{ -\frac{1}{2\beta} (\boldsymbol{\mu}_0 - \boldsymbol{\alpha})'(\boldsymbol{\mu}_0 - \boldsymbol{\alpha}) \right\}.$$

By Bayes' theorem, the joint posterior density of $\boldsymbol{\mu}_n, \dots, \boldsymbol{\mu}_0$ is obtained as

$$\begin{aligned} & p(\boldsymbol{\mu}_n, \dots, \boldsymbol{\mu}_0 | \boldsymbol{y}, \lambda, \sigma^2) \\ & \quad \propto \prod_{t=1}^{n} p(\boldsymbol{y}_t | \boldsymbol{\mu}_t, \sigma^2) \cdot \prod_{t=1}^{n} p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \lambda, \sigma^2) \cdot p(\boldsymbol{\mu}_0) \\ & \quad \propto \sigma^{-2n^2} \lambda^{n^2/2} \beta^{-n} \exp\left[ -\frac{1}{2\sigma^2} \left\{ \sum_{t=1}^{n} (\boldsymbol{y} - F\boldsymbol{\mu}_t)'(\boldsymbol{y} - F\boldsymbol{\mu}_t) \right.\right. \\ & \qquad \left.\left. + \lambda \sum_{t=1}^{n} (\boldsymbol{\mu}_t - G\boldsymbol{\mu}_{t-1})'(\boldsymbol{\mu}_t - G\boldsymbol{\mu}_{t-1}) + \frac{\sigma^2}{\beta} (\boldsymbol{\mu}_0 - \boldsymbol{\alpha})'(\boldsymbol{\mu}_0 - \boldsymbol{\alpha}) \right\} \right]. \end{aligned}$$

We estimate the trend by using the posterior means of $\boldsymbol{\mu}_n, \dots, \boldsymbol{\mu}_0$, which can be obtained by minimizing the cost

$$\sum_{t=1}^{n} (\boldsymbol{y} - F\boldsymbol{\mu}_t)'(\boldsymbol{y} - F\boldsymbol{\mu}_t) + \lambda \sum_{t=1}^{n} (\boldsymbol{\mu}_t - G\boldsymbol{\mu}_{t-1})'(\boldsymbol{\mu}_t - G\boldsymbol{\mu}_{t-1}) + \frac{\sigma^2}{\beta} (\boldsymbol{\mu}_0 - \boldsymbol{\alpha})'(\boldsymbol{\mu}_0 - \boldsymbol{\alpha}).$$

This cost function can be written as

$$\begin{aligned} \text{(2.1)} \quad & \sum_{s=1}^{n} \sum_{t=1}^{n} (y_{st} - x_{st})^2 + \lambda \sum_{s=1}^{n} \sum_{t=0}^{n-1} (4x_{st} - x_{s-1,t} - x_{s+1,t} - x_{s,t-1} - x_{s,t+1})^2 \\ & \qquad + \frac{\sigma^2}{\beta} \sum_{s=1}^{n} \sum_{t=-1}^{0} (x_{st} - \alpha)^2 \end{aligned}$$

with the condition, $x_{0t} = x_{1t}$ and $x_{n+1,t} = x_{nt}$. Therefore, our $b(x_{st}\text{'s})$ is

$$b(x_{st}\text{'s}|\lambda, \sigma^2) = \lambda \sum_{(s,t)\in\partial_S} (4x_{st} - x_{s-1,t} - x_{s+1,t} - x_{s,t-1} - x_{s,t+1})^2 + \frac{\sigma^2}{\beta}\sum_{s=1}^{n}\sum_{t=-1}^{0}(x_{st} - \alpha)^2$$

where $\partial_S = \{(s,t)|1 \le s \le n,\ t = 0, 1 \text{ or } s = 1, n,\ 2 \le t \le n-1\}$.

## 3. Procedure

### 3.1 *The Kalman filter*

In this section, we review the Kalman filter, and show that the maximum likelihood estimation procedure for $\lambda$ can be constructed with the computer memory size of $O(n^2)$. We refer to Anderson and Moore (1979) and Kitagawa (1987).

Let $\boldsymbol{Y}_t = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_t')'$. To clearly explain the derivation of the likelihood defined in the Kalman filter, we first present the prediction and filtering formulas in the state-space approach, which are given, respectively, as

$$p(\boldsymbol{\mu}_t|\boldsymbol{Y}_{t-1}, \lambda, \sigma^2) = \int_{R(\boldsymbol{\mu}_{t-1})} p(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1}, \lambda, \sigma^2)p(\boldsymbol{\mu}_{t-1}|\boldsymbol{Y}_{t-1}, \lambda, \sigma^2)d\boldsymbol{\mu}_{t-1},$$
$$p(\boldsymbol{\mu}_t|\boldsymbol{Y}_t, \lambda, \sigma^2) = p(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \sigma^2)p(\boldsymbol{\mu}_t|\boldsymbol{Y}_{t-1}, \lambda, \sigma^2)/p(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}, \lambda, \sigma^2)$$

where $p(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}, \lambda, \sigma^2) = \int_{R(\boldsymbol{\mu}_t)} p(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \sigma^2)p(\boldsymbol{\mu}_t|\boldsymbol{Y}_{t-1}, \lambda, \sigma^2)d\boldsymbol{\mu}_t$. The time suffix $t$ runs from 1 to $n$ with the initial condition $p(\boldsymbol{\mu}_0|\boldsymbol{Y}_0, \lambda, \sigma^2) = p(\boldsymbol{\mu}_0)$. The likelihood of the model is defined by

$$L_S(\lambda, \sigma^2|\boldsymbol{y}) = \prod_{t=1}^{n} p(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}, \lambda, \sigma^2).$$

The above formulas can be rewritten by using the conditional means and covariances of $\boldsymbol{\mu}_t$'s, since the conditional distributions of $\boldsymbol{\mu}_t$'s are Gaussian in our case. Let $\hat{\boldsymbol{\mu}}_{j/k}$ and $\sigma^2\Omega_{j/k}$ be the conditional mean and covariance matrix of $\boldsymbol{\mu}_j$ given $\boldsymbol{Y}_k$, $\lambda$ and $\sigma^2$, respectively. Then, the prediction formula can be written as

$$\hat{\boldsymbol{\mu}}_{t/t-1} = G\hat{\boldsymbol{\mu}}_{t-1/t-1}, \qquad \Omega_{t/t-1} = \frac{1}{\lambda}H + G\Omega_{t-1/t-1}G'$$

and the filtering formula can be written as

$$\hat{\boldsymbol{\mu}}_{t/t} = \hat{\boldsymbol{\mu}}_{t/t-1} + \Omega_{t/t-1}F'(F\Omega_{t/t-1}F' + I_n)^{-1}(\boldsymbol{y}_t - F\hat{\boldsymbol{\mu}}_{t/t-1}),$$
$$\Omega_{t/t} = \Omega_{t/t-1} - \Omega_{t/t-1}F'(F\Omega_{t/t-1}F' + I_n)^{-1}F\Omega_{t/t-1}.$$

Following this, the log likelihood of the model can be written as

$$-2\log L_S(\lambda, \hat{\sigma}^2(\lambda)|\boldsymbol{y}) = n^2 \log\hat{\sigma}^2(\lambda) + \sum_{t=1}^{n}\log|F\Omega_{t/t-1}F' + I_n| + const.$$

where $\hat{\sigma}^2(\lambda) = n^{-2}\sum_{t=1}^{n}(\boldsymbol{y}_t - F\hat{\boldsymbol{\mu}}_{t/t-1})'(F\Omega_{t/t-1}F' + I_n)^{-1}(\boldsymbol{y}_t - F\hat{\boldsymbol{\mu}}_{t/t-1})$. These formulas agree with those in the Kalman filter. The maximum likelihood estimate of $\lambda$ can be obtained by maximizing $\log L_S(\lambda, \hat{\sigma}^2(\lambda)|\boldsymbol{y})$.

The matrices which need to be stored at a time in the Kalman filter are the two matrices, $\Omega_{t/t-1}$ and $\Omega_{t/t}$. Each of them is a $2n \times 2n$ matrix. Therefore, the required computer memory size is given by $O(n^2)$.

## 3.2 *Smoothing*

Once the maximum likelihood estimates of $\lambda$ and $\sigma^2$ are obtained, the trend can be estimated by minimizing (2.1). However, the trend also can be estimated by using the smoothing formula in the state-space approach. In this section, we review the smoothing formula, and present a procedure to reduce the computer memory size required in smoothing.

The smoothing formula is given by

$$p(\boldsymbol{\mu}_t|\boldsymbol{y}, \lambda, \sigma^2) = \int_{R(\boldsymbol{\mu}_{t+1})} \frac{p(\boldsymbol{\mu}_{t+1}|\boldsymbol{\mu}_t, \lambda, \sigma^2)p(\boldsymbol{\mu}_t|\boldsymbol{Y}_t, \lambda, \sigma^2)}{p(\boldsymbol{\mu}_{t+1}|\boldsymbol{Y}_t, \lambda, \sigma^2)} p(\boldsymbol{\mu}_{t+1}|\boldsymbol{y}, \lambda, \sigma^2)d\boldsymbol{\mu}_{t+1}.$$

The time suffix $t$ runs from $n-1$ to $1$ with the initial condition $p(\boldsymbol{\mu}_n|\boldsymbol{y}, \lambda, \sigma^2) = p(\boldsymbol{\mu}_n|\boldsymbol{Y}_n, \lambda, \sigma^2)$. Since the conditional distributions of $\boldsymbol{\mu}_t$'s are Gaussian in our case, this formula can be rewritten as

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{t/n} &= \hat{\boldsymbol{\mu}}_{t/t} + A_t(\hat{\boldsymbol{\mu}}_{t+1/n} - \hat{\boldsymbol{\mu}}_{t+1/t}),\\ \Omega_{t/n} &= \Omega_{t/t} + A_t(\Omega_{t+1/n} - \Omega_{t+1/t})A_t'\end{aligned}$$

where $A_t = \Omega_{t/t}G'\Omega_{t+1/t}^{-1}$. The trend can be estimated by using $\hat{\boldsymbol{\mu}}_{t/n}$'s, since $p(\boldsymbol{\mu}_t|\boldsymbol{y}, \lambda, \sigma^2)$ can be regarded as the marginal posterior density induced from the joint posterior density $p(\boldsymbol{\mu}_n, \dots, \boldsymbol{\mu}_0|\boldsymbol{y}, \lambda, \sigma^2)$.

To calculate every $\hat{\boldsymbol{\mu}}_{t/n}$ by the above formula, it is necessary that all of the quantities $\hat{\boldsymbol{\mu}}_{t/t}$, $\hat{\boldsymbol{\mu}}_{t+1/n}$ and $A_t$ are known. They can be obtained only after completing the Kalman filter. The number of the matrices $A_t$ is $n$, and each $A_t$ is a $2n \times 2n$ matrix. Therefore, the memory size required in smoothing is given by $O(n^3)$. However, it is possible to practically reduce this size. Let $m$ be the number of $A_t$'s which can be stored at a time. The flow of the procedure to reduce the required memory size is as follows:

0. Set $K \leftarrow n-1$ and $J \leftarrow K - m$.

1. The Kalman filter

   1.0 (*initial prediction*) Calculate $\hat{\boldsymbol{\mu}}_{1/0}$ and $\Omega_{1/0}$ and set $t \leftarrow 1$.

   1.1 (*filtering*) Calculate $\hat{\boldsymbol{\mu}}_{t/t}$ and $\Omega_{t/t}$.

   1.2 (*prediction*) Calculate $\hat{\boldsymbol{\mu}}_{t+1/t}$ and $\Omega_{t+1/t}$.

   1.3 If $t > J$, then calculate $A_t$ and store it with $\hat{\boldsymbol{\mu}}_{t/t}$ and $\hat{\boldsymbol{\mu}}_{t+1/t}$ for smoothing.

1.4 If $t < K$, then set $t \leftarrow t + 1$ and return to Step 1.1.

2. Smoothing

2.0 If $K = n - 1$, then calculate $\hat{\mu}_{n/n}$.

2.1 Calculate $\hat{\mu}_{t/n}$ for $t = K, \ldots, J + 1$.

3. If $J > 0$, then set $K \leftarrow J$ and $J \leftarrow \max(J - m, 0)$ and return to Step 1.0

We call this procedure the split algorithm. The split algorithm reduces the required memory size by repeating the Kalman filter in Steps 1.0 $\sim$ 1.2. Therefore, the amount of computation increases. Roughly speaking, if we take $n = 1001$ and $m = 100$, the amount of computation for the Kalman filter increases by a factor $\{(n-1)/m + 1\}/2 = 5.5$, while the memory size required for $A_t$ decreases by a factor $m/(n-1) = 0.1$. (The factor of increase of the total amount of computation is about half of the above factor, as the amount of computation for smoothing does not increase.) However, if $m$ is large enough, this increment is allowable, since the Kalman filter is a fast method.

## 4. Numerical examples

In this section, we show the results of application of the proposed method to two sets of artificial data. The size of each data set is $129 \times 129$, that is, $n = 129$. In the estimation, a line search method was used to maximize $\log L_S(\lambda, \hat{\sigma}^2(\lambda)|\boldsymbol{y})$, and the split algorithm was employed with $m = 64$. The values of $\alpha$ and $\beta$ were assumed as $\alpha =$ (sample mean of $y_{st}$) and $\beta =$ (sample variance of $y_{st}$). There are several different ways of selecting $\alpha$ and $\beta$. For example, one is to assume a large value for $\beta$ with an appropriate value of $\alpha$ to decrease the effects of these parameters, and another is to select them using likelihood. However, the above assumption has the merit that suppresses an oscillation of the estimated trend near $t = 1$, and it can be calculated easily. Consequently, we employed it.

*Example* 1. The first data set was generated as follows. Let $\phi(s, t|\xi_s, \xi_t, \sigma_s^2, \sigma_t^2, \rho)$ be the probability density function of the bivariate normal distribution, where $\xi_s$, $\xi_t$, $\sigma_s^2$ and $\sigma_t^2$ are the means and the variances of $s$ and $t$, respectively, and $\rho$ is the correlation coefficient. We first generated the true image $x_{st}$ by

$$\begin{aligned} x_{st} = & -10000 \times \phi(s,t|41,41,16^2,16^2,0.3) + 10000 \times \phi(s,t|41,89,16^2,16^2,0.5) \\ & + 7500 \times \phi(s,t|89,41,16^2,16^2,-0.5) + 5000 \times \phi(s,t|89,89,16^2,16^2,0). \end{aligned}$$

Then we generated the artificial data $y_{st}$ by

$$y_{st} = x_{st} + \epsilon_{st}, \qquad \epsilon_{st} \sim i.i.d.N(0,1).$$

The contour map and the stereogram of the true image are shown in Figs. 1 and 2, and those of the artificial data are shown in Figs. 3 and 4.

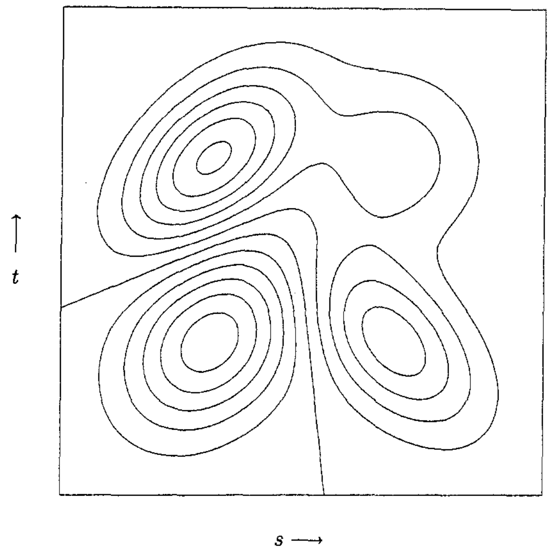


Fig. 1. A contour map of the true image in Example 1.

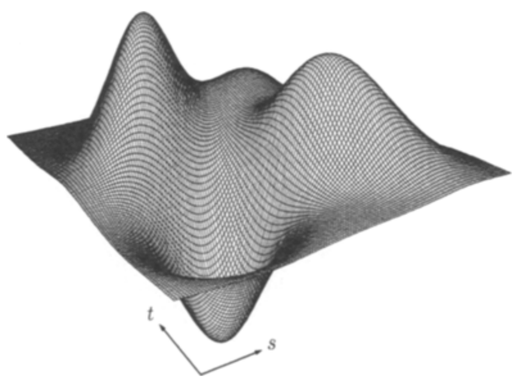


Fig. 2. A stereogram of the true image in Example 1.

The obtained maximum likelihood estimate of $\lambda$ was 33.28, suggesting that the variance of the observation noise is about 33 times as large as that of the system noise. The contour map and the stereogram of the estimated trend are shown in Figs. 5 and 6. Although the estimated trend oscillates near the boundary, especially at $t = 129$, the main shape of the true image seems to be reconstructed well. The oscillation near the boundary may be acceptable, since our primary attention is usually paid to the inside of the domain. The oscillation at $t = 129$

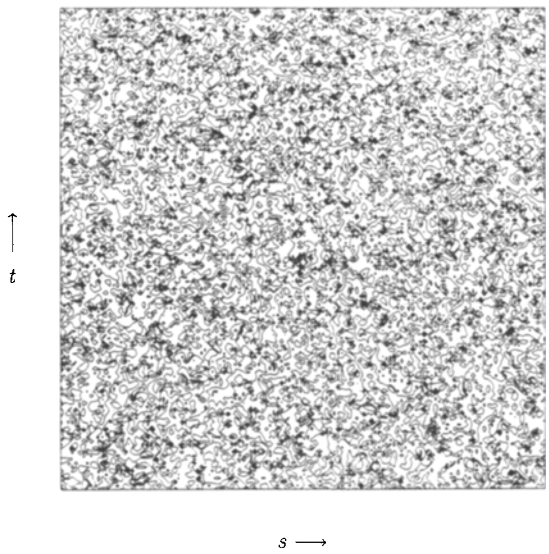


Fig. 3. A contour map of the artifical data in Example 1.

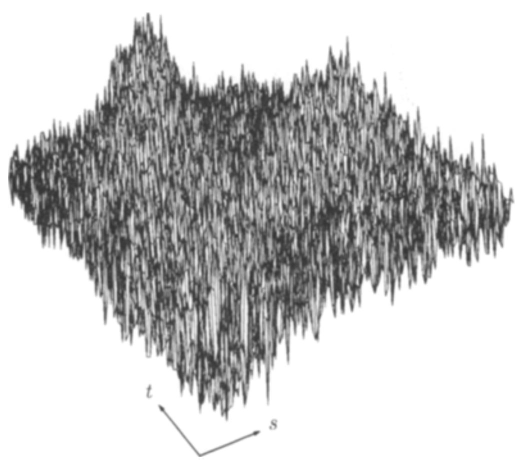


Fig. 4. A stereogram of the artifical data in Example 1.

is mainly caused by the fact that no boundary condition is assumed for $x_{st}$'s at $t = 129$.

*Example* 2. The second data set was generated by

$$y_{st} = x_{st} + \epsilon_{st}, \qquad \epsilon_{st} \sim i.i.d.N(0,1)$$

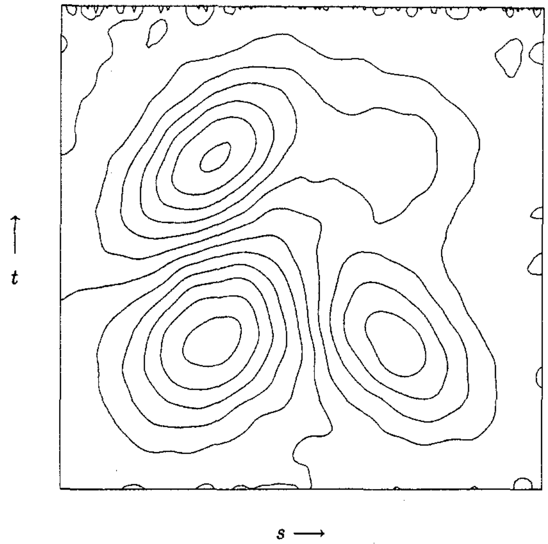


Fig. 5. A contour map of the estimated trend in Example 1.

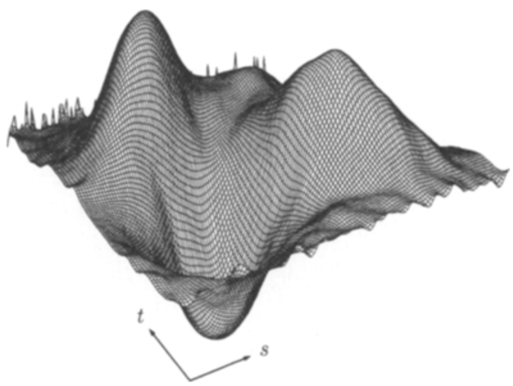


Fig. 6. A stereogram of the estimated trend in Example 1.

with the true image

$$x_{st} = \begin{cases} 2 & \text{if } (s-97)^2 + (t-33)^2 \leq 23^2 \\ 1 & \text{if } (s-97)^2 + (t-33)^2 > 23^2 \quad \text{and} \quad (s-65)^2 + (t-65)^2 \leq 46^2 \\ 0 & \text{otherwise.} \end{cases}$$

The contour map and the stereogram of the true image are shown in Figs. 7 and 8, and those of the artificial data are shown in Figs. 9 and 10. The true image in the first data set was very smooth. This may be a favorable setting to the proposed

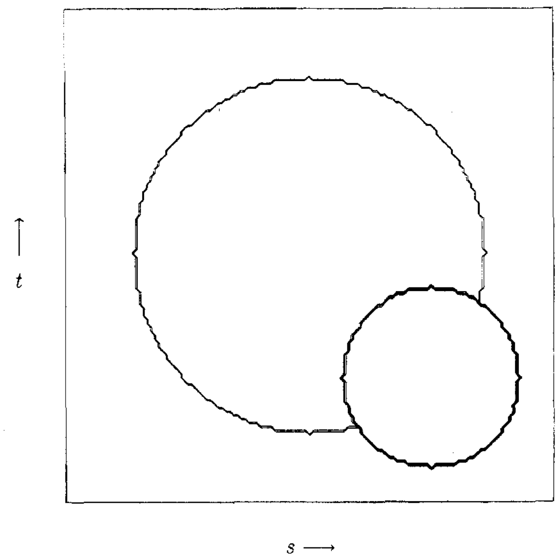


Fig. 7. A contour map of the true image in Example 2.

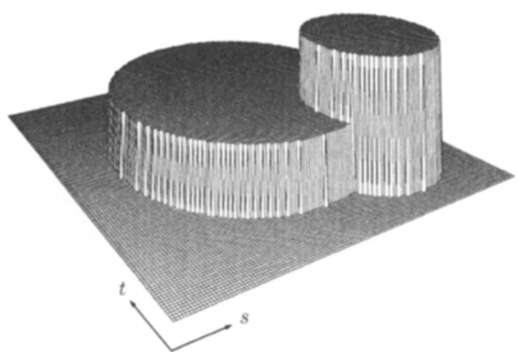


Fig. 8. A stereogram of the true image in Example 2.

model. In the second data set, the true image was assumed to involve stepwise changes, and a relatively large variance of the observation noise was employed. This is a quite unfavorable setting to the proposed model. Using this data set, we checked the performance of the method.

The obtained maximum likelihood estimate of $\lambda$ was 42.22. The contour map and the stereogram of the estimated trend are shown in Figs. 11 and 12. As expected, the proposed method could not trace the stepwise changes of the true image exactly. In addition, small oscillations appeared everywhere. These are

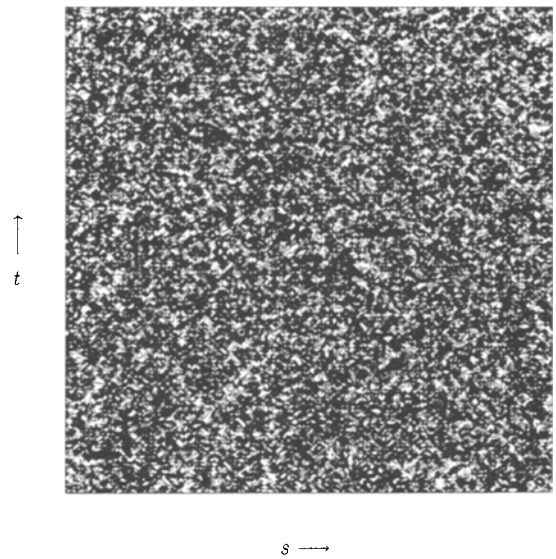


Fig. 9. A contour map of the artifical data in Example 2.

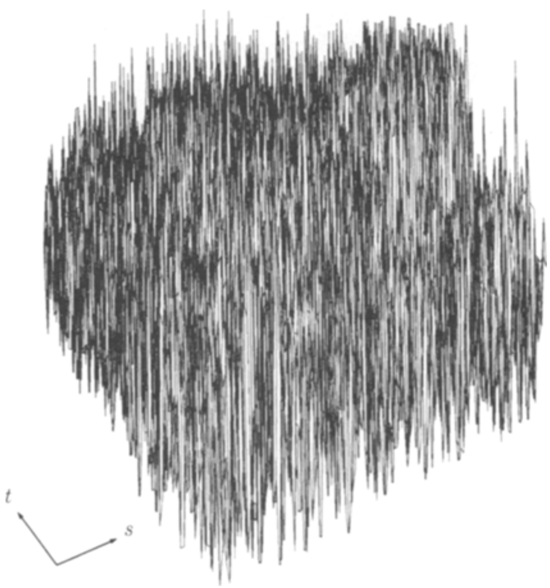


Fig. 10. A stereogram of the artifical data in Example 2.

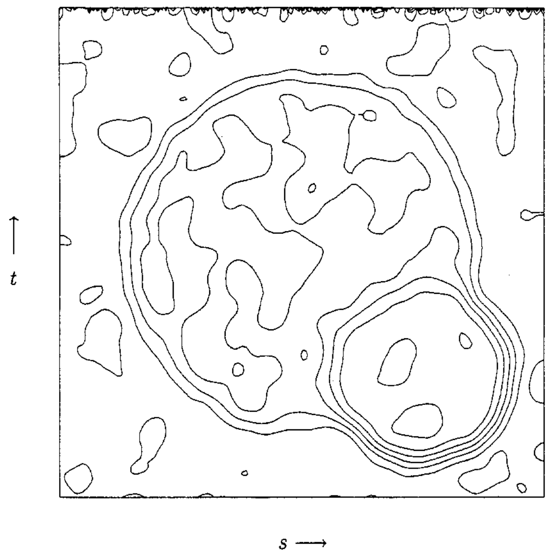


Fig. 11. A contour map of the estimated trend in Example 2.

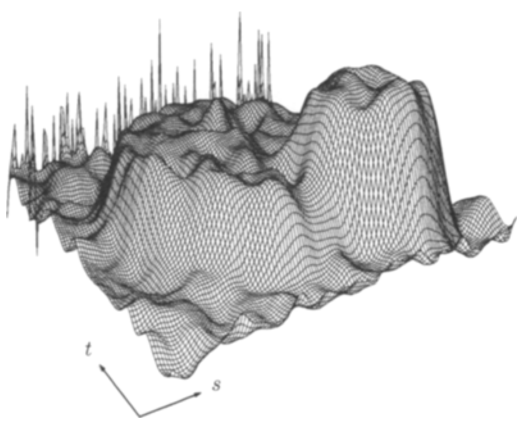


Fig. 12. A stereogram of the estimated trend in Example 2.

caused by the assumption that $\lambda$ is constant. That is, the estimate of $\lambda$ is too large and too small for tracing the stepwise changes and the flat surfaces, respectively. However, in spite of those insufficient results, the proposed method may be said to succeed in reconstructing the main shape of the true image, because the generated data are very dirty. We can at least interpret Figs. 11 and 12. To trace stepwise changes and flat surfaces exactly, a modification of the method will be necessary.

## References

Akaike, H. (1980). Likelihood and the Bayes procedure, *Trabajos de Estadistica*, **31**, 143–166.

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, New Jersey.

Kalman, R. E. (1960). A new approach to linear filtering and prediction, *ASME Journal of Basic Engineering D*, **82**, 35–45.

Kashiwagi, N. and Yanagimoto, T. (1992). Smoothing serial count data through a state-space model, *Biometrics* (in press).

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series, *J. Amer. Statist. Assoc.*, **82**, 1032–1041.

Schweppe, F. C. (1965). Evaluation of likelihood functions for Gaussian signals, *IEEE Trans. Inform. Theory*, **11**, 61–70.

Shiller, R. J. (1973). A distributed lag estimator derived from smoothness priors, *Econometrica*, **41**, 775–788.

Wahba, G. (1979). Convergence rates of "thin plate" smoothing splines when the data are noisy, *Smoothing Techniques for Curve Estimation* (eds. T. Gasser and M. Rosenblatt), 233–245, Springer, Heidelberg.

Whittaker, E. T. (1923). On a new method of graduation, *Proc. Edinburgh Math. Soc.*, **41**, 81–89.