J. Math. Soc. Japan
Vol. 57, No. 4, 2005

# On non-linear filtering problems for discrete time stochastic processes

By Yasunori OKABE and Masaya MATSUURA

(Received Apr. 30, 2004)
(Revised Dec. 10, 2004)

**Abstract.** In this paper, we shall develop the linear causal analysis for the system consisting of two flows in a real inner product space and give an algorithm for calculating the non-linear filter for a discrete stochastic system which is given by two discrete time stochastic processes, to be called a signal process and an observation process, based upon the theory of $KM_2O$-Langevin equations.

## 1. Introduction.

In [**16**] and [**5**], we have solved some problems arising from Masani and Wiener's work ([**4**]) on the non-linear prediction problem for discrete time stochastic processes under Dobrushin-Minlos' regularity condition, based upon the theory of $KM_2O$-Langevin equations for discrete time stochastic processes.

After the Kalman-Buchy's works for the linear filtering problem for a Gaussian system of Markovian type ([**2**], [**3**]), the extended Kalman filter has been derived based upon Taylor approximations of the non-linear system of Markovian type ([**18**]). However, it is said that the linearization of the non-linear system by Taylor approximations provides an insufficiently accurate representation in many cases ([**7**]).

The purpose of this paper is to develop the linear causal analysis for the system consisting of two flows in a real inner product space $W$. By a $d$-flow $\mathbf{Z} = (Z(n); 0 \leq n \leq N)$ in $W$, we mean a function $Z : \{0, 1, \ldots, N\} \to W^d$, where $d, N \in \mathbf{N} \equiv \{1, 2, \ldots\}$. As its application, we shall obtain an algorithm for calculating the non-linear filter for a discrete stochastic system consisting of a signal process and an observation process without Dobrushin-Minlos' regularity condition.

In Section 2, for any $d_1$-flow $\mathbf{X}$ and any $d_2$-flow $\mathbf{Y}$ in $W$, we shall introduce the first kind of the minimum filtering matrix function $D^0(\mathbf{X}|\mathbf{Y}) = (D^0(\mathbf{X}|\mathbf{Y})(n,k); 0 \leq k \leq n \leq N)$, by using the theory of weight transformations developed in [**5**]:

$$P_{\mathbf{M}_0^n(\mathbf{Y})} X(n) = \sum_{k=0}^{n} D^0(\mathbf{X}|\mathbf{Y})(n,k) Y(k) \qquad (0 \leq n \leq N). \tag{1.1}$$

---

2000 *Mathematics Subject Classification.* Primary 60G25; Secondary 60G12, 82C05.

*Key Words and Phrases.* $KM_2O$-Langevin equation, non-linear filtering problem, filtering matrix function.

This work is partially supported by Grand-in-Aid for Science Research (No. 14340030), the Ministry of Education, Science, Sports and Culture, Japan, and Superrobust Computation Project of the 21st Century COE Program "Information Science and Technology Strategic Core".

Moreover, we shall obtain an algorithm for calculating the filtering matrix function $D^0(\mathbf{X}|\mathbf{Y})$.

In Section 3, we shall consider any $d_1$-dimensional square integrable stochastic process $\mathbf{X} = (X(n); n \in \mathbf{N}^*)$ and any $d_2$-dimensional stochastic process $\mathbf{Y} = (Y(n); n \in \mathbf{N}^*)$, where $\mathbf{N}^* \equiv \{0, 1, 2, \ldots\}$. We shall construct a generating system for the stochastic process $\mathbf{Y}$, by modifying the idea in [**16**], where we have constructed a generating system for multi-dimensional stochastic process with time parameter space $\{\ell, \ell+1, \ldots, r\}$ ($\ell$, $r \in \mathbf{Z}, \ell < r$) under Dobrushin-Minlos' regularity condition. Applying the results in Section 2 to the generating system, we shall obtain a formula for calculating the non-linear filters $E(X(n)|\mathscr{B}_0^n(\mathbf{Y}))$.

The authors would like to express their gratitude to the referees for their constructive and valuable advice.

## 2. The minimum filtering matrix functions.

[2.1] Let $\mathbf{X} = (X(n); 0 \le n \le N)$ and $\mathbf{Y} = (Y(n); 0 \le n \le N)$ be any $d_1$-flow and any $d_2$-flow in the real inner product space $W$, respectively. We define the covariance matrix function $R(\mathbf{X}, \mathbf{Y}) = (R(\mathbf{X}, \mathbf{Y})(m, n); 0 \le m, n \le N)$ by

$$R(\mathbf{X}, \mathbf{Y})(m, n) \equiv (X(m), {}^tY(n)), \tag{2.1}$$

where $(\star, {}^t*)$ denotes the inner product matrix of $(d_1, d_2)$-type of the vectors $\star$ and $*$. In particular, we put $R(\mathbf{Y}) \equiv R(\mathbf{Y}, \mathbf{Y})$.

For each integer $n$ $(0 \le n \le N)$, we define a closed subspace $\mathbf{M}_0^n(\mathbf{Y})$ of $W$ by

$$\mathbf{M}_0^n(\mathbf{Y}) \equiv [\{Y_j(m); 1 \le j \le d_2, 0 \le m \le n\}], \tag{2.2}$$

where for any subset $S$ of $W$, $[S]$ stands for the closed subspace of $W$ which is generated by all elements in $S$ and $Y_j(m)$ is the $j$th component of $Y(m)$. Projecting each component of $X(n)$ onto the subspace $\mathbf{M}_0^n(\mathbf{Y})$, we can get a matrix function $D(\mathbf{X}|\mathbf{Y}) = (D(\mathbf{X}|\mathbf{Y})(n, k); 0 \le k \le n \le N)$ such that

$$P_{\mathbf{M}_0^n(\mathbf{Y})} X(n) = \sum_{k=0}^{n} D(\mathbf{X}|\mathbf{Y})(n, k) Y(k) \qquad (0 \le n \le N). \tag{2.3}$$

This matrix function is uniquely determined through relations (2.3) if the flow $\mathbf{Y}$ is non-degenerate, that is, $\{Y_j(n); 1 \le j \le d_2, 0 \le n \le N\}$ is linearly independent in $W$. But this is not the case in general. We denote by $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$ the set of all such matrix functions. Let us fix any element $D(\mathbf{X}|\mathbf{Y})$ of $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$. For each $n$ $(1 \le n \le N+1)$, we define a symmetric matrix $T_+(\mathbf{Y})(n)$ of order $nd_2$, two matrices $\Delta(\mathbf{X}|\mathbf{Y})(n)$ and $S(\mathbf{X}, \mathbf{Y})(n)$ of $(nd_2, d_1)$-type by

$$T_+(\mathbf{Y})(n) \equiv (R(\mathbf{Y})(k-1, \ell-1))_{1 \le k, \ell \le n}, \tag{2.4}$$

$$\Delta(\mathbf{X}|\mathbf{Y})(n) \equiv {}^t(D(\mathbf{X}|\mathbf{Y})(n-1, 0), D(\mathbf{X}|\mathbf{Y})(n-1, 1), \ldots, D(\mathbf{X}|\mathbf{Y})(n-1, n-1)), \tag{2.5}$$

$$S(\mathbf{X}, \mathbf{Y})(n) \equiv {}^t(R(\mathbf{X}, \mathbf{Y})(n-1, 0), R(\mathbf{X}, \mathbf{Y})(n-1, 1), \ldots, R(\mathbf{X}, \mathbf{Y})(n-1, n-1)). \tag{2.6}$$

Taking the inner product of the both-hand sides in (2.3) and $Y(\ell)$ $(0 \le \ell \le n)$, we get

LEMMA 2.1. *For each* $n$ $(1 \le n \le N+1)$, $T_+(\mathbf{Y})(n)\Delta(\mathbf{X}|\mathbf{Y})(n) = S(\mathbf{X},\mathbf{Y})(n)$.

A converse statement of Lemma 2.1 can be proved by using the same method as we used in the proof of Lemma 3.3 in [**5**].

LEMMA 2.2. *Let* $H(n)$ $(1 \le n \le N+1)$ *be any matrix of* $(nd_2, d_1)$*-type such that* $T_+(\mathbf{Y})(n)H(n) = S(\mathbf{X},\mathbf{Y})(n)$. *Devide* $H(n)$ *into submatrices as* $H(n) = {}^t(h(n-1,0), h(n-1,1),\dots,h(n-1,n-1))$, *where* $h(n,k)$*'s are matrices of* $(d_1,d_2)$*-type. Then the matrix function* $h \equiv (h(n,k); 0 \le k \le n \le N)$ *belongs to* $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$.

We shall find a constructive way to obtain a nice element of $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$. Let $\boldsymbol{\xi} = (\xi(n); 0 \le n \le N)$ be any non-degenerate $d_2$-flow in $W$ such that

$$R(\mathbf{X},\boldsymbol{\xi}) = 0 \quad \text{and} \quad R(\mathbf{Y},\boldsymbol{\xi}) = 0. \tag{2.7}$$

For each $w > 0$, we define a $d_2$-flow $\mathbf{Y}^w = (Y^w(n); 0 \le n \le N)$ by

$$Y^w(n) \equiv Y(n) + w\,\xi(n). \tag{2.8}$$

We call the transformation from $\mathbf{Y}$ to $\mathbf{Y}^w$ and the flow $\boldsymbol{\xi}$ the weight transformation with weight $w$ and the additive noise flow, respectively. Then we can easily verify the following Lemma 2.3.

LEMMA 2.3.
(i) $R(\mathbf{Y}^w) = R(\mathbf{Y}) + w^2R(\boldsymbol{\xi})$ $\quad(w > 0)$
(ii) $R(\mathbf{X},\mathbf{Y}^w) = R(\mathbf{X},\mathbf{Y})$ $\quad(w > 0)$
(iii) $T_+(\mathbf{Y}^w)(n) = T_+(\mathbf{Y})(n) + w^2T_+(\boldsymbol{\xi})(n)$ $\quad(w > 0, 1 \le n \le N+1)$
(iv) $S(\mathbf{X},\mathbf{Y}^w)(n) = S(\mathbf{X},\mathbf{Y})(n)$ $\quad(w > 0, 1 \le n \le N+1)$.

Let $\boldsymbol{\eta} = (\eta(n); 0 \le n \le N)$ be any non-degenerate $d_2$-flow in $W$. We define a norm $\|\cdot\|_{\boldsymbol{\eta}}$ on the set $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$ as follows:

$$\|D(\mathbf{X}|\mathbf{Y})\|_{\boldsymbol{\eta}} \equiv \left(\sum_{n=0}^{N}\sum_{j=1}^{d_1}\left\|\eta_j^{D(\mathbf{X}|\mathbf{Y})}(n)\right\|_W^2\right)^{1/2}, \tag{2.9}$$

where $\|\cdot\|_W$ denotes the norm on $W$ induced by the inner product $(\star,*)$ and $\boldsymbol{\eta}^{D(\mathbf{X}|\mathbf{Y})} = (\eta^{D(\mathbf{X}|\mathbf{Y})}(n); 0 \le n \le N)$ is the $d_1$-flow in $W$ defined by

$$\eta^{D(\mathbf{X}|\mathbf{Y})}(n) \equiv \sum_{k=0}^{n} D(\mathbf{X}|\mathbf{Y})(n,k)\eta(k). \tag{2.10}$$

By a direct calculation, we have

LEMMA 2.4. $\|D(\mathbf{X}|\mathbf{Y})\|_{\boldsymbol{\eta}} = \left(\sum_{n=1}^{N+1}\|T_+(\boldsymbol{\eta})(n)^{1/2}\Delta(\mathbf{X}|\mathbf{Y})(n)\|^2\right)^{1/2}$, *where* $\|*\|$ *stands for the Euclidean norm of the matrix* $*$.

We are now in a position to state one of the main theorems in this paper.

Theorem 2.1. *Let $\boldsymbol{\xi}$ be any additive noise flow for the flow* $\mathbf{Y}$ *satisfying* (2.7) *and let* $\mathbf{Y}^w$ *be the flow defined by* (2.8). *Then*

(i) $D(\mathbf{X}|\mathbf{Y}^w)$ *converges as* $w \to 0$ *and the limit* $D^0(\mathbf{X}|\mathbf{Y};\boldsymbol{\xi})$ *belongs to* $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$.

(ii) $D^0(\mathbf{X}|\mathbf{Y};\boldsymbol{\xi})$ *is the unique element of* $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$ *minimizing the norm* $\|\cdot\|_{\boldsymbol{\xi}}$.

To prove this theorem, we show the following theorem on linear algebra.

Theorem 2.2. *Let $A$ and $B$ be any symmetric matrices of order $n$, and let $C$ be any matrix of $(n,m)$-type such that*

(a) $A \geq 0$ *and* $B > 0$,

(b) *there exists a matrix $X$ of $(n,m)$-type for which $AX = C$.*

*For each $\varepsilon > 0$, define a matrix $F_\varepsilon$ of $(n,m)$-type by $F_\varepsilon \equiv (A+\varepsilon B)^{-1}C$. Then*

(i) *$F_\varepsilon$ converges as $\varepsilon \to 0$, and the limit $F_0 \equiv \lim_{\varepsilon\to 0} F_\varepsilon$ satisfies $AF_0 = C$,*

(ii) *$F_0$ is the unique element of $\{X; AX = C\}$ minimizing the norm $\|B^{1/2}X\|$.*

Proof. We have already proved (i) for more general case, and (ii) for the case where $B = I_n$ in Theorems 4.2 and 4.6 of [**5**], respectively. We put $\tilde{A} \equiv B^{-1/2} AB^{-1/2}$ and $\tilde{C} \equiv B^{-1/2}C$. Then we see that $\tilde{A}, B = I_n$ and $\tilde{C}$ fulfil the assumptions (a) and (b). So the general case is reduced to the case $B = I_n$. □

Proof of Theorem 2.1. Let $A \equiv T_+(\mathbf{Y})(n)$, $B \equiv T_+(\boldsymbol{\xi})(n)$ and $C \equiv S(\mathbf{X},\mathbf{Y})(n)$. Then $A$ and $B$ satisfy Theorem 2.2(a). By Lemma 2.1 we see that $C$ satisfies Theorem 2.2(b). So we can apply Theorem 2.2 to obtain Theorem 2.1. □

Immediately from Theorem 2.1, we have

Corollary 2.1. *Let $\boldsymbol{\xi}, \mathbf{Y}$ and $\mathbf{Y}^w$ be as in Theorem* 2.1. *Then*

$$\lim_{w\to 0} P_{\mathbf{M}_0^n(\mathbf{Y}^w)}X(n) = P_{\mathbf{M}_0^n(\mathbf{Y})}X(n) \qquad (0 \leq n \leq N).$$

The following theorem gives a converse statement of Theorem 2.1 in a sense.

Theorem 2.3. *Let us fix any $d_1, d_2 \in \mathbf{N}$ and any $N \in \mathbf{N}^*$. Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be any non-degenerate $d_2$-flows in $W$ satisfying* $\dim(\mathbf{M}_0^N(\boldsymbol{\xi},\boldsymbol{\eta}))^\perp \geq 1$. *Then the following conditions are equivalent*:

(i) $R(\boldsymbol{\xi}) = \lambda R(\boldsymbol{\eta})$ *for some* $\lambda > 0$;

(ii) $D^0(\mathbf{X}|\mathbf{Y};\boldsymbol{\xi}) = D^0(\mathbf{X}|\mathbf{Y};\boldsymbol{\eta})$ *for any $d_1$-flow $\mathbf{X}$ and any $d_2$-flow $\mathbf{Y}$ in* $(\mathbf{M}_0^N(\boldsymbol{\xi},\boldsymbol{\eta}))^\perp$.

For the proof of Theorem 2.3, we need the following Lemma 2.5 and Theorem 2.4.

Lemma 2.5. *For any matrix $T_{11}$ of order $(N+1)d_2$, any matrix $T_{12}$ of $((N+1)d_2, d_1)$-type and any matrix $T_{22}$ of order $d_1$ such that the matrix $T \equiv \begin{pmatrix} T_{11} & T_{12} \\ {}^tT_{12} & T_{22} \end{pmatrix}$ is non-negative definite and* $\operatorname{rank} T \leq \dim W$, *there exist a $d_1$-flow $\mathbf{X}$ and a $d_2$-flow $\mathbf{Y}$ in $W$ such that $T_+(\mathbf{Y})(N+1) = T_{11}$, $S(\mathbf{X},\mathbf{Y})(N+1) = T_{12}$ and $R(\mathbf{X})(N,N) = T_{22}$.*

PROOF. Let $r \equiv \operatorname{rank} T$. Since $T$ is non-negative definite, there exist a matrix $Q$ of order $(N+1)d_2 + d_1$ and a positive definite matrix $D_r$ of order $r$ such that $T = Q\begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix} {}^tQ$. We put $M \equiv Q\begin{pmatrix} D_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix}$. Then $M\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} {}^tM = T$. We define a $d_2$-flow $\mathbf{Y} = (Y(n); 0 \le n \le N)$ in $W$ and a vector $X(N) \in W^{d_1}$ by $({}^tY(0), {}^tY(1), \ldots, {}^tY(N), {}^tX(N)) \equiv (\xi_1, \xi_2, \ldots, \xi_r, 0, \ldots, 0)\, {}^tM$, where $\{\xi_j; 1 \le j \le r\}$ is an orthonormal system in $W$. We also define a $d_1$-flow $\mathbf{X}$ in $W$ by choosing $X(0), X(1), \ldots, X(N-1)$ arbitrarily. Then we can see that Lemma 2.5 holds. □

THEOREM 2.4. *Let $B$ and $G$ be any positive definite matrices of order $n$. Then the following conditions are equivalent*:

(i) $B = \lambda G$ *for some* $\lambda > 0$;

(ii) $F_0$ *in Theorem* 2.2 *is the unique element of* $\{X; AX = C\}$ *minimizing the norm* $\|G^{1/2}X\|$ *for any matrices* $A, C$ *with* $\operatorname{rank}(A) = 1$ *and the conditions in Theorem* 2.2.

PROOF. By Theorem 2.2, we have only to prove that (ii) implies (i). We have already proved it in Theorem 4.8 in [**5**] for the case where $G = I_n$, which deduces the general case by putting $\tilde{A} \equiv G^{-1/2}AG^{-1/2}$ and $\tilde{C} \equiv G^{-1/2}C$. □

PROOF OF THEOREM 2.3. We assume (i). For any $d_1$-flow $\mathbf{X}$ and any $d_2$-flow $\mathbf{Y}$ in $(\mathbf{M}_0^N(\boldsymbol{\xi}, \boldsymbol{\eta}))^\perp$, Lemma 2.4 shows that $\|D(\mathbf{X}|\mathbf{Y})\|_{\boldsymbol{\xi}} = \lambda^{1/2}\|D(\mathbf{X}|\mathbf{Y})\|_{\boldsymbol{\eta}}$ $(D(\mathbf{X}|\mathbf{Y}) \in \mathscr{LMF}(\mathbf{X}|\mathbf{Y}))$. This combined with Theorem 2.1(ii) implies that $D^0(\mathbf{X}|\mathbf{Y}; \boldsymbol{\xi})$ minimizes the norm $\|\cdot\|_{\boldsymbol{\eta}}$. Replacing $\boldsymbol{\xi}$ with $\boldsymbol{\eta}$ in Theorem 2.1(ii), we find that (ii) holds.

We now assume (ii). Let $A$ and $C$ be as in Theorem 2.4. We define a matrix $T_{AC}$ of order $(N+1)d_2 + d_1$ by $T_{AC} \equiv \begin{pmatrix} A & C \\ {}^tC & {}^tCF \end{pmatrix}$, where $F$ is a matrix of $((N+1)d_2, d_1)$-type for which $AF = C$ holds. Then $T_{AC}$ is non-negative definite and $\operatorname{rank} T_{AC} = 1$, because ${}^t\begin{pmatrix} I & -F \\ 0 & I \end{pmatrix} T_{AC} \begin{pmatrix} I & -F \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$. By Lemma 2.5, we see that there exist a $d_1$-flow $\mathbf{X}$ and a $d_2$-flow $\mathbf{Y}$ in $(\mathbf{M}_0^N(\boldsymbol{\xi}, \boldsymbol{\eta}))^\perp$ that satisfy $T_+(\mathbf{Y})(N+1) = A$ and $S(\mathbf{X}, \mathbf{Y})(N+1) = C$. Let $B \equiv T_+(\boldsymbol{\xi})(N+1)$. Then $B$ is positive definite because the noise $\boldsymbol{\xi}$ is non-degenerate. We see $A, B$ and $C$ satisfy the conditions in Theorem 2.2 with $m = d_1$ and $n = (N+1)d_2$. Here Theroem 2.2(b) is ensured by Lemma 2.1. It follows from Lemma 2.1 that the corresponding $F_0$ in Theorem 2.2 equals $\Delta(\mathbf{X}|\mathbf{Y})(N+1)$, which is defined by taking $D(\mathbf{X}|\mathbf{Y}) \equiv D^0(\mathbf{X}|\mathbf{Y}; \boldsymbol{\xi})$ in (2.5). We now take $G \equiv T_+(\boldsymbol{\eta})(N+1)$ and show $G$ satisfies Theorem 2.4(ii). Since $D^0(\mathbf{X}|\mathbf{Y}; \boldsymbol{\xi}) = D^0(\mathbf{X}|\mathbf{Y}; \boldsymbol{\eta})$, $D^0(\mathbf{X}|\mathbf{Y}; \boldsymbol{\xi})$ minimizes the norm $\|\cdot\|_{\boldsymbol{\eta}}$. Hence by Lemma 2.4, $A, C, F_0$ and $G$ thus defined satisfy Theorem 2.4(ii). So by Theorem 2.4 we obtain $T_+(\boldsymbol{\xi})(N+1) = \lambda T_+(\boldsymbol{\eta})(N+1)$, which implies (i). □

We shall analyze the special case where the additive noise flow satisfies the white noise property. For that purpose, we define a norm $\|\cdot\|$ on the set $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$ by

$$\|D(\mathbf{X}|\mathbf{Y})\| \equiv \left( \sum_{n=0}^{N} \sum_{k=0}^{n} \sum_{j=1}^{d_1} \sum_{\ell=1}^{d_2} D_{j\ell}(\mathbf{X}|\mathbf{Y})(n,k)^2 \right)^{1/2}. \tag{2.11}$$

Immediately from (2.9) and (2.11), we have

LEMMA 2.6. *If a $d_2$-flow $\boldsymbol{\eta}$ satisfies the white noise property, that is, $T_+(\boldsymbol{\eta})(N+1) = I_{(N+1)d_2}$, then $\|D(\mathbf{X}|\mathbf{Y})\|_{\boldsymbol{\eta}} = \|D(\mathbf{X}|\mathbf{Y})\|$ for any $D(\mathbf{X}|\mathbf{Y}) \in \mathscr{LMF}(\mathbf{X}|\mathbf{Y})$.*

Combining Theorems 2.1, 2.3 and Lemma 2.6, we obtain

THEOREM 2.5. *Let us fix any* $d_1, d_2 \in \mathbf{N}, N \in \mathbf{N}^*$ *with* $(N+1)d_2 < \dim W$ *and let* $\boldsymbol{\xi}$ *be any non-degenerate* $d_2$*-flow in* $W$*. Then, the following conditions are equivalent*:

(i) $T_+(\boldsymbol{\xi})(N+1) = \lambda I_{(N+1)d_2}$ *for some* $\lambda > 0$;

(ii) *For any* $d_1$*-flow* $\mathbf{X}$ *and any* $d_2$*-flow* $\mathbf{Y}$ *in* $(\mathbf{M}_0^N(\boldsymbol{\xi}))^\perp$*, the matrix function* $D^0(\mathbf{X}|\mathbf{Y};\boldsymbol{\xi})$ *minimizes the norm* $\|\cdot\|$ *in* $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$.

By Theorem 2.3, the minimum norm element of $\mathscr{LMF}(\mathbf{X}|\mathbf{Y})$ with respect to the norm $\|\cdot\|$ does not depend upon the white noise $\boldsymbol{\xi}$. We denote it by $D^0(\mathbf{X}|\mathbf{Y})$ and call it the first kind of the minimum filtering matrix function of the flow $\mathbf{X}$ based upon the flow $\mathbf{Y}$. We remark $D^0(\mathbf{X}|\mathbf{Y})$ satisfies (1.1) by construction. Further, we note that $D^0(\mathbf{X}|\mathbf{Y}^w) = D(\mathbf{X}|\mathbf{Y}^w)$ $(w > 0)$.

[2.2] In order to get an algorithm for calculating the matrix function $D^0(\mathbf{X}|\mathbf{Y})$, we define the flow $\nu_+(\mathbf{Y}) = (\nu_+(\mathbf{Y})(n); 0 \le n \le N)$ by

$$\nu_+(\mathbf{Y})(n) \equiv Y(n) - P_{\mathbf{M}_0^{n-1}(\mathbf{Y})} Y(n), \tag{2.12}$$

where $\mathbf{M}_0^{-1}(\mathbf{Y}) \equiv \{0\}$. We define a matrix function $V_+(\mathbf{Y}) = (V_+(\mathbf{Y})(n); 0 \le n \le N)$ by

$$V_+(\mathbf{Y})(n) \equiv R(\nu_+(\mathbf{Y}))(n,n). \tag{2.13}$$

Then we know that the flow $\nu_+(\mathbf{Y})$ satisfies the orthogonality property (2.14) and the causality relation (2.15) with the flow $\mathbf{Y}$:

$$R(\nu_+(\mathbf{Y}))(m,n) = \delta_{m,n} V_+(\mathbf{Y})(n) \quad (0 \le m, n \le N), \tag{2.14}$$

$$\mathbf{M}_0^n(\mathbf{Y}) = \mathbf{M}_0^n(\nu_+(\mathbf{Y})) \quad (0 \le n \le N). \tag{2.15}$$

Replacing $\mathbf{Y}$ by $\mathbf{Y}^w$ in (2.15) and noting that $\mathbf{Y}^w$ is non-degenerate for each $w > 0$, we can uniquely find two matrix functions $\gamma_+(\mathbf{Y}^w) = (\gamma_+(\mathbf{Y}^w)(n,k); 0 \le k < n \le N)$ and $P_+(\mathbf{Y}^w) = (P_+(\mathbf{Y}^w)(n,k); 0 \le k \le n \le N)$ such that

$$Y^w(n) = -\sum_{k=0}^{n-1} \gamma_+(\mathbf{Y}^w)(n,k) Y^w(k) + \nu_+(\mathbf{Y}^w)(n), \tag{2.16}$$

$$Y^w(n) = \sum_{k=0}^{n} P_+(\mathbf{Y}^w)(n,k) \nu_+(\mathbf{Y}^w)(k). \tag{2.17}$$

We know from Theorems 4.1, 4.5 and 7.3 in [**5**] that these matrix functions $\gamma_+(\mathbf{Y}^w)$ and $P_+(\mathbf{Y}^w)$ converge as $w \to 0$ and that respective limits $\gamma_+^0(\mathbf{Y})$ and $P_+^0(\mathbf{Y})$ satisfy (2.16) and (2.17) with the replacement of $\mathbf{Y}^w$ by $\mathbf{Y}$. In particular, the former equation is the forward KM$_2$O-Langevin equation describing the time evolution of $\mathbf{Y}$:

$$Y(n) = -\sum_{k=0}^{n-1} \gamma_+^0(\mathbf{Y})(n,k)Y(k) + \nu_+(\mathbf{Y})(n). \tag{2.18}$$

By Theorems 4.5 and 7.3 in [**5**], $\gamma_+^0(\mathbf{Y})$ and $P_+^0(\mathbf{Y})$ do not depend upon the white noise flow $\boldsymbol{\xi}$. We note that $\gamma_+^0(\mathbf{Y}^w) = \gamma_+(\mathbf{Y}^w)$ and $P_+^0(\mathbf{Y}^w) = P_+(\mathbf{Y}^w)$ $(w > 0)$. Further, we know that

$$\begin{cases} P_+^0(\mathbf{Y})(n,n) = I_{d_2} & (0 \le n \le N), \\ P_+^0(\mathbf{Y})(n,k) = -\displaystyle\sum_{\ell=k}^{n-1} \gamma_+^0(\mathbf{Y})(n,\ell)P_+^0(\mathbf{Y})(\ell,k) & (0 \le k < n \le N). \end{cases} \tag{2.19}$$

For each $w > 0$, by (2.15) with replaced $\mathbf{Y}$ by $\mathbf{Y}^w$, we can find the unique matrix function $C(\mathbf{X}|\mathbf{Y}^w) = (C(\mathbf{X}|\mathbf{Y}^w)(n,k); 0 \le k \le n \le N)$ such that

$$P_{\mathbf{M}_0^n(\mathbf{Y}^w)}X(n) = \sum_{k=0}^{n} C(\mathbf{X}|\mathbf{Y}^w)(n,k)\nu_+(\mathbf{Y}^w)(k). \tag{2.20}$$

Multiplying (2.20) by $\nu_+(\mathbf{Y}^w)(k)$, we see from (2.14) that

Theorem 2.6. *For each $n, k$ $(0 \le k \le n \le N)$,*

$$C(\mathbf{X}|\mathbf{Y}^w)(n,k) = \left(R(\mathbf{X},\mathbf{Y}^w)(n,k) + \sum_{\ell=0}^{k-1} R(\mathbf{X},\mathbf{Y}^w)(n,\ell)\,{}^t\gamma_+(\mathbf{Y}^w)(k,\ell)\right)V_+(\mathbf{Y}^w)(k)^{-1}.$$

Next, we shall obtain an algorithm for calculating the matrix function $D^0(\mathbf{X}|\mathbf{Y})$ in terms of the matrix functions $C^0(\mathbf{X}|\mathbf{Y})$ and $\gamma_+^0(\mathbf{Y})$.

Theorem 2.7. *For each $n, k$ $(0 \le k \le n \le N)$,*
(i) *A limit $C^0(\mathbf{X}|\mathbf{Y}) \equiv \lim_{w\to 0} C(\mathbf{X}|\mathbf{Y}^w)$ exists,*
(ii) $D^0(\mathbf{X}|\mathbf{Y})(n,k) = C^0(\mathbf{X}|\mathbf{Y})(n,k) + \displaystyle\sum_{\ell=k+1}^{n} C^0(\mathbf{X}|\mathbf{Y})(n,\ell)\gamma_+^0(\mathbf{Y})(\ell,k)$,
(iii) $C^0(\mathbf{X}|\mathbf{Y})(n,k) = \displaystyle\sum_{\ell=k}^{n} D^0(\mathbf{X}|\mathbf{Y})(n,\ell)P_+^0(\mathbf{Y})(\ell,k)$.

Proof. By (2.16), $\nu_+(\mathbf{Y}^w)(k) = Y^w(k) + \sum_{\ell=0}^{k-1}\gamma_+(\mathbf{Y}^w)(k,\ell)Y^w(\ell)$ $(0 \le k \le N)$. Hence, by (2.20), $D(\mathbf{X}|\mathbf{Y}^w)(n,k) = C(\mathbf{X}|\mathbf{Y}^w)(n,k) + \sum_{\ell=k+1}^{n} C(\mathbf{X}|\mathbf{Y}^w)(n,\ell)\gamma_+(\mathbf{Y}^w)(\ell,k)$ $(0 \le k \le n \le N)$. Next, substituting (2.17) into (1.1) with $\mathbf{Y}$ replaced by $\mathbf{Y}^w$, we see that $C(\mathbf{X}|\mathbf{Y}^w)(n,k) = \sum_{\ell=k}^{n} D(\mathbf{X}|\mathbf{Y}^w)(n,\ell)P_+(\mathbf{Y}^w)(\ell,k)$. Thus, letting $w$ tend to 0, we find from Theorem 2.1 that Theorem 2.7 holds. □

By virtue of Theorems 2.6 and 2.7, we can obtain an algorithm for calculating the matrix function $D^0(\mathbf{X}|\mathbf{Y})$ from the matrix function $R(\mathbf{X},\mathbf{Y})$ and the system $\{\gamma_+^0(\mathbf{Y}), V_+(\mathbf{Y})\}$. We note that the latter can be obtained from the matrix function

$R(\mathbf{Y})$ ([**6**]).

[2.3] We shall consider any $d_1$-flow $\mathbf{X} = (X(n); n \in \mathbf{N}^*)$ and $d_2$-flow $\mathbf{Y} = (Y(n); n \in \mathbf{N}^*)$ in a Hilbert space $W$. For each $N \in \mathbf{N}$, we restrict the time domain of both flows $\mathbf{X}$ and $\mathbf{Y}$ to the set $\{0, 1, \ldots, N\}$ and apply the results above. Then, by virtue of Theorem 2.1(ii), we can construct the first (resp. second) kind of the minimum filtering matrix function $D^0(\mathbf{X}|\mathbf{Y}) = (D^0(\mathbf{X}|\mathbf{Y})(n,k); 0 \le k \le n < \infty)$ (resp. $C^0(\mathbf{X}|\mathbf{Y}) = (C^0(\mathbf{X}|\mathbf{Y})(n,k); 0 \le k \le n < \infty)$) of the $d_1$-flow $\mathbf{X}$ based upon the $d_2$-flow $\mathbf{Y}$ such that

$$P_{\mathbf{M}_0^n(\mathbf{Y})} X(n) = \sum_{k=0}^{n} D^0(\mathbf{X}|\mathbf{Y})(n,k) Y(k) \qquad (n \in \mathbf{N}^*), \tag{2.21}$$

$$P_{\mathbf{M}_0^n(\mathbf{Y})} X(n) = \sum_{k=0}^{n} C^0(\mathbf{X}|\mathbf{Y})(n,k) \nu_+(\mathbf{Y})(k) \qquad (n \in \mathbf{N}^*). \tag{2.22}$$

We note that the same relations as in Theorems 2.6 and 2.7 hold for the $d_1$-flow $\mathbf{X}$ and $d_2$-flow $\mathbf{Y}$ with time domain $\mathbf{N}^*$.

## 3. A formula for calculating the non-linear filter.

We shall consider in this section any $d_1$-dimensional square integrable stochastic process $\mathbf{X} = (X(n); n \in \mathbf{N}^*)$ and any $d_2$-dimensional stochastic process $\mathbf{Y} = (Y(n); n \in \mathbf{N}^*)$ defined on a probability space $(\Omega, \mathscr{B}, P)$. For any $n \in \mathbf{N}^*$, we define the non-linear information space $\mathbf{N}_0^n(\mathbf{Y})$ for the stochastic process $\mathbf{Y}$ by

$$\mathbf{N}_0^n(\mathbf{Y}) \equiv \{Y \in L^2(\Omega, \mathscr{B}, P); Y \text{ is } \mathscr{B}_0^n(\mathbf{Y})\text{-measurable}\}, \tag{3.1}$$

where $\mathscr{B}_0^n(\mathbf{Y})$ stands for the smallest $\sigma$-field generated by $Y_j(m)$ $(0 \le m \le n, 1 \le j \le d_2)$.

As an application of Section 2, we shall give a formula for calculating the non-linear filters $E(X(n)|\mathscr{B}_0^n(\mathbf{Y}))$, which are equal to the projection of the vector $X(n)$ on $\mathbf{N}_0^n(\mathbf{Y})$ $(n \in \mathbf{N}^*)$:

$$E(X(n)|\mathscr{B}_0^n(\mathbf{Y})) = P_{\mathbf{N}_0^n(\mathbf{Y})} X(n). \tag{3.2}$$

By taking account of Theorem 10.2 in [**5**] and Theorem 2.2 in [**16**], we shall give

DEFINITION 3.1. A system $\{\mathbf{Y}^{(q)}; q \in \mathbf{N}\}$ of multi-dimensional stochastic processes is said to be a generating system of $\mathbf{N}_0^n(\mathbf{Y})$ $(n \in \mathbf{N}^*)$ if

(a) each $\mathbf{Y}^{(q)} = (Y^{(q)}(n); n \in \mathbf{N}^*)$ is a square integrable stochastic process $(q \in \mathbf{N})$,

(b) $\{\mathbf{Y}^{(q)}; q \in \mathbf{N}\}$ has a nest structure, that is, $\mathbf{Y}^{(q+1)}(n) = {}^t({}^t\mathbf{Y}^{(q)}(n), *)$,

(c) $\mathbf{N}_0^n(\mathbf{Y}) = [\{1\}] \oplus [\bigcup_{q=1}^{\infty} \mathbf{M}_0^n(\mathbf{Y}^{(q)})] \qquad (n \in \mathbf{N}^*)$.

Without Dobrushin-Minlos' condition in [**16**], we shall prove the following theorem.

THEOREM 3.1. *There exists a generating system of* $\mathbf{N}_0^n(\mathbf{Y})$ $(n \in \mathbf{N}^*)$.

PROOF. Define a $d_2$-dimensional stochastic process $\arctan(\mathbf{Y}) = (\arctan(\mathbf{Y})(n);$

$n \in \mathbf{N}^*$) by $\arctan(\mathbf{Y})(n) \equiv {}^t(\arctan(Y_1(n)), \dots, \arctan(Y_{d_2}(n)))$. Since it is a bounded process, we can use the same idea as in [**16**] to construct a generating system for the process $\arctan(\mathbf{Y})$. Noting that $\mathbf{N}_0^n(\mathbf{Y}) = \mathbf{N}_0^n(\arctan(\mathbf{Y}))$ $(n \in \mathbf{N}^*)$, we find that the above system is a desired one. $\square$

REMARK 3.1. If the process $\mathbf{Y}$ is square integrable, we can construct a generating system such that $\mathbf{Y}^{(1)} = \tilde{\mathbf{Y}}$, where $\tilde{\mathbf{Y}} = (\tilde{Y}(n); n \in \mathbf{N}^*)$ is defined by $\tilde{Y}(n) \equiv Y(n) - E(Y(n))$.

In the sequel, we shall fix any generating system. Applying the results in [2.3] of Section 2 to the $d_1$-dimensional flow $\mathbf{X}$ and the $(d_q + 1)$-dimensional flow $\mathbf{Y}^{(q)}$ in the real Hilbert space $L^2(\Omega, \mathscr{B}, P)$, we can construct the minimum filtering matrix functions $D^0(\mathbf{X}|\mathbf{Y}^{(q)})$ and $C^0(\mathbf{X}|\mathbf{Y}^{(q)})$. Applying (2.21) and (2.22) to the flows $\mathbf{X}$ and $\mathbf{Y}^{(q)}$ and letting $q$ tend to $\infty$, we obtain from the property (b) and (c) of generating system that

THEOREM 3.2. *For any $n \in \mathbf{N}^*$,*

(i) $P_{\mathbf{N}_0^n(\mathbf{Y})}X(n) = E(X(n)) + \lim_{q\to\infty} \sum_{k=0}^{n} D^0(\mathbf{X}|\mathbf{Y}^{(q)})(n,k)Y^{(q)}(k)$ *in* $L^2(\Omega, \mathscr{B}, P)$,

(ii) $P_{\mathbf{N}_0^n(\mathbf{Y})}X(n) = E(X(n)) + \lim_{q\to\infty} \sum_{k=0}^{n} C^0(\mathbf{X}|\mathbf{Y}^{(q)})(n,k)\nu_+(\mathbf{Y}^{(q)})(k)$ *in* $L^2(\Omega, \mathscr{B}, P)$.

For each $n \in \mathbf{N}^*$, we define the non-linear filtering error matrix $e_+^{(nl)}(\mathbf{X}|\mathbf{Y})(n)$ of the random variable $X(n)$ conditioned on the $\sigma$-field $\mathscr{B}_0^n(\mathbf{Y})$ by

$$e_+^{(nl)}(\mathbf{X}|\mathbf{Y})(n) \equiv E((X(n) - E(X(n)|\mathscr{B}_0^n(\mathbf{Y}))\ {}^t(X(n) - E(X(n)|\mathscr{B}_0^n(\mathbf{Y})))). \tag{3.3}$$

We shall give the following formula for calculating the non-linear filtering error matrix.

THEOREM 3.3. *For any $n \in \mathbf{N}^*$,*

$$\begin{aligned} e_+^{(nl)}(\mathbf{X}|\mathbf{Y})(n) = {} & R(\mathbf{X})(n,n) - E(X(n))\ {}^tE(X(n)) \\ & - \lim_{q\to\infty}\left\{\sum_{k=0}^{n} C^0(\mathbf{X}|\mathbf{Y}^{(q)})(n,k)V_+(\mathbf{Y}^{(q)})(k)\ {}^tC^0(\mathbf{X}|\mathbf{Y}^{(q)})(n,k)\right\}. \end{aligned}$$

PROOF. By Theorem 3.2(ii), we see that $e_+^{(nl)}(\mathbf{X}|\mathbf{Y})(n) = R(\mathbf{X})(n,n) - E(X(n))\ {}^tE(X(n)) - \lim_{q\to\infty}\sum_{k=0}^{n} C^0(\mathbf{X}|\mathbf{Y}^{(q)})(n,k)E(\nu_+(\mathbf{Y}^{(q)})(k)\ {}^tX(n))$. Further, noting that $\nu_+(\mathbf{Y}^{(q)})(k) \in \mathbf{M}_0^n(\mathbf{Y}^{(q)})$ for any $k$ $(0 \le k \le n)$, we see from (2.22) that $E(\nu_+(\mathbf{Y}^{(q)})(k)\ {}^tX(n)) = V_+(\mathbf{Y}^{(q)})(k)\ {}^tC^0(\mathbf{X}|\mathbf{Y}^{(q)})(n,k)$, which proves Theorem 3.3. $\square$

## References


[1] F. Carravetta, A. Germani and M. Raimondi, Polynomial filtering for linear discrete time non-Gaussian systems, SIAM J. Control Optim., **34** (1996), 1666–1690.

[2] R. E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME J. Basic Eng., **82D** (1960), 35–45.

[ 3 ] R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, Trans. ASME J. Basic Eng., **83D** (1961), 95–108.
[ 4 ] P. Masani and N. Wiener, Non-linear prediction, Probability and Statistics, The Harald Cramér Volume, (ed. U. Grenander), John Wiley, 1959, 190–212.
[ 5 ] M. Matsuura and Y. Okabe, On a non-linear prediction problem for one-dimensional stochastic processes, Japan. J. Math., **27** (1) (2001), 51–112.
[ 6 ] M. Matsuura and Y. Okabe, On the theory of $KM_2O$-Langevin equations for non-stationary and degenerate flows, J. Math. Soc. Japan, **55** (2003), 523–563.
[ 7 ] M. N. A. Nxrgaard, N. K. Poulsen and O. Ravn, Advances in derivative-free state estimation for nonlinear systems, Technical report: IMM-REP-1998-15, Department of Mathematical Modeling, Technical University of Denmark, 1998.
[ 8 ] Y. Okabe, On a stochastic difference equation for the multi-dimensional weakly stationary process with discrete time, "Algebraic Analysis" in celebration of Professor M. Sato's sixtieth birthday, Prospect of Algebraic Analysis, (ed. M. Kashiwara and T. Kawai), Academic Press, 1988, 601–645.
[ 9 ] Y. Okabe and Y. Nakano, The theory of $KM_2O$-Langevin equations and its applications to data analysis (I): Stationary analysis, Hokkaido Math. J., **20** (1991), 45–90.
[10] Y. Okabe and A. Inoue, The theory of $KM_2O$-Langevin equations and its applications to data analysis (II): Causal analysis (1), Nagoya Math. J., **134** (1994), 1–28.
[11] Y. Okabe, Langevin equations and causal analysis, Amer. Math. Soc. Transl., **161** (1994), 19–50.
[12] Y. Okabe and T. Yamane, The theory of $KM_2O$-Langevin equations and its applications to data analysis (III): Deterministic analysis, Nagoya Math. J., **152** (1998), 175–201.
[13] Y. Okabe, On the theory of $KM_2O$-Langevin equations for stationary flows (1): characterization theorem, J. Math. Soc. Japan, **51** (1999), 817–841.
[14] Y. Okabe, On the theory of $KM_2O$-Langevin equations for stationary flows (2): construction theorem, Acta Appl. Math., **63** (2000), 307–322.
[15] Y. Okabe and M. Matsuura, On the theory of $KM_2O$-Langevin equations for stationary flows (3): extension theorem, Hokkaido Math. J., **29** (2000), 369–382.
[16] Y. Okabe and A. Kaneko, On a non-linear prediction problem for multi-dimensional stochastic processes with its applications to data analysis, Hokkaido Math. J., **29** (2000), 601–657.
[17] Y. Okabe, The Fluctuation-Dissipation Principle in Time Series Analysis and Experimental Mathematics, Japan Hyouronsha, 2002.
[18] T. Söderström, Discrete-time Stochastic Systems Estimation and Control, Adv. Textb. Control Signal Process., Springer, 2002.

Yasunori Okabe
Department of Mathematical Informatics
Graduate School of Information Science and Technology
University of Tokyo
Bunkyo-ku, Tokyo, 113-8656
Japan
E-mail: okabe@mist.i.u-tokyo.ac.jp

Masaya Matsuura
Department of Mathematical Informatics
Graduate School of Information Science and Technology
University of Tokyo
Bunkyo-ku, Tokyo, 113-8656
Japan
E-mail: masaya@mist.i.u-tokyo.ac.jp