系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
5 (2) (1985), 94—106

THE LINEAR SYSTEMS LIE ALGEBRA, THE SEGAL-SHALE-WEIL REPRESENTATION AND ALL KALMAN-BUCY FILTERS

Michiel Hazewinkel
(*Center for Math. and Comp. Sci., The Netherlands*)

1. Introduction

Let ls_n be the Lie algebra of all differential operators in n variables with polynomial coefficients of total degree in variables and derivatives $\leqslant 2$. Thus e.g. ls_1 is the Lie algebra with basis

$$x^2,\ x\frac{\partial}{\partial x},\ \frac{\partial^2}{\partial x^2},\ x,\ \frac{\partial}{\partial x},\ 1 \tag{1.1}$$

(The product is of course the commutator product). The symbol ls for this Lie algebra stands for "linear systems". The reason for this appellation derives from the following. Consider a linear stochastic system

$$dx_t = Ax_t dt + Bdw_t,\quad dy_t = Cx_t + dv_t. \tag{1.2}$$

Then an unnormalized version of the density of the conditional expectation of the state x_t given the past observations y_s, $0 \leqslant s \leqslant t$, satisfies a (stochastic) evolution equation

$$d\rho(x,t) = L\rho(x,t)dt + L_1\rho(x,t)dy_{1t} + \cdots + L_p\rho(x,t)dy_{pt} \tag{1.3}$$

with $L, L_1, \cdots, L_p \in ls_n$. And for varying systems (1.2) these operators generate all of ls_n.

The Kalman-Bucy filter for $\hat{x}_t = E[x_t | y_s, 0 \leqslant s \leqslant t]$ is a system of the form

$$dz = \alpha(z)dt + \beta_1(z)dy_{1t} + \cdots + \beta_p(z)dy_{pt} \tag{1.4}$$

where z is short for $(P, \hat{x})$ and $\alpha, \beta_1, \cdots, \beta_p$ are vectorfields on $(P, \hat{x})$-space. Let $V(\underline{\underline{R}}^N)$ denote the Lie algebra of vectorfields on $\underline{\underline{R}}^N$. Then the first main point of this paper is that all Kalman-Bucy filters combine to define a "universal Kalman-Bucy filter" in the shape of an anti-homomorphism of Lie algebras

$$\kappa: ls_n \to V(\underline{\underline{R}}^N),\ N = \frac{1}{2}n(n+1) + n \tag{1.5}$$

(and it is even possible to use this to propagate nongaussian initial densities). Here "anti" means that

$$\kappa[D, D'] = [\kappa(D'), \kappa(D)] \text{ rather than } \kappa[D, D'] = [\kappa(D), \kappa(D')].$$

This also establishes that the Kalman filter does indeed define an antihomomorphism of Lie algebras

Received June 6, 1984.

from the Lie algebra generated by L, L_1, $\cdots$, L_p in (1.3) (the so-called estimation Lie algebra) to a suitable Lie algebra of vectorfields, as it should according to a philosophy (almost a theorem now) first proposed by Brockett and Clark [1].

The structure of ls_n is simple. It is an extension of the real symplectic Lie algebra Sp_n by the Heisenberg Lie algebra h_n. Let Sp_n be the symplectic Lie group. Then there is a famous and somewhat mysterious representation of Sp_n (or more precisely its 2-fold covering $\widetilde{S}p_n$) which turns up in many distinct areas of mathematics, e.g. number theory and quantum mechanics. It is called the Segal-Shale-Weil representation or sometimes the oscillator representation. The second main point of this paper is that this Segal-Shale-Weil representation and the "filter anti-representation" (1.5) above are intimately related. This extends and strengthens the links between filtering theory and quantum mechanics which had been noted before [11], cf. also various contributions in [5].

It seems likely that the fact that all Kalman-Bucy filters fit together nicely will be useful both for theory and applications. In fact it is definitely of importance in a class of nonlinear filtering problems coming from identification and tracking [4, 10] where the estimation Lie algebra is a lways a subalgebra of a current algebra $ls_n \otimes R$ where R is a ring of polynomials. Further applications of the "universal filter" (1.5) and/or its relations with the Segal-Shale-Weil representation seem likely.

2. The Linear Systems Lie Algebra ls_n

2.1. *Definition of* ls_n. Let $n \in N$. If α is a multi-index $\alpha = (\alpha_1, \cdots, \alpha_n)$, $\alpha_i \in \underline{N} \cup \{0\}$, then $|\alpha|$ denotes $\alpha_1 + \cdots + \alpha_n$ and we write

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \quad \partial_\beta = \frac{\partial^\beta}{\partial x^\beta} = \frac{\partial^{\beta_1}}{\partial x_1^{\beta_1}} \cdots \frac{\partial^{\beta_n}}{\partial x_n^{\beta_n}}.$$

With these notations ls_n is by definition the Lie algebra of all differential operators of total degree $\leqslant 2$, i. e. all differential operators $\Sigma c_{\alpha,\beta} x^\alpha \partial_\beta$ with $c_{\alpha,\beta} = 0$ unless $|\alpha| + |\beta| \leqslant 2$. These operators are considered to act on some suitable space of (real or complex valued) smooth functions on $\mathbb{R}^n$, say the Schwartz space $S(\mathbb{R}^n)$ of rapidly decreasing smooth functions on $\mathbb{R}^n$. The product (Lie bracket) of D_1, D_2 is then of course given by the commutator

$$[D_1, D_2](\phi) = D_1(D_2\phi) - D_2(D_1\phi), \quad \phi \in S(\mathbb{R}^n).$$

It is an elementary observation that ls_n is closed under this commutator product.

I shall call ls_n the *linear systems Lie algebra*. The reason for this name will become clear later (in section 4 below).

2.2. *The Heisenberg Lie algebra* h_n. Let h_n be the subspace of ls_n spanned by the operators of total degree $\leqslant 1$, i.e. the operators $x_1, \cdots, x_n; \partial_1, \cdots, \partial_n; 1$ (with an obvious notation). The products in h_n are of course the Heisenberg commutation relations

$$[\partial_i, x_j] = \delta_{ij}, \ [x_i, x_j] = [\partial_i, \partial_j] = [x_i, 1] = [\partial_i, 1] = 0 \tag{2.3}$$

where δ_{ij} is the Kronecker δ. The Lie algebra h_n is called the Heisenberg Lie algebra.

2.4. *The symplectic Lie algebra* sp_n. Let J be the $2n \times 2n$ matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

where I stands for the $n\times n$ unit matrix. The Lie algebra sp_n consists of all $2n \times 2n$ matrices M which satisfy $MJ + JM^T = 0$ (where M^T is the transpose of M). The product on sp_n is the commutator matrix product $[M, M'] = MM' - M'M$.

2.5. *Structure of ls_n.* It is an easy observation that $h_n \subset ls_n$ is an ideal, i.e. $[D, D'] \in h_n$ for all $D \in ls_n, D' \in h_n$. The quotient Lie algebra ls_n/h_n is isomorphic to sp_n. This can e. g. be seen as follows. Let $E_{i,j}$ denote the matrix with a 1 at spot (i,j) and 0 everywhere else. Then the homomorphism of vectorspaces defined by

$$x_i x_j \to E_{i,n+j} + E_{j,n+i}, \qquad i, j = 1, \cdots, n,$$

$$x_i \frac{\partial}{\partial x_j} \to E_{i,j} - E_{n+j,n+i}, \qquad i, j = 1, \cdots, n,$$

$$\frac{\partial^2}{\partial x_i \partial x_j} \to E_{n+i,j} - E_{n+j,i}, \qquad i, j = 1, \cdots, n,$$

$$h_n \to 0,$$

is a surjective homomorphism of Lie algebras as is easily checked and induces an isomorphism $ls_n/h_n \simeq sp_n$. Thus we have an exact sequence

$$0 \to h_n \xrightarrow{l} ls_n \xrightarrow{\pi} sp_n \to 0. \tag{2.6}$$

A lift of π (i. e. a homomorphism of Lie algebras $\sigma\colon sp_n \to ls_n$ such that $\pi\circ\sigma = \mathrm{id}$) is given by $\sigma(E_{i,n+j} + E_{j,n+i}) = x_i x_j$, $\sigma(E_{n+i,j} - E_{n+j,i}) = -\dfrac{\partial^2}{\partial x_i \partial x_j}$, $\sigma(E_{i,j} - E_{n+j,n+i}) = x_i \dfrac{\partial}{\partial x_j} + \dfrac{1}{2}\delta_{ij}$. This defines an action of sp_n on h_n and also on $h_n/Z \simeq \mathbb{R}^{2n}$ (as an abelian Lie algebra) where Z is the one dimensional centre of h_n and ls_n. Identifying $\mathbb{R}^{2n}$ with h_n/Z by means of $e_i \to x_i$, $e_{n+i} \to -\partial_i$, $i = 1, \cdots, n$, this action becomes the usual action of sp_n as a Lie algebra of $2n \times 2n$ matrices on $\mathbb{R}^{2n}$.

3. The Filter Anti-Representation of ls_n

3.1. *Description of the anti-representation.* If M is a smooth manifold, $F(M)$ denotes the smooth functions on M and $V(M)$ denotes the Lie algebra of vectorfields on M (considered as the Lie algebra of derivations $F(M) \to F(M)$). If $M = \mathbb{R}^n$ then in the coordinates $(x_1, \cdots, x_n)$ every vectorfield on $\mathbb{R}^n$ can be written as $\sum_i f_i(x) \dfrac{\partial}{\partial x_i}$, where the $f_i(x)$ are smooth functions.

Now consider $\mathbb{R}^N$ with $N = \dfrac{1}{2}n(n+1) + n + 1$ with coordinates $P_{ij} = P_{ji}, i, j = 1, \cdots, n; m_i, i = 1, \cdots, n$. Consider the homomorphism of real vectorspaces

$$\kappa\colon ls_n \to V(\mathbb{R}^N) \tag{3.2}$$

defined by the formulas

$$1 \to \frac{\partial}{\partial c}, \tag{3.3}$$

$$x_i \to m_i \frac{\partial}{\partial c} + \sum_{t=1}^{n} P_{it} \frac{\partial}{\partial m_t}, \tag{3.4}$$

$$\frac{\partial}{\partial x_i} \to -\frac{\partial}{\partial m_i}, \tag{3.5}$$

$$x_i x_j \to (m_i m_j + P_{ij}) \frac{\partial}{\partial c} + \sum_t (m_i P_{jt} + m_j P_{it}) \frac{\partial}{\partial m_t}$$
$$+ \sum_{s,t} P_{is} P_{jt} \frac{\partial}{\partial P_{st}} + \sum_t P_{it} P_{jt} \frac{\partial}{\partial P_{tt}}, \tag{3.6}$$

$$x_i \frac{\partial}{\partial x_j} \to -m_i \frac{\partial}{\partial m_j} - \delta_{ij} \frac{\partial}{\partial c} - P_{ij} \frac{\partial}{\partial P_{ij}} - \sum_t P_{it} \frac{\partial}{\partial P_{jt}}, \tag{3.7}$$

$$\frac{\partial^2}{\partial x_i \partial x_j} \to \frac{\partial}{\partial P_{ij}} \text{ if } i \neq j, \quad \frac{\partial^2}{\partial x_i^2} \to 2 \frac{\partial}{\partial P_{ii}}. \tag{3.8}$$

3.9. **Theorem.** *The vectorspace homomorphism $\kappa: ls_n \to V(\mathbf{R}^N)$ defined by the formulae (3.3)—(3.8) is an injective anti-homomorphism of Lie algebras (i.e. it satisfies $\kappa[D, D'] = [\kappa(D'), \kappa(D)]$ for all $D, D' \in ls_n$).*

The proof of this theorem is a straightforward but perhaps somewhat tedious calculation. As an example we have $[\partial_i, x_i] = 1$ and

$$\left[m_i \frac{\partial}{\partial c} + \sum_t P_{it} \frac{\partial}{\partial m_t}, -\frac{\partial}{\partial m_i}\right] = \frac{\partial}{\partial c}$$

which fits. As another example if $i \neq j$ we have

$$\left[\frac{\partial^2}{\partial x_i \partial x_j}, x_i x_j\right] = x_i \frac{\partial}{\partial x_i} + x_j \frac{\partial}{\partial x_j} + 1.$$

Now

$$\left[\frac{\partial}{\partial P_{ij}}, (m_i m_j + P_{ij}) \frac{\partial}{\partial c}\right] = \frac{\partial}{\partial c},$$
$$\left[\frac{\partial}{\partial P_{ij}}, \sum_t (m_i P_{jt} + m_j P_{it}) \frac{\partial}{\partial m_t}\right] = m_i \frac{\partial}{\partial m_i} + m_j \frac{\partial}{\partial m_j},$$
$$\left[\frac{\partial}{\partial P_{ij}}, \sum_{s,t} P_{is} P_{jt} \frac{\partial}{\partial P_{s,t}}\right] = \sum_t P_{jt} \frac{\partial}{\partial P_{jt}} + \sum_s P_{is} \frac{\partial}{\partial P_{si}},$$
$$\left[\frac{\partial}{\partial P_{ij}}, \sum_t P_{it} P_{jt} \frac{\partial}{\partial P_{tt}}\right] = P_{jj} \frac{\partial}{\partial P_{jj}} + P_{ii} \frac{\partial}{\partial P_{ii}}.$$

So indeed

$$\left[\kappa(x_i x_j), \kappa\left(\frac{\partial^2}{\partial x_i \partial x_j}\right)\right] = -\frac{\partial}{\partial c} - m_i \frac{\partial}{\partial m_i} - m_j \frac{\partial}{\partial m_j} - \sum_t P_{jt} \frac{\partial}{\partial P_{jt}}$$
$$- \sum_s P_{is} \frac{\partial}{\partial P_{is}} - P_{jj} \frac{\partial}{\partial P_{jj}} - P_{ii} \frac{\partial}{\partial P_{ii}}$$
$$= \kappa\left(x_i \frac{\partial}{\partial x_i}\right) + \kappa\left(x_j \frac{\partial}{\partial x_j}\right) \kappa(1).$$

The remaining identities are checked similarly.

3.10. **Remark.**

$$\frac{\partial}{\partial x_i} \to -\frac{\partial}{\partial x_i}, \quad x_i \frac{\partial}{\partial x_j} \to -x_j \frac{\partial}{\partial x_i}, \quad x_i x_j \to x_i x_j, \quad \frac{\partial^2}{\partial x_i \partial x_j} \to \frac{\partial^2}{\partial x_i \partial x_j},$$

$x_i \to x_i, 1 \to 1$ defines an anti-automorphism of ls_n. Thus changing the sign in formulas (3.5) and (3.7) defines a representation of ls_n in $V(\mathbf{R}^N)$.

4. DMZ Equations and Kalman Filters

4.1. *The Duncan-Mortenson-Zakai equation and the estimation Lie algebra.* Consider a general nonlinear stochastic system (in Ito form)

$$dx_t = f(x_t)dt + G(x_t)dw_t, \quad dy_t = h(x_t)\,dt + dv_t, \quad x_t \in \underline{\underline{R}}^n, \tag{4.2}$$

where f, G, h are suitable vector and matrix valued functions and w_t and v_t are independent unit covariance Wiener processes also independent of the initial random vector x_0. Given sufficiently nice f, G, h, an unnormalized version $\rho(x, t)$ of the probability density $p(x,t)$ of the state x_t given the past observations $y_s, 0 \leqslant s \leqslant t$, satisfies the (forced) diffusion equation (Fisk-Stratonovič form)

$$d\rho = L\rho dt + \sum_{j=1}^{p} h_j \rho dy_t, \tag{4.3}$$

where h_j is the j-th component of h and L is the second order differential operator

$$L\phi = \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}((GG^T)_{ij}\phi) - \sum_i \frac{\partial}{\partial x_i}(f_i\phi) - \frac{1}{2}\sum_i h_i^2\phi. \tag{4.4}$$

Here f_i is the i-th component of f and $(GG^T)_{ij}$ the (i,j)-entry of the matrix product GG^T. Equation (4.3) is called the Duncan-Mortensen-Zakai equation. Cf. e.g. [3] for a derivation. The Lie algebra of differential operators (on $S(\mathbf{R}^n)$ say) generated by L and $h_1, h_2, \cdots, h_p$ is called the estimation Lie algebra.

4.5. *Exact filters and Liealgebra anti-homomorphisms.* Now let

$$d\xi_t = \alpha(\xi_t)dt + \beta_1(\xi_t)dy_{1t} + \cdots + \beta_p(\xi_t)dy_{pt}, \ \hat{x}_t = \gamma(x_t) \tag{4.6}$$

be a stochastic system (in Fisk-Stratonovič form) driven by y_t which calculates the conditional expectation

$$\hat{x}_t = E[x_t | y_s, \ 0 \leqslant s \leqslant t] \tag{4.7}$$

of the state given the past observations. I. e. (4.6) is a filter for $\hat{x}_t$. Then as Brockett and Clark observed [1] we have two ways of calculating $\hat{x}_t$, one via (4.3) and one via (4.6). Minimal realization theory then suggests that there will be a corresponding homomorphism of Lie algebras from the estimation Lie algebra L of the system to the Lie algebra of vectorfields generated by the vectorfields $\alpha, \beta_1, \cdots, \beta_p$ in (4.6) given by $A \to \alpha, h_i \to \beta_i, i = 1, \cdots, p$. This is called the Brockett-Clark homomorphism principle. In [1] this was verified to be indeed the case for the case of the Kalman filter of one of the simplest possible linear systems, namely $dx_t = dw_t, dy_t = x_t dt + dv_t$.

As a matter of fact a filter like (4.6) for $\hat{x}_t$ (or for some other statistic) should give rise to an anti-homomorphism from the Lie algebra of differential operators to the Lie algebra of vectorfields generated by the vectorfields in the filter. The reason is that $A\rho$ and $h_j\rho$ in (4.3) must be interpreted as vectorfields on $S(\mathbf{R}^n)$ and the mapping which assigns to a linear operator the corresponding linear vectorfield is an injective anti-homomorphism of Lie algebras.

4.8. *The Kalman-Bucy filter.* Now consider an n-dimensional linear system with m inputs and p outputs

$$(\Sigma) \quad dx_t = Ax_t dt + B dw_t, \quad dy_t = Cx_t dt + dv_t. \tag{4.9}$$

The Kalman-Bucy filter for $\hat{x}_t$ is given by the equations

$$d\hat{x}_t = A\hat{x}_t dt + P_t C^T(dy_t - C\hat{x}_t dt), \tag{4.10}$$

$$dP_t = (AP_t + P_t A^T + BB^T - P_t C^T C P_t)dt. \tag{4.11}$$

Write m_t for $\hat{x}_i$ and $P_t = (P_{ij})$. Then the part of the right hand side of (4.10) involving dy_{kt} and contributing to dm_{it} is equal to

$$\sum_j P_{ij} c_{kj} dy_{kt}.$$

It follows that if we write (4.10), (4.11) in the form (4.6) then the vectorfields $\beta_1, \cdots, \beta_p$ are equal to

$$\beta_k = \sum_{r,s} P_{rs} c_{ks} \frac{\partial}{\partial m_r}, \qquad k = 1, \cdots, p. \tag{4.12}$$

Similarly the α vectorfield of (4.10)—(4.11) is equal to

$$\begin{aligned}\alpha = &\sum_{i,j} a_{ij} m_j \frac{\partial}{\partial m_i} - \sum_{i,j,r,s} P_{ij} c_{rj} c_{rs} m_s \frac{\partial}{\partial m_i} \\ &+ \sum_{r,i\leqslant j} a_{ir} P_{rj} \frac{\partial}{\partial P_{ij}} + \sum_{r,i\leqslant j} P_{ir} a_{jr} \frac{\partial}{\partial P_{ij}} + \sum_{r,i\leqslant j} b_{ir} b_{jr} \frac{\partial}{\partial P_{ij}} \\ &- \sum_{r,s,t,i\leqslant j} P_{ir} c_{sr} c_{st} P_{tj} \frac{\partial}{\partial P_{ij}}.\end{aligned} \tag{4.13}$$

4.14. *Estimation Lie algebra and Kalman-Bucy filter.* Consider again the linear system (4.9). The operators which occur in the DMZ equation for this system are

$$h_i = \sum_r c_{ir} x_r, \tag{4.15}$$

$$L = \frac{1}{2} \sum_{i,j,r} b_{ir} b_{jr} \frac{\partial^2}{\partial x_i \partial x_j} - \sum_{r,i} a_{ir} x_r \frac{\partial}{\partial x_i} - \frac{1}{2} \sum_{i,j,r} c_{ri} c_{rj} x_i x_j - \sum_i a_{ii}. \tag{4.16}$$

Let $L(\Sigma)$ be the estimation Lie algebra of the linear system (4.9). This is obviously, cf. (4.16), a sub-Lie algebra of ls_n, and for varying Σ the various $L(\Sigma)$ generate all of ls. Whence the name "linear systems Lie algebra" for ls_n.

As in section 3 above let $N = \frac{1}{2} n(n+1) + n + 1$. Consider the projection $\mathbf{R}^N \to \mathbf{R}^{N-1}$ which maps (m, P, c) to (m, P). Under this projection the vectorfields occurring in the right hand sides of (3.3)—(3.8) map to vectorfields on $\mathbf{R}^{N-1}$. The vectorfields arising in this way are the same ones except that the $\frac{\partial}{\partial c}$ terms are removed. Let

$$\kappa' : ls_n \to V(\underline{\underline{R}}^{N-1}) \tag{4.17}$$

be the resulting anti-homomorphism of Lie algebras.

4.18. **Theorem.** *The restriction of κ' to $L(\Sigma)$ maps the operator L of (4.16) to the vectorfield α of (4.13) and the operators h_i of (4.15) to the vectorfields β_i of (4.12). In other words the restriction of κ' to $L(\Sigma) \subset ls_n$ is the Kalman-Bucy filter for the system (Σ).*

The proof of theorem 4.18 is an entirely straightforward verification, lightly complicated by the fact that $P_{ij}=P_{ji}$ must be taken into account which is not automatically done by the notation used. Thus the coefficient of $\frac{\partial^2}{\partial x_i \partial x_j}$ in L in (4.16) is equal to $\sum_r b_{ir}b_{jr}$ if $i \neq j$ and $\frac{1}{2}\Sigma b_{ir}^2$ if $i=j$, and under κ' which takes

$$\frac{\partial^2}{\partial x_i \partial x_j} \to \frac{\partial}{\partial P_{ij}}, \quad \frac{\partial^2}{\partial x_i^2} \to 2\frac{\partial}{\partial P_{ii}},$$

this gives the fifth term of α in (4.13). Similarly the coefficient of $x_r \frac{\partial}{\partial x_i}$ in(4.16) is $-a_{ir}$. The morphism κ' takes

$$x_r \frac{\partial}{\partial x_i} \to -m_r \frac{\partial}{\partial m_i} - P_{ri}\frac{\partial}{\partial P_{ii}} - \sum_t P_{rt}\frac{\partial}{\partial P_{it}}$$

and these terms account for the first, third and fourth terms in (4.13). Finally the coefficient of x_ix_j in (4.16) is $-\sum_r c_{ri}c_{rj}$ if $i \neq j$ and $-\frac{1}{2}\sum c_{ri}^2$ if $i=j$. The morphism κ' takes x_ix_j into

$$\sum_t (m_iP_{jt}+m_jP_{it})\frac{\partial}{\partial m_t} + \sum_{s,t} P_{is}P_{jt}\frac{\partial}{\partial P_{st}} + \sum_t P_{it}P_{jt}\frac{\partial}{\partial P_{tt}}$$

and this accounts for the second and sixth terms in (4.13). Similarly (and rather easier) one checks that κ' takes the h_i of(4.15)into the β_i of (4.12). This proves that κ' indeed restricts to the Kalman-Bucy filter on $L(\Sigma)$.

4.19. **Remarks.** Another way to state theorem 4.18 is to say that all possible Kalman-Bucy filters combine to define an anti-representation of ls_n which is faithful modulo the one-dimensional centre. The lifted anti-representation κ is faithful on ls_n itself and permits us to propagate also nongaussian initial densities. Cf. also section 6 below.

As a corollary of theorem 4.18 we of course obtain that $L \to \alpha$, $h_i \to \beta_i$ (with L, α, h_i, β_i respectively given by (4.16), (4.13), (4.15), (4.12)) does indeed define an anti-homomorphism of Lie algebras,as it should.

4.20. *Identification as a nonlinear filtering problem.* Consider a linear system

$$dx = Axdt + Bdw, \; dy = Cxdt + dv_t, \tag{4.21}$$

in which A, B, C are unknown. The problem is to find the best estimates of both x_t and the matrices A, B, C given $y_s, 0 \leqslant s \leqslant t$. By adding to (4.21) the state equations

$$dA = 0, \; dB = 0, \; dC = 0 \tag{4.22}$$

(so that A, B, C are viewed as random variables (constant in time)), we obtain a nonlinear system and the nonlinear filtering problem of finding the conditional expectation of the extended state (x, A, B, C) is the identification of linear system problem (or at least one version of it).

One potentially interesting statistic is the conditional expectation of the random variable $x_t|(A, B, C)$. The family of all Kalman filters (4.10)—(4.11) (for varying (A, B, C)) computes this. According to the Brockett-Clark anti-homomorphism principle, there should be a corresponding anti-homomorphism of Lie algebras. This is the morphism κ' when the image is viewed as vectorfields on $\mathbb{R}^{N-1}\times\mathbb{R}^{n^2+np+nm}$ with no $\frac{\partial}{\partial A}, \frac{\partial}{\partial B}$ or $\frac{\partial}{\partial C}$ terms. Thus the Brockett-Clark principle also holds in this case.

4.23. *Example*. For special linear systems $L\rightarrow\alpha, h_i\rightarrow\beta_i$ may accidentally also define a homomorphism of Lie algebras. This happens e. g. for all one-dimensional systems and all systems (4.9) for which the A matrix is zero. In general this is not the case as the following example shows

$$\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dt + \begin{pmatrix} 0 \\ 1 \end{pmatrix} dw,$$

$$\begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dt + \begin{pmatrix} dv_1 \\ dv_2 \end{pmatrix}.$$

5. The Segal-Shale-Weil Representation

This section simply lists some well-known facts on the basis of [7] with a few elaborations.

5.1. *The symplectic group.* Let J be as in 2.4 above. Then the symplectic group Sp_n consists of all real $2n\times 2n$ matrices M such that $MJM^T=J$. The Lie algebra of Sp_n is the Lie algebra sp_n which we encountered in section 2 above.

A certain representation of Sp_n or more precisely of its two-fold covering $\tilde{S}p_n$ on $L^2(\mathbb{R}^n)$ which is called the Segal-Shale-Weil representation is of considerable importance in several areas of mathematics, notably number theory [14] and quantum mechanics [12,13]. As we shall see it is also closely related to all Kalman-Bucy filters.

5.2. *Definition of the Segal-Shale-Weil representation.* One well-known way to obtain this representation is via the Stone-Von Neumann uniqueness theorem. Let H_n denote the Heisenberg group, $H_n=\mathbb{R}^n\times\mathbb{R}^n\times S^1$, where S^1 is the circle, with the multiplication $(x,y,z)(x',y',z')=(x+x',y+y',e^{-2\pi i\langle x,y'\rangle}zz')$. The Lie algebra of H_n is h_n (which we also encountered in section 2 above). This Lie algebra can also be described as $h_n=\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}$ and then the Lie bracket defines a bilinear form $\mathbb{R}^{2n}\times\mathbb{R}^{2n}\rightarrow\mathbb{R}$ which is given by the matrix J. Thus Sp_n can be seen as a group of automorphisms of h_n and H_n which moreover is the identity on the centre $S^1\subset H_n$.

One version of the Stone-Von Neumann theorem says that up to unitary equivalence there is a unique irreducible representation of H_n whose character on S^1 is the identity. Now let ρ be the standard (Schrödinger) representation of H_n in $L^2(\mathbb{R}^n)$ which is given by

$$(x,0,0)\rightarrow M_x, M_x f(x') = e^{2\pi i\langle x,x'\rangle}f(x'),$$
$$(0,y,0)\rightarrow T_y, T_y f(x') = f(x'-y),$$
$$(0,0,z)\rightarrow S_z, S_z f(x') = zf(x').$$

Now let $g\in Sp_n$ and consider Sp_n as a group of automorphisms of H_n. Then $h\rightarrow\rho(g(h))$ is also an irreducible representation of H_n with the same central character. By the uniqueness theorem there is an intertwining operator $\omega(g)$ such that $\omega(g)\rho(h)\omega(g)^{-1}=\rho(g(h))$. These $\omega(g)$ are unique up to a scalar factor. It remains to see whether these scalar factors can be fixed up to yield a representation of Sp_n on $L^2(\mathbb{R}^n)$ (instead of on $P(L^2(\mathbb{R}^n))$). This can almost be done and the result is the Segal-Shale-Weil representation of the two-fold covering

$\tilde{S}p_n$ of Sp_n in $L^2(\mathbf{R}^n)$.

5.3. *More or less explicit description of the Segal-Shale-Weil representation.* Let

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp_n, \tag{5.4}$$

where A, B, C, D are $n \times n$ matrices. Then the A, B, C, D satisfy $AB^T = BA^T$, $CD^T = DC^T$, $AD^T - BC^T = I$. Important special elements in Sp_n are

$$\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \begin{pmatrix} A & 0 \\ 0 & (A^{-1})^T \end{pmatrix}, \begin{pmatrix} I & N \\ 0 & I \end{pmatrix}, N \text{ symmetric} \tag{5.5}$$

and it is not especially difficult to show that these generate all of Sp_n. Thus in principle to describe the Segal-Shale-Weil representation it suffices to describe the unitary operators corresponding to these matrices. These are as follows

$$\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \to \text{ Fourier transform } F: L^2(\underline{R}^n) \to L^2(\underline{R}^n), \tag{5.6}$$

$$\begin{pmatrix} A & 0 \\ 0 & (A^{-1})^T \end{pmatrix} \to (f(x) \to |\det A|^{\frac{1}{2}} f(A^T x)), \tag{5.7}$$

$$\begin{pmatrix} I & N \\ 0 & I \end{pmatrix} \to (f(x) \to e^{\pi i N(x)} f(x)), \tag{5.8}$$

where $N(x)$ is the quadratic form defined by the symmetric matrix N.

5.9. *The Lie algebra representation defined by the Segal-Shale-Weil representation.* First consider a symmetric matrix $N = (n_{ij})$. Then

$$\frac{d}{dt}\left(\begin{pmatrix} I & 0 \\ tN & I \end{pmatrix} f(x)\right)\Big|_{t=0} = (\pi i N(x))(f(x)).$$

Next let B be an $n \times n$ matrix, $A = e^{tB}$. Then

$$\frac{d}{dt}\begin{pmatrix} e^{tB} & 0 \\ 0 & (e^{-tB})^T \end{pmatrix}(f)\Big|_{t=0} = \left(+\frac{1}{2}\operatorname{Tr}(B) + \sum_i (B^T x)_i \frac{\partial}{\partial x_i}\right)(f).$$

Finally consider the one-parameter subgroup

$$S_t = \begin{pmatrix} I\cos t & I\sin t \\ -I\sin t & I\cos t \end{pmatrix}$$

of Sp_n whose tangent vector at $t = 0$ is J (and which also passes through J). Writing

$$S_t = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\begin{pmatrix} I\cos^{-1} t & 0 \\ 0 & I\cos t \end{pmatrix}\begin{pmatrix} I & I\sin t\cos t \\ I & 0 \end{pmatrix}$$
$$\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}\begin{pmatrix} I & I\operatorname{tg} t \\ 0 & I \end{pmatrix}$$

it is not difficult to write down $(S_t f(x)$ and to calculate the derivative at $t = 0$. The result is

$$\pi i(x_1^2 + \cdots + x_n^2) - \frac{i}{4\pi}\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\right).$$

It readily follows that the Lie algebra of operators arising from the Segal-Shale-Weil representation is the one with basis

$$\pi i x_k x_j, \ \frac{i}{4\pi}\frac{\partial^2}{\partial x_k \partial x_j}, \ x_k\frac{\partial}{\partial x_j}+\frac{1}{2}\delta_{kj},$$

which is of course isomorphic to sp_n, for example to the incarnation of sp_n as the subalgebra $\sigma(sp_n)\subset ls_n$ via the isomorphism induced by the coordinate change $x_k \to (\sqrt{\pi i})x_k$.

6. Kalman-Bucy Filters and the Segal-Shale-Weil Representation

6.1. *Outline of the connection*. Given that the Kalman-Bucy filters combine to give an anti-representation of $sp_n \subset ls_n$ with sp_n realized as a Lie algebra of differential operators and that the differentiated version of the Segal-Shale-Weil representation is also a representation of this same Lie algebra of differential operators, it would be odd if they were not rather closely related. Indeed, as the attentive reader will have seen coming, the filter anti-representation is essentially a real and local version of the Segal-Shale-Weil representation.

The connection is essentially given by assigning to a pair (P,m), $m\in \mathbf{R}^n$, P a symmetric positive definite matrix, the corresponding normal density

$$\frac{1}{\sqrt{(2\pi)^n|P|}}\, e^{-\frac{1}{2}P^{-1}(x-m)}, \tag{6.2}$$

where $|P|$ is the absolute value of the determinant of P and $P^{-1}(y)$ is the quadratic form defined by P^{-1}. These functions form a total system in $L^2(\mathbf{R}^n)$ meaning that the finite linear combinations are dense, so that to define a representation of say Sp_n of $L^2(\mathbf{R}^n)$ it suffices to know what the representation does on these special functions. For the Segal-Shale-Weil representation one uses more generally $n\times n$ matrices Q whose real part is positive definite. And in fact it seems that Weil originally constructed his representation essentially in this way (cf. his comments, also referenced under [14], on the paper in question).

To spell things out in more detail and to avoid equations in P^{-1} (and calculating trouble) it is useful to use the Fourier transform.

6.3. *Some Fourier transform facts*. Let F denote the Fourier transform. Then we need the following more or less well-known facts

$$F = \frac{\partial}{\partial x_k} = 2\pi i x_k F, \quad F x_k = -\frac{1}{2\pi i}\frac{\partial}{\partial x_k}F,$$

where e.g. Fx_k stands for the composition of the operator "multiplication with x_k" with the operator F. The second fact we need is the formula

$$F^{-1}\left(\frac{1}{\sqrt{(2\pi)^n|P|}}\, e^{-\frac{1}{2}P^{-1}(x-m)+c}\right) = e^{c+\langle 2\pi i m, x\rangle - 2\pi^2 P(x)}. \tag{6.5}$$

Finally it is useful to note that the set of all functions of the form

$$p(x)e^{-Q(x)}, \tag{6.6}$$

where $p(x)$ is a (complex) polynomial and Q a (complex) polynomial of degree 2 whose real homogeneous part of degree 2 is positive definite, is stable under the Fourier transform, multiplication with polynomials and partial differentiation with respect to x_k.

6.7. *Obtaining the filter anti-representation.* Consider a function of the type $e^{c+\langle 2\pi i m, x\rangle}$ (m and P real). Imagine that m and P vary with time and try to see what this invo an evolution equation of the type

$$\frac{\partial}{\partial t}\left(e^{c+\langle 2\pi i m,x\rangle-2\pi^2 P(x)}\right)=L\,e^{c+\langle 2\pi i m,x\rangle-2\pi^2 P(x)}$$

where L is a differential operator from ls_n. As is easy to see this yields a system of o differential equations for m_k and P_{rs} provided that L is in ls_n. This idea is also due ckett. These first order differential equations which have polynomial right hand sides least locally uniquely solvable.

As one of the most complicated examples for $n = 2$ consider $\frac{\partial^2}{\partial x_1 \partial x_2} \in ls_n$. exp(—) for the e-power in (6.8) we find

$$\frac{\partial}{\partial t}\exp(-) = \exp(-)(\dot{c} + 2\pi i\dot{m}_1x_1+2\pi i\dot{m}_2x_2 - 2\pi^2(\dot{P}_{11}x_1^2 + 2\dot{P}_{12}x_1x_2 + \dot{P}_{22}x_2^2)),$$

$$\begin{aligned}\frac{\partial^2}{\partial x_1\partial x_2}\exp(-) = \exp(1)[&-4\pi^2 m_1m_2 - 8\pi^3 i m_2x_1P_{11} - 8\pi^3 i m_2x_2P_{12}\\ &- 8\pi^3 i m x_1x_2P_{22} + 16\pi^4x_1P_{11}x_2P_{22} + 16\pi^4x_2^2P_{12}P_{22} - 8\pi^3 i m_1x_1P_{12}\\ &+ 16\pi^4x_1^2P_{11}P_{12} + 16\pi^4x_1x_2P_{12}^2 - 4\pi^2P_{12}].\end{aligned}$$

Comparing these two expressions yields the differential equations

$$\begin{aligned}\dot{c} &= -4\pi^2m_1m_2 - 4\pi^2P_{12},\\ 2\pi i\dot{m}_1 &= -8\pi^3 i m_2P_{11} - 8\pi^3 i m_1P_{12},\\ 2\pi i\dot{m}_2 &= -8\pi^3 i m_2P_{12} - 8\pi^3 i m_1P_{22},\\ -2\pi^2\dot{P}_{11} &= 16\pi^4P_{11}P_{12},\\ -2\pi^2\dot{P}_{22} &= 16\pi^4P_{12}P_{22},\\ -4\pi^2\dot{P}_{12} &= 16\pi^4P_{11}P_{22} + 16\pi^4P_{12}^2.\end{aligned}$$

Writing down the associated vectorfield and using (6.4) and (6.5) the result is that th evolution of an unnormalized normal probability density $e^cN(m, P)$ with mean m and ance P in an evolution equation

$$\frac{\partial}{\partial t}\,e^cN(m,P) = x_1x_2e^cN(m, P)$$

is given by

$$\frac{\partial}{\partial t}(c, m, P) = \alpha(c, m, P), \qquad ($$

where α is the vectorfield

$$\begin{aligned}&(m_1m_2 + P_{12})\frac{\partial}{\partial c} + (m_2P_{11} + m_1P_{12})\frac{\partial}{\partial m_1} + (m_2P_{12} + m_1P_{22})\frac{\partial}{\partial m_2}\\ &+ 2P_{11}P_{12}\frac{\partial}{\partial P_{11}} + 2P_{12}P_{22}\frac{\partial}{\partial P_{22}} + (P_{11}P_{22} + P_{12}^2)\frac{\partial}{\partial P_{12}}, \qquad (\end{aligned}$$

which is of course the special case $n = 2$ of formula (3.6).

6.12. *Obtaining the Segal-Shale-Weil representation.* To obtain the Segal-Shale-Weil representation one can proceed in almost precisely the same way. Now of course one admits complex m and P (with the real part of P positive definite) and one uses

$$i\frac{\partial}{\partial x_j \partial x_k},\ ix_jx_k \text{ instead of } x_jx_k,\ \frac{\partial^2}{\partial x_j \partial x_k}.$$

6.13. *Finite escape time.* The class of functions $e^{c+\langle 2\pi im, x\rangle - 2\pi^2 P(x)}$ is stable under Fourier transform, multiplication with $e^{iQ(x)}$, Q a real quadratic form and under $x \rightarrow Ax$, A invertible, i.e. they are stable under the transformations corresponding to the special elements (5.5) of Sp_n. As these elements generate Sp_n it follows that there will be no finite escape time phenomena for the equations of the Segal-Shale-Weil case analogous to (6.10).

In the real case, i. e. the Kalman-Bucy filter case this can not be guaranteed. Indeed finite escape time does occur (cf. also [9]) and it is easy to see why. In this case $\begin{pmatrix} I & N \\ 0 & I \end{pmatrix}$ acts on $f(x)$ by multiplication with $e^{N(x)}$ and depending on $f(x)$ this may or may not result in a function $e^{N(x)}f(x)$ which is not Fourier transformable.

Writing elements of Sp_n as products of the special elements (5.5) gives more or less explicit solutions of Riccati equations for elements not too far from the identity and this also gives a good deal of information about in what directions (of sp_n or ls_n) finite escape time phenomena do not occur. Of course the one parameter subgroups of LS_n (the Lie group of ls_n) involve many more directions than those defined by "classically" studied Riccati equations. For complex linear systems the "Sp_n representation directions" are such that no finite escape time occurs either backwards or forwards. I do not know if this has system theoretic implications.

References

[1] Brockett, R. W., Clark, J. M. C., The geometry of the conditional density equations, Proc. of the Oxford on Stochastic Systems, Oxford, 1978.

[2] Brockett, R. W., Geometrical methods in stochastic control and estimation, In: M. Hazewinkel, J. C. Willems (eds), Stochastic Systems: the mathematics of filtering and identification and applications, Reidel Publ. Cy., 1981, 441—478.

[3] Davis, M. H. A., Marcus, S. I., An introduction to nonlinear filrering, In: Hazewinkel, M., Willems J. C., (eds), Stochastic systems: the mathematics of filtering and identification and applications, Reidel Publ. Cy., 1981, 53—76.

[4] Honzon, B., Hazewinkel, M., On identification of linear system and the estimation Lie algebra of the associated nonlinear filtering problem. In: Proc. 6-th IFAC Conf. on identifications, Washington, June 1982, Pergamon Press, 1983, 63—68.

[5] Hazewinkel, M., Willems, J. C. (eds), Stochastic systems: the mathematics of filtering and identification and applications, Reidel Publ. Cy., 1981.

[6] Hazewinkel, M., Marcus, S. I., On Lie algebras and nonlinear filtering, Report 8019, Econometric Inst., Erasmus Univ. Rotterdam, 1980. To appear in Stochastics.

[6] Howe, R. E., On the role of the Heisenberg group in harmonic analysis, *Aull. Amer. Math. Soc.*, 3 (1980), 821—844.

[8] Jazwinsky, A. H., Stochastic processes and filtering theory, Acad. Pr., 1970.

[9] Kalman, R. E., Contributions to the theory of optimal control, Bol. Soc. Mat. Mexicana 1960, 102—119.

[10] Krishnaprasad, P. S., Marcus, S. I., Identification and tracking: a class of nonlinear filtering problems, In: Proc. JACC, June 1981.

[11] Mitter, S. K., On the analogy between mathematical problems of nonlinear filtering and quantum

physics, *Ricerche di Automatica*, **10**: 2(1979), 163—216.

[12] Segal, I. E., Transforms for operators and symplectic automorphisms over a locally compact abelian group, *Math. Scand.*, **13**(1963), 31—43.

[13] Shale, D., Linear symmetrics of free boson fields, *Trans. Amer. Math., Soc.*, **103** (1962), 149—167.

[14] Weil, A., Sur certains groupes d'operateurs unitaires, In: Collected Papers Vol. III, 1—71, Springer 1980; Comments on this paper, ibid, 443—447.