A Scaling Theory for Linear Systems

ROGER W. BROCKETT, FELLOW, IEEE, AND P. S. KRISHNAPRASAD, MEMBER, IEEE

***Abstract*—In this paper we develop a theory of scaling for rational (transfer) functions in terms of transformation groups. In particular, we identify two different four-parameter scaling groups which play natural roles in studying linear systems and investigate the effect of scaling on Fisher information and related statistical measures in system identification. The scalings considered include change of time scale, feedback, exponential scaling, magnitude scaling, etc.**

The scaling action of the groups studied in this paper is tied to the geometry of transfer functions in a rather strong way as becomes apparent in our examination of the invariants of scaling. As a result, the scaling process also provides new insight into the parameterization question for rational functions.

I. Introduction

In the study of heat transfer, fluid mechanics, and various other aspects of applied physics, it is common to introduce dimensionless groupings of parameters which enter into the mathematical model and to use these numbers as an aid to understanding similitude and scaling. The Reynolds number of fluid mechanics is a very well-known example. In control theory, there has been as yet no focused attack on the question of what can be accomplished by scaling. Although specific questions regarding magnitude scaling, feedback, etc. are well-understood, the literature does not make clear what the relations are between the various types of scalings which one can use. The objective of this paper is to do so.

The scaling theory developed in this paper has two roles: firstly, in setting up geometric structure on rational functions and secondly, in creating a basis for analyzing identification algorithms. In the first part of the paper we define and analyze the basic scalings. We have also collected together some of the basic topological facts from [1]. In the second part of the paper we introduce the notion of invariant inference and apply some of the scaling methods to the identification problem. The scaling theory has its origin in [1] and has an implicit role in [2]. Some of the results here have appeared in a preliminary form in [3].

Manuscript received January 22, 1979; revised June 20, 1979. Paper recommended by B. Francis, Chairman of the Linear Systems Committee. This work was supported in part by the U.S. Office of Naval Research, Joint Services Electronics Program, under Contract N00014-75-C-0648, in part by the Army Research Office under Grant DAAG29-76-C-0139, and in part by NASA under NSG-2265.

R. W. Brockett is with the Division of Applied Sciences, Harvard University, Cambridge, MA 02178.

P. S. Krishnaprasad is with the Department of Systems and Computer Engineering, Case Western Reserve University, Cleveland, OH 44106.

II. Scaling on Rational Functions

The term scaling is used in a variety of ways in the physical sciences and engineering. We make this notion precise as follows. (We treat [4] and [5] as basic references for the differential geometric terms and results used here.)

1) Let the objects to be scaled be given the structure of a differentiable manifold M.

2) Let G be a Lie group.

Then, a scaling of M by G is a smooth action ϕ of G on M, i.e., a smooth map

$$\phi: G \times M \to M$$
$$(g,x) \to gx$$

satisfying the following.

a) For any $g \in G$, the map

$$\zeta_g: M \to M$$
$$x \to gx$$

is a diffeomorphism.

b) For each $x \in M$ and $g_1, g_2 \in G$ the relation

$$(g_1 g_2)x = g_1(g_2 x) \quad \text{holds.}$$

c) For each $x \in M$, the relation $ex = x$ holds, where e = identity of G.

We call any point $x_1 \in M$, a scaled version of $x_2 \in M$ if there exists some element $g \in G$ such that $\phi(g, x_1) = x_2$. In other words, the action of G on M partitions M into equivalence classes (orbits), points on the same orbit being thought of as scaled versions of each other. In order for the scaling operation to be useful, the orbit space should have sufficient structure and this in turn depends on the choice of the group G and the action ϕ.

In the present context, the differentiable manifold to be scaled is Rat(n), defined as follows. Consider the set of rational functions of the form $g(s) = q(s)/p(s)$, where $q(s) = q_{n-1}s^{n-1} + q_{n-2}s^{n-2} + \cdots + q_1 s + q_0$ and $p(s) = s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0$ are relatively prime polynomials, as an open subspace of $\mathfrak{R}^{2n}$ as the coefficients (q_i, p_i) vary over the reals. This subspace together with the differentiable manifold structure from $\mathfrak{R}^{2n}$ is called Rat(n). There are several one-parameter scaling operations that can be applied to rational transfer functions in order to bring the numerical values of the coefficients within a convenient numerical range. They appear naturally in input–ouput experiments, as a consequence of an experimenter's choice

0018-9286/80/0400-0197$00.75 ©1980 IEEE

of scales. We consider below two overlapping sets of scalings $\mathcal{Q}$ and $\mathcal{B}$ obtained from the basic scalings (here $g(s)\in \text{Rat}(n)$, $\mathcal{R}$ = real line, $\mathcal{R}^+$ = positive half-line).

1) $s\to\alpha s$, $\alpha\in\mathcal{R}^+$ (frequency scaling);
2) $s\to s+\sigma$, $\sigma\in\mathcal{R}$ (shift of imaginary axis);
3) $g(s)\to mg(s)$, $m\in\mathcal{R}^+$ (magnitude scaling);
4) $g(s)\to g(s)/(1+kg(s))$, $k\in\mathcal{R}$ (output feedback);
5) $g(s)=c(sI-A)^{-1}b\to c(sI-A)^{-1}e^{A\tau}b$, $\tau\in\mathcal{R}$.

The scalings are collected into sets $\mathcal{Q}$ and $\mathcal{B}$ as $\mathcal{Q}=\{1,2,3,4\}$ and $\mathcal{B}=\{1,2,3,5\}$. Even though item 5) looks as though it depends on the choice of representation $[A,b,c]$ and not just $g(s)$, it is in fact independent of the choice of representation since, in terms of the weighting pattern [Laplace transform inverse of $g(s)$], it corresponds to a shift of the origin of the time axis.

The first point to note is that the scalings in set $\mathcal{Q}$ and $\mathcal{B}$ generate, respectively, two four-parameter Lie groups $G_{\mathcal{Q}}$ and $G_{\mathcal{B}}$ that act on Rat(n). We will describe them explicitly in the next section. First we need a lemma.

Lemma 1: The one-parameter scalings 1)–5) defined above preserve the McMillan degree. Further, all except 5) leave the number of zeros unchanged.

Proof: The verification for the first three cases is quite elementary and we leave it to the reader.

Under output feedback, $q(s)\to q(s)$ and $p(s)\to p(s)+kq(s)$. In order for the McMillan degree to change, it is necessary for $p+kq$ and q to have a common zero, say z_i. But this means p and q have that same common zero which is a contradiction.

Lastly, let $[A,b,c]$ be a minimal realization of $w(t)$ and let $[A,e^{A\tau}b,c]$ be a realization of $w(t+\tau)$. Since $[A,b]$ is a cyclic pair, so is $[A,e^{A\tau}b]$. Hence, $[A,e^{A\tau}b,c]$ is also a minimal triple and the degree is preserved under shift. However, unlike the first four scalings, it is possible for one or many of the zeros to disappear into ∞ as can be seen from the following simple example. Let

$$g(s)=\frac{q_1s+q_0}{s^2}\in\text{Rat}(2).$$

(Thus, $q_0\neq 0$.)

We have a minimal realization

$$A=\begin{bmatrix}0&1\\0&0\end{bmatrix};\ b=\begin{bmatrix}0\\1\end{bmatrix};\ c=\begin{bmatrix}q_0 & q_1\end{bmatrix}.$$

Under the shift, we have the deformation

$$\frac{q_1s+q_0}{s^2}\to\frac{(q_0\tau+q_1)s+q_0}{s^2}.$$

Thus, for $\tau=-q_1/q_0$, the lone zero goes to ∞.

Remark 1: One of the reasons for dividing up the five scalings into two sets $\mathcal{Q}$ and $\mathcal{B}$ as above is that the scalings of set $\mathcal{Q}$ all act as linear transformations on the coefficients $(q_0,\cdots,q_{n-1},p_0,p_1,\cdots,p_{n-1})$ of the polynomials in the rational function representation of $g(s)$ while the scalings in set $\mathcal{B}$ act linearly on the (Markov) parameters h_i of the Laurent expansion

$$g(s)=\frac{h_0}{s}+\frac{h_1}{s^2}+\frac{h_2}{s^3}+\cdots$$

as will follow from the following calculations.

Remark 2: One may ask how the various one-parameter group actions can be combined. The answer to this depends on the calculation of certain Lie algebras in Section III. Before we close this section we recall one of the main results in [1] that we need.

Theorem 1: Rat(n) has $(n+1)$ connected components. The components are distinguished by the Cauchy index which takes values in the set $\{-n,-n+2,\cdots,n-2,n\}$.

Recall [7] that the Cauchy index of a rational function $g(s)$ is defined as

$$I_{-\infty}^{\infty}(g)=(\text{number of jumps of } g \text{ from } -\infty \text{ to} +\infty) - (\text{number of jumps of } g \text{ from } +\infty \text{ to } -\infty)$$

as s ranges over the reals from $-\infty$ to $+\infty$.

Denoting as Rat(p,q) the connected component of Rat$(p+q)$ of Cauchy index $(p-q)$, we see that Lemma 1 implies that the orbits of the scalings and hence the orbits of $G_{\mathcal{Q}}$ and $G_{\mathcal{B}}$ lie entirely within some Rat(p,q).

III. The Lie Algebras $\tilde{G}_{\mathcal{Q}}$ and $\tilde{G}_{\mathcal{B}}$

Recalling the notation of the previous section we have, for every smooth action ϕ by a group G on a differentiable manifold M, the following: to every $m\in M$, there corresponds a C^∞ map also denoted by m, of G into M defined as

$$m(g)=gm=\phi(g,m).$$

Let dm denote the corresponding derivative map from the tangent space at the identity (Lie algebra $\tilde{G}$) of G. Then we have a Lie algebra homomorphism,

$$\lambda:\ \tilde{G}\to\mathcal{V}(M)=\text{Lie algebra of smooth vector fields on } M$$

$$X\mapsto\bar{X}=\lambda X \text{ defined by } (\lambda X)(m)=dm(X(e)).$$

For the scalings of set $\mathcal{Q}$ we can establish these (linear) vector fields X_α (frequency scaling), X_σ (shift of imaginary axis), X_m (magnitude scaling), and X_k (output feedback) in the standard basis associated with the coordinates $(q_0,q_1,\cdots,q_{n-1},p_0,p_1,\cdots,p_{n-1})$ as

$$X_\alpha=-\sum_{i=0}^{n-1}(n-i)\left[q_i\frac{\partial}{\partial q_i}+p_i\frac{\partial}{\partial p_i}\right];$$

$$X_m=\sum_{i=0}^{n-1}q_i\frac{\partial}{\partial q_i}$$

$$X_\sigma=\sum_{i=0}^{n-2}(i+1)q_{i+1}\frac{\partial}{\partial q_i}+\sum_{i=0}^{n-2}(i+1)p_{i+1}\frac{\partial}{\partial p_i}+n\frac{\partial}{\partial p_{n-1}}$$

and

$$X_k=\sum_{i=0}^{n-1}q_i\frac{\partial}{\partial p_i}.$$

(Readers unfamiliar with this notation might want to look at Brockett [20, Appendix].) Here the $\{\partial/\partial q_i, \partial/\partial p_i\}$ are the standard basis vector fields for $\mathfrak{A}(\mathrm{Rat}(n))$. To combine the scalings of set $\mathcal{A}$, we need to compute the Lie algebra $\tilde{G}_{\mathcal{A}}$ generated by the vector fields X_α, X_m, X_σ, X_k. Given $X=\sum_i f_i \partial/\partial x^i$ and $Y=\sum_j g_j \partial/\partial x^j$, we define the Lie bracket of the vector fields X and Y as

$$[X,Y]=\sum_{i,j}\left(f_i\frac{\partial}{\partial x^i}g_j - g_i\frac{\partial}{\partial x^i}f_j\right)\frac{\partial}{\partial x^i}.$$

We then obtain the commutation rules,

$$[X_m,X_k]=X_k; \qquad [X_k,X_\sigma]=0$$
$$[X_m,X_\sigma]=0; \qquad [X_k,X_\alpha]=0$$
$$[X_m,X_\alpha]=0; \qquad [X_\alpha,X_\sigma]=X_\sigma.$$

Thus, the Lie algebras generated are $\{X_m,X_k\}_{\mathrm{LA}}\simeq \mathrm{af}(1)$ and $\{X_\alpha,X_\sigma\}_{\mathrm{LA}}\simeq \mathrm{af}(1)$ where af(1) is the affine Lie algebra on the real line. Recall that there is only one non-Abelian two-dimensional Lie algebra up to isomorphism [6]. Thus, $\tilde{G}_{\mathcal{A}}\equiv \mathrm{af}(1)\oplus\mathrm{af}(1)$, or the group $G_{\mathcal{A}}$ generated by the scalings of set $\mathcal{A}$, is locally isomorphic to a direct product of two copies of the affine group af(1). We have a matrix representation

$$G_{\mathcal{A}}\sim\begin{bmatrix}1&\sigma&0&0\\0&\alpha&0&0\\0&0&1&mk\\0&0&0&m\end{bmatrix}\qquad \alpha,m\in\mathfrak{R}^+;\ k,\sigma\in\mathfrak{R}$$

with the action

$$\phi_{\mathcal{A}}: G_{\mathcal{A}}\times \mathrm{Rat}(n)\to\mathrm{Rat}(n)$$
$$((\alpha,m,\sigma,k),g(s))\to\frac{mg(\alpha s+\sigma)}{1+kmg(\alpha s+\sigma)}.$$

Thus, orbits of $G_{\mathcal{A}}$ are four parameter families $F_g^{\mathcal{A}}$ of rational functions of degree n

$$F_g^{\mathcal{A}}=\left\{\frac{mg(\alpha s+\sigma)}{1+mk\,g(\alpha s+\sigma)};\begin{matrix}\alpha\in\mathfrak{R}^+, m\in\mathfrak{R}^+\\ k\in\mathfrak{R},\sigma\in\mathfrak{R}\end{matrix}\right\}.$$

Unlike the scalings of set $\mathcal{A}$, the scalings of set $\mathcal{B}$ do not divide up into commuting parts. To compute the Lie algebra $\tilde{G}_{\mathcal{B}}$ of the scalings of set $\mathcal{B}$, we find it convenient to work in the coordinates $(h_0,h_1,\cdots,h_{n-1},p_0,p_1,\cdots,p_{n-1})$ where $\{p_i\}$ are in the same as before and $\{h_i\}$ are the Markov parameters. Recall the useful identity [7, vol. 2, p. 234]

$$h_{n+j}=-\sum_{i=0}^{n-1}p_ih_{i+j},\qquad j=0,1,2,\cdots.$$

In these coordinates, the scaling vector fields of set $\mathcal{B}$ are

$$X_\alpha=\sum_{i=0}^{n-1}-\left[(i+1)h_i\frac{\partial}{\partial h_i}+(n+1-i)p_i\frac{\partial}{\partial p_i}\right]$$

$$X_\sigma=\sum_{i=0}^{n-1}ih_{i-1}\frac{\partial}{\partial h_i}+\sum_{i=0}^{n-2}-(i+1)p_{i+1}\frac{\partial}{\partial p_i}-n\frac{\partial}{\partial p_{n-1}}$$

$$X_m=\sum_{i=0}^{n-1}h_i\frac{\partial}{\partial h_i};$$

$$X_\tau=\sum_{i=0}^{n-2}h_{i+1}\frac{\partial}{\partial h_i}-\left(\sum_{i=0}^{n-1}p_ih_i\right)\frac{\partial}{\partial h_{n-1}}.$$

The commutation rules are then

$$[X_m,X_\tau]=0;\qquad [X_m,X_\sigma]=0;\qquad [X_m,X_\alpha]=0$$
$$[X_\alpha,X_\sigma]=-X_\sigma;\quad [X_\alpha,X_\tau]=-X_\tau;\quad [X_\tau,X_\sigma]=X_m.$$

It can be verified that the Lie algebra $\{X_\tau,X_m,X_\alpha,X_\sigma\}_{\mathrm{LA}}$ is isomorphic to the Lie algebra of matrices of the form

$$\begin{bmatrix}0&x_1&x_2\\0&x_3&x_4\\0&0&0\end{bmatrix}\qquad \text{all } x_i\in\mathfrak{R}$$

which we denote as $\tilde{G}_{\mathcal{B}}$. One action generated in this way is isomorphic to a four-parameter subgroup of $Gl(3;\mathfrak{R})$, the isomorphism given by

$$(\alpha,\sigma,m,\tau)\to\begin{bmatrix}1&\tau&\ln(m)\\0&\alpha&\sigma\\0&0&1\end{bmatrix}.$$

α, $m\in\mathfrak{R}^+$, and σ, τ, $\in\mathfrak{R}$. Thus, an orbit of $G_{\mathcal{B}}$ is a four-parameter family of weighting patterns, $F_w^{\mathcal{B}}=\{me^{\sigma t}w(\alpha t+\tau)\colon \alpha,m\in\mathfrak{R}^+,\sigma,\tau\in\mathfrak{R}\}$. In the following section we take up some geometric questions related to the scalings.

IV. Geometrical Aspects of the $G_{\mathcal{A}}$ and $G_{\mathcal{B}}$ Action

Although the one-parameter scalings themselves had no fixed points, it may well happen that for some $t\in G_{\mathcal{A}}$, $tx=x$ for some $x\in\mathrm{Rat}(n)$, and $t\neq$identity, i.e., the action may fail to be free. For the orbits of $G_{\mathcal{A}}$ to be integral submanifolds of Rat(n) of constant dimension, it is necessary and sufficient that

$$X_\alpha(p),\ X_m(p),\ X_\sigma(p),\text{ and } X_k(p)$$

form a set of independent vectors $\in T_x$ (Rat(n)) the tangent space to Rat(n) at x, for each $x\in\mathrm{Rat}(n)$. That this is not true for all of Rat(n) is verified by noting that $t=(\alpha,\alpha^n,0,0)\in G_{\mathcal{A}}$ leaves $g(s)=1/s^n$ fixed for any $\alpha>0$.

Let $T_{\mathcal{A}}$ denote the $2n\times 4$ matrix of tangent vectors $[X_m,X_k,X_\sigma,-X_\alpha]$ at some point of Rat(n) with coordinates $(q_0,q_1,\cdots,q_{n-1},P_0,\cdots,P_{n-1})$. We have in the standard basis

$$T_{\mathcal{G}}=\left[\begin{array}{cccc} q_0 & 0 & q_1 & nq_0 \\ q_1 & & 2q_2 & (n-1)q_1 \\ & \vdots & \vdots & \\ \vdots & & (n-1)q_{n-1} & \vdots \\ q_{n-1} & 0 & 0 & q_{n-1} \\ \hdashline 0 & q_0 & p_1 & np_0 \\ & & 2p_2 & (n-1)p_1 \\ \vdots & \vdots & \vdots & \\ & & (n-1)p_{n-1} & \vdots \\ 0 & q_{n-1} & n & p_{n-1} \end{array}\right].$$

The condition for free $G_{\mathcal{G}}$ action is that

$$\operatorname{Rank}(T_{\mathcal{G}})=4 \qquad (n>2).$$

Since at least for some $i\in\{0,1,\cdots,n-1\}$, $q_i\neq 0$, this condition is equivalent to the condition that

$$\tilde{T}_{\mathcal{G}}=\begin{bmatrix} q_0 & q_1 & nq_0 \\ q_1 & 2q_2 & (n-1)q_1 \\ \vdots & \vdots & \vdots \\ q_{n-2} & (n-1)q_{n-1} & 2q_{n-2} \\ q_{n-1} & 0 & q_{n-1} \end{bmatrix}$$

is of rank 3 $(n>2)$.

Now let $(q_0,q_1,\cdots,q_{n-1})=(0,\cdots,0,q_f,\cdots,q_l,0,\cdots,0)$ were $q_f=$first nonzero numerator coefficient and $q_l=$last nonzero numerator coefficient. The rank condition is guaranteed by the requirement that the first and last columns of $\tilde{\tilde{T}}_{\mathcal{G}}$ are independent, where

$$\tilde{\tilde{T}}_{\mathcal{G}}=\begin{bmatrix} q_f & (f+1)q_{f+1} & (n-f)q_f \\ q_{f+1} & & (n-f-1)q_{f+1} \\ \vdots & \vdots & \vdots \\ q_{l-1} & lq_l & (n-l+1)q_{l+1} \\ q_l & 0 & (n-l)q_l \end{bmatrix}$$

which reduces to the condition that

$$\det\begin{pmatrix} q_f & (n-f)q_f \\ q_l & (n-l)q_l \end{pmatrix}\neq 0, \qquad \text{i.e., } f\neq l.$$

Thus, if there are two or more distinct zeros for all members of a particular connected component of $\operatorname{Rat}(n)$, then the group $G_{\mathcal{G}}$ acts freely and effectively on that component. It is not hard to see from [1] that every member of $\operatorname{Rat}(p,q)$ for $|p-q|\geqslant 3$ has at least two distinct zeros. This condition can be weakened as follows.

1) If n is *odd*, the region with index$=\pm 1$ can have poles of odd multiplicity n. If so, all $p_i=0$ and even if there is one zero (say $q_1\neq 0$), the rank condition is not satisfied. However,

2) if n is *even*, in the region with Cauchy index$=\pm 2$, each rational function must have at least two distinct poles with one interlacing zero. This implies that *not* all $p_i=0$ and since there is at least one zero, say $q_1\neq 0$, the matrix $T_{\mathcal{G}}$ satisfies the rank condition. In the region with index$=0$, the presence of elements of the form $g(s)=1/s^n$, does not admit of free $G_{\mathcal{G}}$-action.

Summing up we have free $G_{\mathcal{G}}$-action on $\operatorname{Rat}(p,q)$ *iff* $|p-q|>1$. So far, our arguments have been specialized for the case $n\geqslant 3$, because of the way we passed from a rank condition for $T_{\mathcal{G}}$ to a rank condition for $\tilde{T}_{\mathcal{G}}$. But the conclusions are valid for the case $n=2$ and this can be verified in a direct and interesting way. Here

$$T_{\mathcal{G}}=\begin{bmatrix} q_0 & 0 & q_1 & 2q_0 \\ q_1 & 0 & 0 & q_1 \\ 0 & q_0 & p_1 & 2p_0 \\ 0 & q_1 & 2 & p_1 \end{bmatrix}.$$

Now $\det(T_{\mathcal{G}})=2q_1(q_0^2+q_1(p_0q_1-p_1q_0))$. On the other hand, for

$$g(s)=\frac{q(s)}{p(s)}=\frac{q_1s+q_0}{s^2+p_1s+p_0}\in\operatorname{Rat}(2)$$

the resultant [8] is given by

$$\operatorname{Res}(q,p)=\det\begin{bmatrix} 1 & p_1 & p_0 \\ 0 & q_1 & q_0 \\ q_1 & q_0 & 0 \end{bmatrix}$$
$$=-q_0^2-q_1(q_1p_0-q_0p_1)\neq 0$$

since $g(s)=q(s)/p(s)\in\operatorname{Rat}(2)$.

Thus, $\det(T_{\mathcal{G}})=-2q_1\operatorname{Res}(q,p)\neq 0$ *iff* $q_1\neq 0$. Then the $G_{\mathcal{G}}$-action is free only in the two components $\operatorname{Rat}(2,0)$ and $\operatorname{Rat}(0,2)$, where each element has at least one zero. In the case $n=1$, the dimensionality requirement rules out free $G_{\mathcal{G}}$-action.

We can examine the action of $G_{\mathcal{B}}$ in the same way as we did for $G_{\mathcal{G}}$ by imposing a rank condition on the matrix of tangent vectors $T_{\mathcal{B}}=[X_m,X_\tau,X_\sigma,-X_\alpha]_{(x)}$, $x\in\operatorname{Rat}(n)$. Surprisingly enough, the condition for free $G_{\mathcal{B}}$-action turns out to be the same as for $G_{\mathcal{G}}$. We omit the calculation (see [9] for details) except for noting that the multiplicities of the poles play roughly the same role as did the number of zeros in the case of $G_{\mathcal{G}}$-action. The case $n=2$ once again helps illustrate this. For this case,

$$\det(T_{\mathcal{B}})=(p_1^2-4p_0)\cdot(q_0^2+q_1(q_1p_0-q_0p_1))$$
$$=D(p)\cdot\operatorname{Res}(q,p)$$

where

$$T_{\mathcal{B}}=\left[\begin{array}{cccc} q_0 & -q_1p_0 & q_1 & 2q_0 \\ q_1 & q_0-q_1p_1 & 0 & q_1 \\ \hdashline 0 & 0 & p_1 & 2p_0 \\ 0 & 0 & 2 & p_1 \end{array}\right]$$

and $D(p)$ and $\mathrm{Res}(q,p)$ stand for the discriminant and the resultant of p and q/p, respectively. In $\mathrm{Rat}(2,0)$ and $\mathrm{Rat}(0,2)$ the poles are distinct and hence $D(p)\neq 0$. The resultant does not vanish of course. Thus, $G_{\mathfrak{B}}$ acts freely on these two components. In $\mathrm{Rat}(1,1)$ the $G_{\mathfrak{B}}$-action is not free because of the occurrence of repeated poles. We collect together the results of this section in the following.

Theorem 2: The groups $G_{\mathfrak{A}}$ and $G_{\mathfrak{B}}$ act freely (and hence effectively) on $\mathrm{Rat}(p,q)$ *iff* $|p-q|>1$.

In the remainder of this section, we show how this result is useful in understanding the geometry of $\mathrm{Rat}(p,q)$. Specifically, for $|p-q|>1$, $\mathrm{Rat}(p,q)$ is representable as the bundle space of two distinct principal bundles with structure groups $G_{\mathfrak{A}}$ and $G_{\mathfrak{B}}$. The concept of a fiber bundle has already begun to play a significant role in physics and system theory (Yang–Mills fields, canonical forms, etc.). Before we address the specific questions concerning $\mathrm{Rat}(p,q)$, it is helpful to define a few of the main concepts in the theory of fiber bundles. (Husemoller [5] and Nomizu [22] are standard references; Simms [10] has a compact exposition leading to quantization ideas.)

A bundle over a manifold is a triple (B,π,M) where B and M are smooth manifolds and $\pi:B\to M$ is a smooth projection onto M. The preimage $\pi^{-1}(x)=B_x$ is called the *fiber* over x. We say that (B,π,M) is a fiber bundle with fiber F if each fiber is a closed submanifold of B diffeomorphic to some fixed manifold F and the following local triviality condition holds.

For each $x\in M$ there is a neighborhood U and a diffeomorphism ϕ: $\pi^{-1}(U)\to U\times F$ such that the following diagram commutes:

$$\begin{array}{ccc} \pi^{-1}(U) & \stackrel{\phi}{\longrightarrow} & U\times F \\ \pi & \searrow \quad \swarrow & pr_1 \\ & U & \end{array}$$

Here pr_1 is the projection onto the first factor. The manifold B is called the bundle space of the fiber bundle.

Whereas this definition is quite general, for our purposes it is adequate to work with the special case of a *principal bundle*. Recall that a right action of a Lie group G on a smooth manifold B is simply a map ϕ: $B\times G\to B$ such that for any $g\in G$ the map R_g: $B\to B$ given by $p\to\phi(p,g)=p\cdot g$ is a diffeomorphism and we have the composition rule $R_{g_2}\circ R_{g_1}=R_{g_1g_2}$.

Suppose we are given smooth manifolds B and M and π: $B\to M$ a smooth map onto M. Let G be a Lie group acting on B to the right. Then (B,π,M,G) is a *principal G-bundle* (or B is a bundle over M with structure group G) if:

1) G acts freely on B, that is, $R_g(p)=p$ for some p, implies $g=e$ the identity;

2) M is the quotient space of B w.r.t. the equivalence relation induced by G. Thus, if $\pi(p_1)=\pi(p_2)$ for $p_1,p_2\in B$, then there is a $g\in G$ such that

$$R_g(p_1)=p_1\cdot g=p_2;$$

3) B is locally trivial over M; that is, every point $x\in M$ has a neighborhood U and a diffeomorphism ϕ: $\pi^{-1}(U)\to U\times G$ such that

$$\phi(p)=(\pi(p),\eta(p))$$

and

$$\eta(p\cdot g)=\eta(p)\cdot g.$$

Each fiber of B looks like a copy of the group G except that there is no canonical association of a point of a fiber with the identity element in G.

As a consequence of condition 3) above, one has an open covering $\{U_\alpha\}$ of B together with the diffeomorphisms

$$\begin{aligned}\phi_\alpha:\ \pi^{-1}(U_\alpha)&\to U_\alpha\times G\\ p&\to(\pi(p),\eta_\alpha(p)).\end{aligned}$$

If $p\in\pi^{-1}(U_\alpha\cap U_\beta)$, then

$$\eta_\beta(pg)\cdot\eta_\alpha(pg)^{-1}=\eta_\beta(p)\cdot\eta_\alpha(p)^{-1}.$$

Thus, the map $p\in\pi^{-1}(U_\beta\cap U_\alpha)\to\eta_\beta(p)\cdot\eta_\alpha(p)^{-1}$ is constant on fibers. Hence, it induces a differentiable mapping

$$w_{\beta\alpha}:\ U_\alpha\cap U_\beta\to G.$$

The mappings $w_{\beta\alpha}$ for $U_\alpha\cap U_\beta$ nonempty, are called transition functions associated with the covering $\{U_\alpha\}$. It is an easy exercise to verify that

$$(*)\quad w_{\gamma\alpha}(p)=w_{\gamma\beta}(p)\cdot w_{\beta\alpha}(p).$$

Condition (*) (known sometimes as the *cocycle condition*) brings the study of fiber bundles into the realm of algebraic topology (see, e.g., Steenrod [21]). It is important to note that given the transition functions satisfying condition (*), it is possible to reconstruct the bundle (see Nomizu [22] for this). The simplest examples of principal bundles are of course the trivial bundles, that is, where the bundle space B is diffeomorphic to $M\times G$. These are characterized by the existence of a *cross section*. Given a principal bundle (B,π,M,G) a cross section of the bundle is a map $\gamma:M\to B$, which commutes with the projection

$$\begin{array}{ccc} & B & \\ \pi & \downarrow\ \uparrow & \gamma \\ & M & \end{array}$$

that is, $\pi\circ\gamma=i_M$ the identity diffeomorphism on M. If a principal bundle admits a cross section γ then given a $p\in B$ with $\pi(p)=x\in M$, there is a unique $g\in G$ such that $\gamma(x)\cdot g=p$. This is a consequence of the freeness of the action of G on B. In this case the map ϕ_γ: $B\to M\times G$ given by $\phi_\gamma(p)=(x,g)$ is a diffeomorphism and the bundle is trivial. The converse is an equally easy exercise. In case a principal bundle is not trivial one still has *local sections* defined on open subsets of M.

Returning now to the main discussions, we consider the following example. Let $\hat{R}$ denote the set of all minimal triples $[A,b,c]$ where A is an $n\times n$ matrix, $b\in\mathfrak{R}^{n\times 1}$, and $c\in\mathfrak{R}^{1\times n}$ is a row vector. $\hat{R}$ is obtained from $\mathfrak{R}^{n^2+2n}$ by excluding the set K defined by

$$K=\{[A,b,c]: \det[b,Ab,\cdots,A^{n-1}b]\cdot\det[c',A'c',\cdots,A'^{n-1}c]=0\}.$$

Clearly, K is a closed set in the usual $\mathfrak{R}^{n^2+2n}$ topology and $\hat{R}$ inherits an (open) subspace topology from $\mathfrak{R}^{n^2+2n}$. The map

$$\pi: \hat{R}\to \mathrm{Rat}(n)$$
$$[A,b,c]\to c(sI-A)^{-1}b=g(s)$$

is a projection onto Rat(n). The group $Gl(n)$ acts on $\hat{R}$ to the *right* by the transformations

$$A\to P^{-1}AP$$
$$b\to P^{-1}b$$
$$c\to cP \qquad \text{for all } P\in Gl(n).$$

That the action is free is a consequence of controllability/observability. The map

$$\lambda: \mathrm{Rat}(n)\to\hat{R}$$

is defined by $\lambda(g(s))=[\bar{A},\bar{b},\bar{c}]$ where

$$g(s)=\frac{q_{n-1}s^{n-1}+\cdots+q_1s+q_0}{s^n+p_{n-1}s^{n-1}+\cdots+p_1s+p_0}$$

and

$$\bar{A}=\begin{bmatrix} 0 & & & \\ 0 & & & \\ \vdots & & I_{n-1} & \\ 0 & & & \\ \hline -p_0 & -p_1 & \cdots & -p_{n-1}\end{bmatrix};\quad \bar{b}=\begin{bmatrix}0\\0\\ \vdots\\0\\1\end{bmatrix};$$

$$\bar{c}=(q_0,q_1,\cdots,q_{n-1}).$$

The map λ is well-defined on Rat(n) and is a global section. Thus, the set $\hat{R}$ of the minimal triples can be given the structure of a principal bundle with structure group (fiber) $Gl(n)$: $\hat{R}\simeq\mathrm{Rat}(n)\times Gl(n)$.

The existence of a standard controllable form made it possible to write down the section immediately. In the case of Rat(p,q), with $G_{\mathfrak{A}}$ and $G_{\mathfrak{B}}$ acting freely on it, it is then natural to look for some suitable canonical form for orbits. The appropriate base space for a bundle with Rat(p,q) as bundle space and $G_{\mathfrak{A}}$ as structure group is picked in the following way.

Consider Rat(p,q) with $|p-q|>1$. Then $G_{\mathfrak{A}}$ acts freely on Rat(p,q) (from Theorem 2). Further,

1) using $s\to s+\sigma$, the barycenter of the zeros can be brought to the origin $s=0$. Hence, $q_{m-1}=0$, where $m=$ number of zeros $(\leqslant n-1)$ is an invariant on $G_{\mathfrak{A}}$ orbits (from Lemma 1). Here $\sigma\in\mathfrak{R}$;

2) using $s\to\alpha s$, and $g(s)\to\beta g(s)$ we can make $q_m=q_l=1$, where q_l is the last nonzero element in the sequence $(q_m,\cdots,q_l,q_{l-1},\cdots,q_0)$. We only have to solve for β and $\alpha>0$, the pair of equations

$$\beta(q_m/\alpha^{n-m})=1$$
$$\beta(q_l/\alpha^{n-l})=1;$$

3) using feedback, $g(s)\to g(s)/(1+kg(s))$ we can set $\bar{p}=(p_{n-1},\cdots,p_0)'$ orthogonal to $\bar{q}=(0,\cdots,0,q_m,\cdots,q_0)'$ by solving for k,

$$\langle\bar{p},\bar{q}\rangle=-k\langle\bar{q},\bar{q}\rangle.$$

Thus, a given $G_{\mathfrak{A}}$ orbit has the canonical form

$$g(s)=\frac{s^m+0\cdot s^{m-1}+\cdots+q_{l+1}s^{l+1}+s^l}{s^n+p_{n-1}s^{n-1}+\cdots+p_1s+p_0}$$

(with the orthogonality condition implicit). From the freeness of $G_{\mathfrak{A}}$-action on Rat(p,q), we note that there is only one element of $G_{\mathfrak{A}}$ bringing a particular $g(s)$ to the above form. Let Q be the set of all rational functions in the above canonical form $\in$Rat(p,q). It is clearly a closed subset of Rat(p,q) and hence when given the subspace topology, it acquires the Hausdorff as well as manifold structure from Rat(p,q). Then we have Rat(p,q) as a (trivial) principal $G_{\mathfrak{A}}$-bundle over Q

$$(\mathrm{Rat}(p,q),\pi,Q,G_{\mathfrak{A}}).$$

The map

$$\lambda: Q\to\mathrm{Rat}(p,q)$$
$$g(s)\mapsto g(s)$$

is the required global cross section. Proceeding as above, it is possible to give Rat(p,q) an alternative bundle structure with $G_{\mathfrak{B}}$ as structure group. For this we use a different canonical form. See [9] for details.

In the remaining part of the paper we examine some aspects of the scaling theory that have implications for the identification problem.

V. The Identification Problem and Scaling

Since the scalings introduced in this paper are all within control of an experimenter, the question naturally arises as to what extent system identification algorithms respect the scalings. The question of invariance of a statistical problem under the action of a Lie group has received some attention in the literature in connection with classical statistical problems, e.g., the simple measurement model, regression model (cf. Fraser [11]). In these cases the parameter space is itself a homogeneous space. Our own concern is with inference problems on stochastic processes and the identification problem for linear systems is a canonical example. Consider a Gaussian stochastic process $\{x_t: t\in T\}$ arising from experiments which suggest a spectral density of the rational type. Our goal is to represent it as a linear system driven by white noise. Thus, in case the time index set $T=\mathfrak{R}$, then we are looking for a representation of the form

$$(*)\quad \begin{aligned} dz(t) &= Az(t)\,dt + b\,dw(t) \\ x(t) &= cz(t) \end{aligned}$$

where $A: \mathfrak{R}^n \to \mathfrak{R}^n$, $b \in \mathfrak{R}^n$, $c: \mathfrak{R}^n \to \mathfrak{R}$, $z(t) \in \mathfrak{R}^n$, and $w(t)$ is a standard Wiener process. When $T=\mathfrak{Z}$ the integers, the corresponding representation is

$$(**)\quad \begin{aligned} z_{t+1} &= Fz_t + gu_t \\ x_t &= hz_t \end{aligned}$$

where $\{u_t\}$ is a standard Gaussian white noise. Of course, such a representation is determined only up to its transfer function $g(s)=c(sI-A)^{-1}b$. Thus, the basic parameter space is Rat(n) taken together with the group actions of $G_{\mathcal{Q}}, G_{\mathcal{B}}$ and possibly other groups as well. Picking a member of Rat(n) that maximizes a certain *support function* evaluated on sample paths $\{x_t\}$ solves the identification problem.

Recall that the sample paths $\{x_t\}$ of $(*)$ belong to $\mathcal{C}^{(n-m-1)}$, the space of functions on $\mathfrak{R}$ with $(n-m-1)$ continuous derivatives, where n is the McMillan degree and m is the number of zeros of $g(s)=c(sI-A)^{-1}b$ (cf. Hajek [12]). Group $G_{\mathcal{Q}}$ acts on Rat(n) leaving the number of zeros and the McMillan degree invariant (from Lemma 1). This action induces an action on the space $\mathfrak{X}$ of sample paths of $(*)$, which has to be interpreted in the almost sure sense. Thus, e.g., $g(s) \mapsto mg(s)$ induces $x(\cdot) \to mx(\cdot)$ on $\mathfrak{X}$.

In general, let $\mathcal{G}(\theta,\hat{\theta},x,T)$ denote the identification criterion or support function, i.e., if θ represents the true transfer function, then $\mathcal{G}(\theta,\hat{\theta},x,T)$ is a measure of the support for an estimate $\hat{\theta}$ of θ given by a realization (sample path) $x(\cdot)$ of duration T. (In the discrete-time case, let T denote the number of sample points.) A commonly used support function is the log-likelihood function or its time average. The ideas of group invariant identification can be introduced as follows.

Definition 1: Let G be a group which acts on Rat(n). An identification procedure is said to be G-invariant if for all $g \in G$

$$\mathcal{G}(\theta,\hat{\theta},x,T) = \mathcal{G}(g(\theta),g(\hat{\theta}),g(x),g(T)) \quad \text{w.p.l.}$$

That is to say, the believability of an estimate $\hat{\theta}$ does not depend on the choice of "scale" insofar as G defines scale changes.

Definition 2: An identification procedure is said to be asymptotically G-invariant if, for all $g \in G$,

$$\lim_{T\to\infty} \mathcal{G}(\theta,\hat{\theta},x,T) = \lim_{T\to\infty} \mathcal{G}(g(\theta),g(\hat{\theta}),g(x),g(T)) \qquad \text{w.p.l.}$$

Let us examine the G-invariance of the discrete-time problem $(**)$ of this section with

$$g(z)=\frac{q(z)}{p(z)}=\frac{1+q_{m-1}z^{-1}+\cdots+q_0z^{-m}}{1+p_{n-1}z^{-1}+\cdots+p_0z^{-n}} \qquad m \leqslant n-1.$$

To be able to write down the likelihood ratio even approximately, one would want to assume stationarity. Let $g(z)$ have its poles and zeros within the open unit disk. In input–output form, $(**)$ is written as

$$x_t+p_{n-1}x_{t-1}+\cdots+p_0x_{t-n} = u_t+q_{m-1}u_{t-1}+\cdots+q_0u_{t-m}.$$

There is no loss of generality in setting $q_0 \neq 0$. Then the log-likelihood for the system is

$$\log(L) = -\frac{T}{2}\log(2\pi) - T\log(\sigma) - \frac{1}{2}\sum_{t=1}^{T}\frac{\epsilon_t^2}{\sigma^2} + \eta_T$$

where $E\eta_T^2 \leqslant a < \infty$ and "a" is not dependent on T. Here ϵ_t is the residual sequence generated by

$$\hat{q}(z)\epsilon_t = \hat{p}(z)x_t$$

where $\hat{g}(z)=\hat{q}(z)/\hat{p}(z)$ denotes an estimate of $g(z)$, and $\sigma^2=E[\epsilon_t^2]$. Let $\theta=(p_{n-1},\cdots,p_0,q_{m-1},\cdots,q_0)'$ and let $V_T = 1/T\sum_{t=1}^{T}\epsilon_t^2$. V_T depends on both θ and its estimate $\hat{\theta}$. The identification criterion or support function is then

$$\begin{aligned} \mathcal{G}(\theta,\hat{\theta},x,T) &= \frac{1}{T}\log(L) \\ &= -\log(2\pi) - \log(\sigma) - \frac{1}{2\sigma^2}V_T. \end{aligned}$$

If the estimate $\hat{g}(z)$ has also its poles and zeros in the open unit disk, then it can be shown that the residual sequence is ergodic (cf. Doob [13]). Hence,

$$\lim_{t\to\infty} V_T = E(\epsilon_t^2) = V(\theta,\hat{\theta}) \qquad \text{w.p.l.}$$

Further, by an application of the Parseval–Plancheral theorem,

$$V = \frac{1}{2\pi i}\oint_C \frac{g(z)g(z^{-1})}{\hat{g}(z)\hat{g}(z^{-1})}\frac{dz}{z}$$

where C is the unit circle. If we replace V_T by its asymptotic approximation V in $\log(L)$ and σ by its best supported value, $\sigma=\sqrt{V}$, then

$$\log(L) \simeq -\frac{T}{2}\log(V),$$

or the identification criterion in the asymptotic limit is

$$\lim_{T\to\infty}\mathcal{G}(\theta,\hat{\theta},x,T) = -\frac{1}{2}\log(V) \qquad \text{w.p.l.}$$

Hence, the best supported estimate minimizes V. We are interested in scaling on V.

It is important to note that $G_{\mathcal{Q}}$-action does not preserve the conditions of asymptotic stability and minimum phase, necessary for likelihood inference. However, given $q(s)/p(s) \in \text{Rat}(n)$ we may restrict $g \in G_{\mathcal{Q}}$ to those that do preserve these properties. This results in a pseudogroup action on Rat(n) and we call the associated scalings admissible. For example, the exponential scaling (in continuous time),

$$g(s) \to g(s+\sigma), \qquad \sigma \in \mathfrak{R},$$

acting on $g(s)=\alpha/s+\lambda$, $\lambda>0$ will not preserve asymptotic stability for $\sigma<-\lambda$. In what follows, let us denote by G_S, the subgroup of $G_{\mathcal{Q}}$ defined by the one-parameter groups of magnitude scaling and exponential scaling,

$$g(s) \to mg(s) \qquad m>0$$
$$g(s) \to g(s+\sigma) \qquad \sigma \in \mathfrak{R}.$$

The discrete-time analogs of these scalings are, respectively,

$$g(z) \to mg(z)$$
$$g(z) \to g(\beta z) \qquad m, \beta \in \mathfrak{R}^+.$$

The admissible G_S-action for discrete-time is given by $\alpha \in \mathfrak{R}^+$ and $\beta > \beta_0$. We have the following result.

Lemma 2: The likelihood function is asymptotically G_S-invariant for admissible G_S-action.

Proof: It follows from the discussion above that we only have to show

$$V(\theta, \hat{\theta}) = V(t(\theta), t(\hat{\theta}))$$

for (admissible) $t \in G_S$. Now under constant scaling,

$$V(t(\theta), t(\hat{\theta})) = \frac{1}{2\pi i} \oint_C \frac{m^2 g(z) g(z^{-1})}{m^2 \hat{g}(z) \hat{g}(z^{-1})} \frac{dz}{z}$$
$$= \frac{1}{2\pi i} \oint_C \frac{g(z) g(z^{-1})}{\hat{g}(z) \hat{g}(z^{-1})} \frac{dz}{z}$$
$$= V(\theta, \hat{\theta}).$$

Under exponential scaling,

$$V(t(\theta), t(\hat{\theta})) = \frac{1}{2\pi i} \oint_C \frac{g(mz) g(m^{-1}z^{-1})}{\hat{g}(mz) \hat{g}(m^{-1}z^{-1})} \frac{dz}{z}$$
$$= \frac{1}{2\pi i} \oint_C \frac{g(mz) g(m^{-1}z^{-1})}{\hat{g}(mz) \hat{g}(m^{-1}z^{-1})} \frac{d(mz)}{(mz)}$$
$$= \frac{1}{2\pi i} \oint_{\tilde{C}} \frac{g(z) g(z^{-1})}{\hat{g}(z) \hat{g}(z^{-1})} \frac{dz}{z}$$

where $\tilde{C}$ = image of the unit circle under the map $z \to mz$. Since the exponential scaling is admissible, the annulus between C and $\tilde{C}$ contains no poles or zeros of $g(z)$ or $\hat{g}(z)$. Therefore, the integral on the right may be replaced by an integral over C which coincides with $V(\theta, \hat{\theta})$.

Remark: The scaling actions of Lemma 2 are rather "weak" structural changes and the result is perhaps not unexpected. The G_S-action on Rat(n) is "rigid" in the sense that the relative support for θ over $\hat{\theta}$ remains unchanged. On the other hand, introduction of feedback is a more complex change and one may expect that a result similar to Theorem 1 *does not* hold. That this is so is seen from the following example.

The action of interest is (for admissible k),

$$g(z) \to \frac{g(z)}{1+kg(z)} \qquad k \in \mathfrak{R}.$$

Consider the first-order system

$$g(z) = \frac{1}{z+\lambda}.$$

Then

$$V(\lambda, \tilde{\lambda}) = \frac{1}{2\pi i} \oint_C \frac{(z+\hat{\lambda})(z^{-1}+\hat{\lambda})}{(z+\lambda)(z^{-1}+\lambda)} \frac{dz}{z}$$

$$(|\lambda|<1, \ |\hat{\lambda}|<1 \text{ for asymptotic stability}).$$

A short calculation shows that

$$V(\lambda, \hat{\lambda}) = 1 + \frac{(\hat{\lambda}-\lambda)^2}{1-\lambda^2}.$$

Under feedback, $g(z) \to 1/z+(k+\lambda)$ with the admissibility requirement $|k+\lambda|<1$. Then

$$V(\lambda+k, \hat{\lambda}+k) = 1 + \frac{(\hat{\lambda}+k-\lambda-k)^2}{1-(\lambda+k)^2}$$
$$= 1 + \frac{(\hat{\lambda}-\lambda)^2}{1-(\lambda+k)^2}.$$

Clearly V is not invariant under feedback. Now it is natural (cf. Edwards [14]) to think of the log-likelihood as a measure of the information for discrimination between the hypotheses λ (true) and $\hat{\lambda}$ (alternative), which can be made to increase or decrease under feedback depending on the choice of k. For the choice of feedback $k=-\lambda$, V has a stationary point but that is a minimum. The "optimal" choice would be to pick k, to bring $\lambda+k$ close to 1. A better understanding of this requires the analysis of the Fisher information which follows in the next section.

The analysis of this section can be extended to continuous-time problems, were the parameter space is Rat$_{SM}(n)$, the set of rational functions of degree n with poles and zeros in the open left half-plane. The corresponding expression for $V(\theta, \hat{\theta})$ is

$$V(\theta, \hat{\theta}) = \int_{-\infty}^{\infty} \log \frac{g(j\omega) g(-j\omega)}{\hat{g}(j\omega) \hat{g}(-j\omega)} dw.$$

We also note that a recently introduced identification criterion due to Akaike shares some of the invariance properties discussed here.

In the next section we consider some of the invariance questions for identification in a local setting. The reason for doing so is that global distance measures on Rat(n) such as $V(\theta, \hat{\theta})$ poorly reflect the geometry of the space itself.

VI. Fisher Information and Scaling

Here we study the effect of scaling on Fisher information. Since several well-known and useful measures of information such as Kullback–Liebler information and Renyi's entropy are locally equivalent to the Fisher information, our results have wider applicability. Further, there is the well-known relation to parameter estimation accuracy via the Cramér–Rao lower bound. For the system (**) of Section V, the Fisher information is

$$E\left\{-\frac{\partial^2}{\partial\hat{\theta}^i \partial\hat{\theta}^j}\log L\right\}\Bigg|_{\hat{\theta}=\theta}=\frac{T}{2\sigma^2}\frac{\partial^2 V(\theta,\hat{\theta})}{\partial\hat{\theta}^i\partial\hat{\theta}^j}\Bigg|_{\hat{\theta}=\theta.}$$

We define the *Fisher information rate* F to be

$$F_{ij}\triangleq\frac{\partial^2 V(\theta,\hat{\theta})}{\partial\hat{\theta}^i\partial\hat{\theta}^j}\Bigg|_{\hat{\theta}=\theta.}$$

We intend to show that F defines a metric. Let $M=D\operatorname{Rat}_{SM}(n;m)$ denote the subset of $\operatorname{Rat}(n)$ consisting of rational functions $g(s)$ with n poles and m zeros all lying within the open unit disk and satisfying

$$\lim_{|s|\to\infty} s^{(n-m)}g(s)=1.$$

Then M is an analytic manifold of dimension $k=m+n$ with subspace topology from $\operatorname{Rat}(n)$ [9]. Pick some $g_0(\cdot)\in M$ and let $\theta=(\theta^1,\theta^2,\cdots,\theta^k)$ be local coordinates in some neighborhood U of M containing $g_0(\cdot)$. Then the real-valued function

$$f: M\to\Re^1$$
$$g(\cdot)\mapsto\frac{1}{2\pi i}\oint_C\frac{g_0(z)g_0(z^{-1})}{g(z)g(z^{-1})}\frac{dz}{z}$$

satisfies the following [15]:

a) $f(g)\geqslant 1$ for $g(s)\in M$ and $f(g)=1$ iff $g_0=g$

b) The gradient $df_g=0$ at $g\in M$ iff $g=g_0$ (unique critical point);

hence,

c) $H_f(g_0)=\dfrac{\partial^2 f}{\partial\theta^i\partial\theta^j}\Big|_{\theta=\theta_0}$ is positive definite.

It follows from part b) that if (ϕ, W) is another coordinate chart $g_0\in W$, then in these coordinates the matrix representation of H_f is given by

$$\frac{\partial\theta^p}{\partial\phi^i}\frac{\partial\theta^q}{\partial\phi^j}\frac{\partial^2 f}{\partial\theta^p\partial\theta^q}\Bigg|_{\theta=\theta_0}\quad\text{(sum over repeated indexes).}$$

Thus, $H_f(g_0)$ is an invariantively defined symmetric bilinear form on the tangent space TM_{g_0} of index 0. Noting that the matrix elements of H_f are rational functions of θ_0^i, as g_0 varies over M, we have a smooth cross section of $\mathfrak{T}_2^0(M)$—the set of tensor fields on M of type (0,2). Further, the positive definiteness implies that we have a *Riemannian metric on M*. We call this the Fisher metric on M, e.g., $n=1$, $m=0$ $M=\{g(\cdot): g(z)=1/(z+\lambda)|\lambda|<1\}$ and $ds^2=2/(1-\lambda^2)d\lambda^2$ defines the Fisher metric.

The behavior of this metric in certain directions has important implications for identifiability. To see this, let $\overline{M}$ be the closure of M by admitting cancellation. Then $\overline{M}\simeq\Re^k$. Suppose $g_0\in\overline{M}-M$ and in particular $g_0\in D\operatorname{Rat}(n-1;m-1)$. Since f has an obvious extension to $\overline{M}$ satisfying conditions as before, one can consider the critical points of

$$f:\overline{M}\to\Re^1$$
$$g(\cdot)\mapsto\frac{1}{2\pi i}\oint_C\frac{g_0(z)g_0(z^{-1})}{g(z)g(z^{-1})}\frac{dz}{z}$$
$$g_0(\cdot)\in D\operatorname{Rat}(n-1;m-1).$$

Now in $\overline{M}, g_0=g$ (*iff* $df_g=0$) has infinitely many solutions (lying on an algebraic surface). Now we appeal to the classic result.

Morse Lemma [16]: Let p be a nondegenerate critical point of f: $M\to\Re$, a real-valued function on a manifold. Then there is a local coordinate system $(y^1,\cdots,y^n)$ in a neighborhood U of p with

$$y^i(p)=0 \qquad \text{for all } i$$

and such that the identity

$$f=f(p)-(y^1)^2-\cdots-(y^\lambda)^2+(y^{\lambda+1})^2+\cdots+(y^n)^2$$

holds *throughout* U where $\lambda=$ index of f at p.

This has an immediate *corollary:* nondegenerate critical points are isolated. In the present case with g_0 having common factors, the critical points are clearly nonisolated. Hence, they are degenerate, that is, the Fisher metric tends to become degenerate as we approach rational functions with common factors. This is a local version of the indistinguishability of rational functions with common factors as far as the measure $V(\theta,\hat{\theta})$ is concerned.

Remark: The approach outlined above in establishing the equivalence of the two notions of unidentifiability arising from pole-zero cancellations and degeneracy of Fisher information appears to be new.

Returning to our original concern of G-invariant identification, we examine the scaling action on the Fisher metric. The methods are quite general.

Suppose ϕ: $M\to N$ is a C^∞ map of manifolds. We say ϕ is *regular* if the Jacobian

$$\phi_{*_m}: T_m(M)\to T(N)_{\phi(m)}$$

is one-to-one for each $m\in M$. Such maps pull back structures in the following sense (see, e.g., [17]).

Theorem 3: Let $T \in \mathfrak{T}_s^0(N)$ the collection of C^∞, covariant tensor fields of degree s on N. Then for every regular map $\phi: M \to N$, there exists a unique C^∞ tensor field $\in \mathfrak{T}_s^0(M)$ denoted as

$$S = \phi^* T \qquad \text{on } E = \phi^{-1}\,(\text{domain } T)$$

such that for every $m \in E$

$$\phi_m^* T(\phi(m)) = S(m).$$

Remark 1: Here ϕ_m^* has been extended as a homomorphism

$$\phi_m^*: T^*_{\phi(m)}(N) \to T^*_m(M)$$

defined by

$$\langle \phi_{*m} v, t \rangle = \langle v, \phi_m^* t \rangle \qquad \text{for all } v \in T_m(M),\ t \in T^*_{\phi(m)}(N).$$

Thus, for $v_1, v_2, \cdots, v_s \in T_m(M)$

$$S(v_1, v_2, \cdots, v_s) = T(\phi_{*m} v_1, \phi_{*m} v_2, \cdots, \phi_{*m} v_s).$$

Remark 2: Symmetry and skew symmetry of s-linear forms are preserved under the pullback ϕ^*.

Consider the scalings and feedback on $M = D\,\mathrm{Rat}_{SM}(n; m)$,

$$z \to \beta z$$

$$g(z) \to g(z)/(1 + kg(z))$$

where β and k are suitably restricted. They define one parameter pseudogroup actions on M and commute with each other. Denoting these regular maps (diffeomorphisms) also as β and k, and the Fisher metric as

$$ds^2 = F_{pq}\, d\theta^p d\theta^q$$

for a local chart $(\theta; U)$, we have

$$F^\beta = \beta^* F \quad \text{and} \quad F^k = k^* F,$$

also Riemannian metrics (pullback preserves the index of a bilinear form). Using Theorem 3, matrix representations are obtained as

$$F_{ij}^\beta = F^\beta\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right)_{(m)}$$

$$= F\left(\phi_{*m}\frac{\partial}{\partial\theta^i}, \phi_{*m}\frac{\partial}{\partial\theta^i}\right)_{\phi(m)}$$

and similarly for F^k. However, since the Fisher metric is rather special in that it is defined via a certain Hessian, the change of matrix representation is easily given as

$$F_{ij}^\beta(\theta) = \frac{\partial\theta^p}{\partial\bar{\theta}^i}\frac{\partial\theta^q}{\partial\bar{\theta}^j} F_{pq}(\bar{\theta} = \beta(\theta)).$$

We simply note here that this transformation tells us to what extent the scalings affect identification accuracy (cf. [9], [18]).

VII. Final Remarks

In this paper, we have described a theory of scaling for linear systems using methods from Lie theory. The implications of this theory to questions on identification of systems are also discussed. In particular, we have introduced the concept of invariant identification. Some of these questions have been developed further in a conference paper [19]. The particular choice of scalings has been motivated as much by practical considerations as by their geometrical properties established here. The theory lends itself to generalization to the case of nonrational functions [9], and for multivariable systems where the larger scaling groups may provide a great deal more information on the structure of minimal systems.

References

[1] R. W. Brockett, "Some geometric questions in the theory of linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 449–455, Aug. 1976.

[2] ——, "The Lie groups of simple feedback systems," in *Proc. IEEE Decision and Control Conf.*, New York, 1976.

[3] R. W. Brockett and P. S. Krishnaprasad, "Scaling rational functions and linear system identification," in *Proc. 1977 Conf. on Information Sciences and Systems*, Johns Hopkins Univ., pp. 501–506.

[4] R. L. Bishop and R. J. Crittendon, *Geometry of Manifolds*. New York: Academic, 1969.

[5] D. Husemoller, *Fibre Bundles*. New York: Springer-Verlag, 1975.

[6] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*. New York: Springer-Verlag, 1972.

[7] F. R. Gantmakher, *The Theory of Matrices*. New York: Chelsea, 1960.

[8] B. L. Van der Waerden, *Modern Algebra*. New York: Wiley, 1948.

[9] P. S. Krishnaprasad, Ph.D. dissertation, Harvard Univ., 1977.

[10] D. J. Simms and N. M. J. Woodhouse, *Lectures on Geometric Quantization* (*Lecture Notes in Physics*). Berlin, Germany: Springer/Verlag, 1976.

[11] D. A. S. Fraser, *The Structure of Inference*. New York: Wiley, 1968.

[12] J. Hajek, "On linear statistical problems in stochastic processes," *Czechoslav. Math. J.*, vol. 12, pp. 404–444, 1962.

[13] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.

[14] A. W. F. Edwards, *Likelihood*. London, England: Cambridge Univ. Press, 1972.

[15] K. J. Aström and T. Söderström, "Uniqueness of maximum likelihood estimates of parameters from ARMA Model," *IEEE Trans. Automat. Contr.*, vol. 19, pp. 769–773, 1976.

[16] J. W. Milnor, *Morse Theory: Annals of Mathematics Studies*. Princeton, NJ: Princeton Univ. Press, 1969.

[17] R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds*. New York: Macmillan, 1968.

[18] T. Gustavson, L. Ljung, and T. Söderström, "Identification of processes in closed loop—Identifiability and accuracy aspects," *Automatica*, vol. 13, pp. 59–75, Jan. 1977.

[19] P. S. Krishnaprasad, "Geometry of parametric models—Some probabilistic questions," in *Proc. 15th Allerton Conf. on Control, Communication and Computing*, 1977, pp. 661–670.

[20] R. W. Brockett, "Nonlinear systems and differential geometry," *Proc. IEEE*, vol. 64, pp. 61–72, Jan. 1976.

[21] N. E. Steenrod, *The Topology of Fiber Bundles*. Princeton, NJ: Princeton Univ. Press, 1951.

[22] K. Nomizu, *Lie Groups and Differential Geometry*. Tokyo, Japan: Math. Soc. Japan, 1956.

Roger W. Brockett (S'62–M'63–SM'73–F'74) received the Ph.D. degree from Case Western Reserve University, Cleveland, OH.

From 1963 to 1969 he taught in the Department of Electrical Engineering at the Massachusetts Institute of Technology, Cambridge. He is presently a Gordon McKay Professor of Applied Mathematics in the Division of Applied Sciences, Harvard University, Cambridge, MA, were he participates in the Interdisciplinary Decision and Control Program through teaching and research. He is Associate Editor of *SIAM Journal on Control and Optimization*, *SIAM Journal on Mathematical Analysis*, and *Nonlinear Analysis*. He is author of the book *Finite Dimensional Linear Systems* (New York: Wiley) and coeditor (with D. Q. Mayne) of *Geometric Methods in System Theory* (Dordrecht, The Netherlands: Reidel, 1973). He has held visiting positions at Warwick University, Imperial College, University of Rome, Washington University, University of Groningen, University of Florida, University of Nagoya, and the Australian National University.

Dr. Brockett is recipient of the American Automatic Control Council's Donald P. Eckman Award. In 1975 he held a Guggenheim Fellowship for study in mathematical system theory.

P. S. Krishnaprasad (S'73–M'77) received the Ph.D. degree from Harvard University, Cambridge, MA, in 1977.

He is currently an Assistant Professor of Systems Engineering at Case Western Reserve University, Cleveland, OH. His interests in system theory center around the differential and algebro-geometric aspects of the identification problem. His recent work also makes close contact with problems in analytical mechanics.

Dr. Krishnaprasad is a member of the American Mathematical Society.

Invertibility of C^∞ Multivariable Input–Output Systems

DEBORAH REBHUHN

***Abstract*—Scholars have discussed C^∞ input–output systems in a number of settings. Recent work of Hirschorn's has given sufficient conditions for local invertibility of systems with multivariable inputs. Here we give another such condition which appears to be independent of Hirschorn's.**

I. Introduction

Let N, P be C^∞ paracompact manifolds and let ϕ: $N \rightarrow P$ be a C^∞ mapping. Let $\tilde{x} \in N$ be a fixed point. As usual, for $x \in N$, $y \in P$ we let TN, TP, T_xN, T_yP, respectively, represent the tangent bundles to N and P and the tangent spaces at x and y of N and P. For $i = 1, 2, \cdots, m$ we let A, B^i be C^∞ vector fields on N. Let $\mathcal{U}$ be the set of C^∞ curves in $\mathbb{R}^m$ with a domain that includes a neighborhood of zero in $\mathbb{R}$. The elements of $\mathcal{U}$ will be called inputs. We will say $u', u'' \in \mathcal{U}$ differ by a finite order if for some nonnegative integer j,

$$\frac{d^j}{dt^j} u'(0) \neq \frac{d^j}{dt^j} u''(0).$$

I.1 Definition: For any $u \in \mathcal{U}$ and $x \in N$, we can define the differential equation

$$(*) \quad \frac{d}{dt}(x_u(x,t)) = A(x_u(x,t)) + \sum_{i=1}^{m} u^i B^i(x_u(x,t))$$
$$x_u(x,0) = x.$$

1) The curve $x_u(x,t)$ will be called the response to u starting at x. When $x = \tilde{x}$, we will write $x_u(t)$ instead of $x_u(\tilde{x}, t)$.

2) The curve $y_u(x,t) = \phi(x_u(x,t))$ will be called the output of u determined by x. When $x = \tilde{x}$, we will write $y_u(t)$ instead of $y_u(x,t)$ and call $y_u(t)$ the output of u.

I.2 Definition: The system $u \rightarrow y_u$ is called locally invertible at $\tilde{x} \in N$ if there is an open neighborhood $\tilde{N}$ of $\tilde{x}$ in N such that whenever $x_{u'}$ and $x_{u''}$ lie in $\tilde{N}$ for u' and u'' distinct, then $y_{u'}$ and $y_{u''}$ are distinct.

We will give sufficient conditions to guarantee that when u', u'' differ by a finite order, then $y_{u'}$ and $y_{u''}$ are

Manuscript received December 26, 1978; revised August 17, 1979. Paper recommended by M. Vidyasagar, Chairman of the Stability, Nonlinear, and Distributed Systems Committee.

The author was with the Department of Mathematics, Vassar College, Poughkeepsie, NY 12601. She is now with Bell Laboratories, Holmdel, NJ 07733.

0018-9286/80/0400-0207$00.75 ©1980 IEEE