UDC: 65.012.4
Original scientific paper

ADAPTIVE CONTROL BASED ON THE APPLICATION OF A SIMPLIFIED UNIFORM STRUCTURES AND LEARNING PROCEDURES

József K. Tar, Miklós Rontó
Budapest Polytechnic, John von Neumann Faculty of Informatics
E-mail: jktar@zeus.banki.hu| ronto@matavnet.hu

The present state of creating a new branch of Soft Computing (SC) for particular problem classes, possibly wider than the control of mechanical systems, is reported in this article. Like "traditional" SC it evades the development of analytical system models, and uses uniform structures, but these structures originate from various Lie groups. The advantages are a drastic reduction in size and an increase in lucidity. The generally "stochastic or semi-stochastic" "learning" or parameter tuning seems to be replaceable by simple explicit algebraic procedures of limited steps, too. The idea originated from mechanical systems' control while considering their general internal symmetry group, and later it was further developed by using specific general features of it on a much wider scale. Convergence considerations are given for MIMO and SISO systems, too. Simulation examples are presented for the control of the inverted pendulum with the use of the Generalized Lorentzian Matrices. It is concluded that the method is promising and probably imposes acceptable convergence requirements in many cases.

Keywords: adaptive control, Soft Computing, Lie groups, partial system-identification, Lorentz group.

1. INTRODUCTION

The increasing computational power of modern ICs make it possible to consider the implementation of several types of SC tools, and this basic idea was almost completely developed in the sixties. Nowadays SC roughly means a combination of neural networks and fuzzy controllers that are enhanced with a high parallelism of operation and supported by several deterministic, stochastic or combined parameter-tuning methods. This tuning is often called "learning".

The advantage of this approach is that the development of any intricate and complicated analytical system model can be evaded. Instead of the typical problem, classes have been identified for the solution where typical uniform architectures have been crystallized. Typical examples are e.g. Multilayer Perceptrons for non-linear mapping and forecasting; Hopfield Networks (e.g. Cellular Neural Networks -CNNs-) for singular mapping and fixed-point problems; Elman Networks for realizing nonlinear mapping with built in system dynamics (e.g. car dynamic's modelling); Kohonen Networks to realize self organizing maps for classification purposes and to reveal correlation, etc. Fuzzy systems usually use membership functions of a typical (e.g. trapezoidal, triangular or Gaussian) form, fuzzy relations, and standard methods for fuzzyfication and defuzzyfication.

The "first phase" when using these methods, identification of the problem class and finding the appropriate structure, is usually relatively easy. However, the following phase is much more difficult. In the case of neural networks, finding the proper number of neurons in the case of a perceptron is not trivial. Certain solutions start with quite a big initial network and apply dynamic pruning for getting rid of the unimportant nodes and connections [1]. Alternative approaches start with small networks in which the number of nodes is increased step by step (e.g. [2-3]). Due to the possibility of "local optima", in the case of the fully deterministic "backpropagation training", the inadequacy of a given number of neurons cannot be concluded simply. As a consequence the "learning methods" were seriously improved in the last decade, including stochastic elements (e.g. [4-7]).

In spite of the latest development in traditional SC, it can be stated that for strongly coupled non-linear MIMO systems under external dynamic interaction "unknown" by the controller (as e.g. mechanical devices' control) this approach still has several drawbacks. The number of necessary fuzzy rules strongly increases with the degree of freedom, and the intricacy of the problem. The same is true for the necessary neurons in a neural network approach. Furthermore, an external dynamic interaction, on which normally no satisfactory information is available, influences the system's behavior in a time-varying manner. Both the big size of the necessary structures, and consequently the huge number of parameters to be tuned, as well as the time-varying nature of the learning process in the "traditional soft computing approach", all still mean serious problems.

Regarding the "source" of these problems, one is likely to be of the opinion that the"generality" and the "uniformity" of the "traditional SC structures" prevents the application of plausible simplifications which may be characteristic to narrower but still wide enough sets of typical tasks. This naturally brings up the idea that several branches of SC could be developed for narrower problem classes if more specific features of these classes could be identified, and then taken into account in the uniform structures. In this way the "mathematical framework" of the modelling approach could be made more simple and lucid. The development of the analytical model of the particular system under consideration still would be avoided as well as in the case of traditional SC.

The first steps in this direction were made in connection with classical mechanical systems [8], while further refinements were published in [9-10]. The basic observation was based on the principles of Classical Mechanics (CM) summarized in [11]: on the tangent space of the state of mechanical systems in general a special geometry, the Symplectic Geometry (SG) can be defined. This geometry has an inherent symmetry described by the Symplectic Group (SGr), in a way similar to the the Orthogonal Group (OG) is inherent in the Euclidean Geometry. Because of this different maps connected to each other by Symplectic Transformations (ST) and were applied to the tangent space of the states of a mechanical system and were interpreted as a means of system-identification in an adaptive control. This means that the elements of SGr served the "structures" that where uniform in the field of CM. For "learning", the parameters were determined by an explicit algebraic procedure called the "Standard Symplectizing Algorithm" (SSA). Eventually, the result of the system identification was a symplectic matrix in each control step. This matrix mirrored the "net effect" of the system's inaccurately modelled dynamics and the "unknown" external dynamic perturbations. In control no effort was made to distinguish between these factors, and this meant that "system identification" was regarded" to be situation dependent".

Simulation investigations revealed that this approach is promising, though it suffers from two definit deficiencies: a) since the "momentum part" of the canonical coordinates cannot be measured by common industrial sensors, this method uses physical quantities with the observation that it has phenomenological difficulties; b) to get a definite solution to an ambiguous problem, the SSA algorithm sometimes results in transformations that is too big for the norm; this sometimes generates computational problems.

To improve the situation the idea of "Minimum Operation Symplectic Transformations" (MOST) was invented [12]. This is a convenient way to keep the appropriate symplectic matrix as close to the identity transformation as possible. Mathematically it corresponds to "additional restrictions", making an originally ambiguous problem unambiguous. The phenomenological problem was evaded by a simple "trick" of replacing the "momentum part" with the joint coordinate accelerations weighted with a fictitious inertia.

Though the MOST transformations were found to be efficient and fitted the needs of mechanical systems, a 2DOF system needs 2DOFx2DOF-sized symplectic matrices. The question of a further reduction in the size of the "identification matrix" naturally arises. From a mathematical point of view, a MOST transformation can be regarded as a mapping which interconnects the "desired" and the "realized" joint coordinate accelerations of the controlled system. The trajectory tracking properties are expressed in pure kinematic terms determining the "desired accelerations". Using the "desired data" and a rough system-model, the robot's drives are forced to exert certain generalized forces causing observable accelerations to react. The improvement of the "rough model" takes place by comparing the desired behavior with the observed one. In the next control cycle this improved model replaces the original rough one. The function of this transformation is to describe the functional relation between a narrow segment of the desired and the known behavior, under particular conditions that are determined by external interactions.

It is natural to suppose that such a functional relation could be described by mathematical means other than symplectic matrices. However, from a purely mathematical point of view, it seems to be very convenient if this functional relation can beis described by some Lie group that is more or less similar to the SGr. To proceed in this way the following steps must be taken:

- Identification of those properties of the SGr which are generally valid in the case of other Lie groups of potential use;
- Investigation into the convergence properties when the Lie group used does not describe any of the internal symmetry of the comtrolled system;

- Investigation into the convergence properties when the Lie group used does not describe any of the internal symmetry of the comtrolled system;

Later, these considerations will be discussed in detail. Finally, simulation examples are presented for the control of the inverted pendulum as the most "popular paradigm". (Similar results for a 3 DOF robot arm under unmodelled environmental interaction were recently published, and they compared the MOST algorithm, the stretched orthogonal group, and the generalized Lorentz group).

2. THE NATURAL UNIFORM STRUCTURES IN CLASSICAL MECHANICS

In CM the canonical equations of motion have the form of

$$\dot{x}_i = \sum_s \Im_{is} \frac{\partial H(\mathbf{x})}{\partial x_s} + \tilde{Q}_i^{Free}, \mathbf{x} = [\mathbf{q}, \mathbf{p}], \Im = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \tag{1}$$

in which $H(\mathbf{x})$ Hamiltonian denotes the full energy of the system, $\mathbf{p}$ and $\mathbf{q}$ is the "momentum", and the "coordinate" part of the physical state $\mathbf{x}$, $\tilde{\mathbf{Q}}^{Free}$ denotes the effect of the external interactions not taken into account in $H(\mathbf{x})$. Hamiltonian can be constructed from the Lagrangian of the system. For building up the Lagrangian the concepts of Newton's inertial systems and kinetic energy can be applied as a phenomenological basis. By following this step a set of possible physical states can be regarded as a differentiable manifold on which the measures of a special map are used in Eq. (1). To describe the above system's state-propagation in the neighborhood of a given state any local, invertible coordinate transformation of the form of $\mathbf{x'}=\mathbf{x'}(\mathbf{x})$ can be used, and this will result in another local map. However, if the Jacobian of the coordinate transformation is a symplectic matrix, that satisfies the restriction of

$$S_{ij} = \frac{\partial x'_i}{\partial x_j}, \mathbf{S}\Im\mathbf{S}^T = \Im \tag{2}$$

the form of Eq. (1) will be conserved, too. Due to the special properties of $\Im$ Eq. (2) can be replaced with

$$\Im^{-1} = \Im^T = -\Im, \Rightarrow \mathbf{S}^{-1} = \Im\mathbf{S}^T\Im^T, \mathbf{S}^T = \Im^T\mathbf{S}^{-1}\Im \Rightarrow \mathbf{S}^T\Im\mathbf{S} = \Im \tag{3}$$

The idea of "adaptive control" and system-identification, consisted of finding a proper, symplectic matrix while interconnecting the observed and the desired state propagation when instead of $\dot{\mathbf{p}}$ the quantity $\mathbf{M}\ddot{\mathbf{q}}$ was used with a fixed positive definite **M**. The proper symplectic matrix was found in accordance with the following considerations: The full set of the linearly independent vectors $\{\mathbf{s}^{(i)}\}$ is called symplectic if it satisfy the constraints:

$$\mathbf{s}^{(i)T}\Im\mathbf{s}^{(j)} = \Im_{ij} \tag{4}$$

that is, according to Eq. (3), the columns of a symplectic matrix just form a symplectic set. In the first step, 2DOF pieces of linearly independent vectors were taken arbitrarily (e.g. the elements of the unit matrix). One of these vectors was replaced by $\dot{\mathbf{x}}^D$ and $\dot{\mathbf{x}}^R$, the desired and the realized state-drift, respectively, that the new sets also remained full and linearly independent (it is possible that the first vector was replaced, or in the system-identification problem was now formulated as a matrix equation: it is sought for a symplectic matrix **S** which satisfies the equation

$$\left[\dot{\mathbf{x}}^D \mid \ldots\right] = \mathbf{S}\left[\dot{\mathbf{x}}^R \mid \ldots\right] \tag{5}$$

that, among others **S,** transforms the observed drift vector into the desired one. Its solution is trivial and computationally cost-effective:

$$\mathbf{S} = \left[\dot{\mathbf{x}}^D \mid \ldots\right] \; \left[\dot{\mathbf{x}}^R \mid \ldots\right]^{-1} = \left[\dot{\mathbf{x}}^D \mid \ldots\right]\Im\left[\dot{\mathbf{x}}^R \mid \ldots\right]^T \Im^T \tag{6}$$

Because of the definition of the symplectic matrices it is very important that the inverse can be calculated by applying only the operation of one transposition and two simple matrix multiplications. While in general calculation of the inverse of a matrix one needs to really increase the number of algebraic operations in a general case, within the frame of the symplectic group inversion, does not create a serious problem.

It is trivial that in the columns denoted by "..." in Eqs. (5,6) there is a great number of ambiguous parameters, which are inherited in the resulting S, too. By replacing the SSA with the MOST algorithm, one makes this ambiguity cease in a convenient and useful way that it keeps S as close to the identity operator as possible. Consequently only the result of the first identification will be much further from the unit transformation, and the other matrices obtained via the step-by-step corrections in the form of $S(n+1)=C(n+1)S(n)$ differed only slightly from each other, that is the correction matrix **C** was very close to the unit matrix. The group properties of the symplectic matrices automatically guarantee that after each correction we also have a symplectic matrix. Later it will be shown that similar properties can be utilized in the case of another groups, too.

3. POTENTIAL LIE GROUPS FOR REPLACING THE SYMPLECTIC GROUP

Generally above considerations correspond to the following logical steps: *1) use a rough initial dynamic model; 2) calculate the necessary generalized forces for the desired joint coordinate accelerations on the basis of this rough model; 3) observe the realized accelerations; 4) create some convenient algebraic means for mapping the observed behavior to the realized one; 5) use the result of this mapping in the next control step.* On the basis of physical considerations for the mechanical systems, this convenient algebraic means was constructed from the elements of the SGr.

From a purely mathematical point of view this group is not the only possible mathematical means by which such a task can be conveniently solved. Let **G** be a nonsingular quadratic, otherwise arbitrary constant matrix. Let the set $\{\mathbf{v}^{(i)}\}$ be a linearly independent full set of vectors that correspond to the dimensions of **G**. Let this set be called "*special according to* **G**" if it satisfies the restrictions

$$\mathbf{v}^{(i)T}\mathbf{G}\mathbf{v}^{(j)} = G_{ij} \tag{7}$$

It is trivial, that the elements of this set can form the columns of a special matrix **V**, and that this matrices satisfy the equation

$$\mathbf{v}^{(i)T}\mathbf{G}\mathbf{v}^{(j)} = G_{ij} \tag{7}$$

It is trivial, that the elements of this set can form the columns of a special matrix **V**, and that this matrices satisfy the equation

$$\mathbf{V}^T\mathbf{G}\mathbf{V} = \mathbf{G} \Rightarrow \mathbf{V}^{-1} = \mathbf{G}^{-1}\mathbf{V}^T\mathbf{G}, \tag{8}$$

i.e. generally the calculation of the inverse of these matrices is very easy and computationally cost-effective, and furthermore, these matrices form a group as well as the symplectic matrices. Furthermore, such matrices may have the determinant of only ± 1. If we restrict ourselves to the unimodular sub-group, its generators **H** have to satisfy the restriction

$$\mathbf{G}\mathbf{H} + \mathbf{H}^T\mathbf{G} = \mathbf{0} \tag{9}$$

and by using this, special Lie-groups can be constructed. The special cases in which **G** corresponds to **I**, $\Im$, and $\mathbf{g} = \langle 1,1,1,-c^2 \rangle$, result in the *Orthogonal,the Symplectic,* and the *Lorentz Group,* respectively ("*c*" is the velocity of light). The appropriate special sets are the *orthonormal,* the *symplectic,* and the *Lorentzian sets.* In these examples **G** is either symmetric (**I**, **g**) or skew-symmetric ($\Im$), consequently **H** can be constructed of skew-symmetric or symmetric **J** matrices, respectively, as $\mathbf{H}=\mathbf{G}^{-1}\mathbf{J}$.

All the considerations used to construct a mapping between the observed and the desired behavior can trivially be repeated in the case of another groups, if we suppose, that at least one element of these special sets can be an arbitrary non-zero vector. This definitely does not hold for the *Orthogonal Group* because the appropriate sets consist of paired orthogonal unit vectors. Therefore, instead of the orthogonal group the combined group of positive scalings and rotations can be used:

$$\mathbf{T} = s\mathbf{O}, s > 0, \mathbf{O}^T\mathbf{I}\mathbf{O} = \mathbf{I}, \mathbf{T}^{-1} = s^{-1}\mathbf{O}^T, \mathbf{H} = \dot{\mathbf{T}}\mathbf{T}^{-1} = \dot{s}s^{-1}\mathbf{I} + \dot{\mathbf{O}}\mathbf{O}^T = \xi\mathbf{I} + \Omega \tag{10}$$

which we can call the "*Stretched Orthogonal Group*". To make a solution unambiguous, first the columns of the unit matrix can be considered as the appropriate linearly independent set. The desired and the realized drift vectors can be normalized, then the whole set can be rigidly rotated, so that the rotation moves the first vector into the normalized vectors while leaving their orthogonal sub-space unchanged. Finally the scaling factor "*s*" can be determined on the basis of the norms. In this case the generators can be constructed from the skew-symmetric matrices Ω and the unit matrix **I**. (Though these **T** transformations trivially form a group, their application would be computationally "dangerous" because they are not unimodular, and the zero matrix can be approached by them.)

For using a "*Generalized Lorentz Group (GLG)*" exempt of this difficulty, a "fictitious dimension" can be "added" to the DOF dimensional problem first of all. So for **G** the diagonal matrix $\mathbf{g}=\langle 1,\ldots,1,-c^2 \rangle$ can be used. Now let the DOF dimensional vector **f** stand for the desired/observed joint coordinate acceleration, and let us start with the columns of the DOFxDOF dimensional unit matrix. In the first step, let this set rigidly rotated so that its first vector becomes parallel with **f** while their orthogonal sub-set remains unchanged. Let $\mathbf{e}^{(f)} = \mathbf{f}/\sqrt{\mathbf{f}^T\mathbf{f}} = \mathbf{f}/f$. It is trivial that the columns of the following matrix form a generalized Lorentzian set:

$$\left[\begin{array}{c|c|c|c|c} \mathbf{e}^{(f)}\sqrt{f^2/c^2+1} & \mathbf{e}^{(2)} & \ldots & \mathbf{e}^{(DOF)} & \mathbf{f} \\ \hline f/c^2 & 0 & \ldots & 0 & \sqrt{f^2/c^2+1} \end{array}\right] \tag{11}$$

In this solution the *physically interpreted vector* **f** is accomplished with a fictitious (DOF+1)th component, and it is placed into the last column of a generalized Lorentzian. By following rigidlyEq. (6) the proper GLG matrix, that transforms the observed acceleration into the desired one, can be calculated as

$$\mathbf{L} = \begin{bmatrix} \cdots & \mathbf{f}^D \\ \cdots & \sqrt{f^{D2}/c^2+1} \end{bmatrix} \mathbf{g}^{-1} \begin{bmatrix} \cdots & \mathbf{f}^R \\ \cdots & \sqrt{f^{R2}/c^2+1} \end{bmatrix}^T \mathbf{g} \tag{12}$$

In comparison with the combined group of stretches/shrinks and the symplectic group, this approach always uses unimodular matrices which can never approach the zero matrix, while in size they are much smaller --of (DOF+1)x(DOF+1) dimensions-- than the appropriate symplectic matrices of the dimensions of (2DOF)x(2DOF). For control and technical purposes "c" may be an arbitrary positive constant. The computational complexity of the method can be measured mainly by the algorithm creates the GLG matrices. The Scilab 2.5 syntax is given below:

```
// function [Lambda]=lorl(a,cc)
// Creates a Lorentzian set of the input vector
function [Lambda]=lorl(a,cc);
[DIM,col]=size(a);
Lambda=zeros(DIM+1,DIM+1); // the room reservation for he output Lorentzian
ort=eye(DIM,DIM); // the upper left block of the Lorentzian
na=norm(a,'fro');
small=1e-8;
if na>=small
     be=ort(:,1);
     ae=a/na;
     Fi=acos(ae'*be); // the required degree of rotation
     sFi=sin(Fi);
     cFi=cos(Fi);
     c=ae-(be'*ae)*be; // the part of ae orthogonal to be
     cnorm=norm(c,'fro');
     if (cnorm>small) // then rotation is necessary as follows
        c=c/cnorm;
        ort=eye(DIM,DIM)-be*be'-c*c'+cFi*(c*c'+be*be')+sFi*(-be*c'+c*be');
       // b is made parallel with a a-
     end; // if cnorm>small
```

```
        Lambda(1:DIM,1:DIM)=ort;
        Lambda(1:DIM,DIM+1)=a;
        Lambda(DIM+1,DIM+1)=sqrt(na^2/cc^2+1);
        Lambda(1:DIM,1)=ae*sqrt(1+na^2/cc^2);
        Lambda(DIM+1,1)=na/cc^2;
else // if na>=small
        Lambda=eye(DIM+1,DIM+1);
end; //if na>=small
```

This algorithm is two-times applied twice in the control cycle. It also contains the necessary rotation, too.

As it we saw in [12], the proper matrices of the MOST algorithm also can be constructed by additions/subtractions/rotations and stretches/shrinks, that is in general the same algebraic operations and considerations can be done for each of the groups that were considered. Without giving the proof here, we will only define the proper components of the appropriate symplectic matrices for a comparison. Let the 2DOF dimensional vector **f** consist of the DOF dimensional blocks $\mathbf{f}=[\mathbf{f}^{(1)T},\mathbf{f}^{(2)T}]^T$, then the matrix S defined below is symplectic:

$$\left[\begin{array}{cccccc|ccccc} \mathbf{f}^{(1)} & \mathbf{u}^{(2)} & \mathbf{e}^{(3)} & \dots & \mathbf{e}^{(DOF)} & & -\tilde{\mathbf{f}}^{(2)} & -\tilde{\mathbf{u}}^{(1)} & \mathbf{0} & \dots & \mathbf{0} \\ \hline \mathbf{f}^{(2)} & \mathbf{u}^{(1)} & \mathbf{0} & \dots & \mathbf{0} & & \tilde{\mathbf{f}}^{(1)} & \tilde{\mathbf{u}}^{(2)} & \mathbf{e}^{(3)} & \dots & \mathbf{e}^{(DOF)} \end{array}\right]. \tag{13}$$

The components of S are defined as follows:

$$\mathbf{u}^{(1)}=\frac{1}{\mathbf{f}^{(1)^T}\mathbf{f}^{(1)}}\left[\mathbf{f}^{(1)}-\frac{\mathbf{f}^{(2)^T}\mathbf{f}^{(1)}}{\mathbf{f}^{(2)^T}\mathbf{f}^{(2)}}\mathbf{f}^{(2)}\right],\mathbf{u}^{(2)}=\frac{1}{\mathbf{f}^{(2)^T}\mathbf{f}^{(2)}}\left[\mathbf{f}^{(2)}-\frac{\mathbf{f}^{(1)^T}\mathbf{f}^{(2)}}{\mathbf{f}^{(1)^T}\mathbf{f}^{(1)}}\mathbf{f}^{(1)}\right] \tag{14}$$

and

$$\tilde{\mathbf{f}}^{(1)}=\frac{\mathbf{f}^{(1)}}{\mathbf{f}^{(1)^T}\mathbf{f}^{(1)}+\mathbf{f}^{(2)^T}\mathbf{f}^{(2)}},\tilde{\mathbf{f}}^{(2)}=\frac{\mathbf{f}^{(2)}}{\mathbf{f}^{(2)^T}\mathbf{f}^{(2)}+\mathbf{f}^{(1)^T}\mathbf{f}^{(1)}},\text{ and} \tag{15}$$

$$\tilde{\mathbf{u}}^{(1)}=\frac{\mathbf{u}^{(1)}}{\mathbf{u}^{(1)^T}\mathbf{u}^{(1)}+\mathbf{u}^{(2)^T}\mathbf{u}^{(2)}},\tilde{\mathbf{u}}^{(2)}=\frac{\mathbf{u}^{(2)}}{\mathbf{u}^{(2)^T}\mathbf{u}^{(2)}+\mathbf{u}^{(1)^T}\mathbf{f}^{(1)}}. \tag{16}$$

The $\{\mathbf{e}^{(i)}\}$vectors form a DOF dimensional (arbitrary) orthonormal set is rotated so rigidly that its 1st vector $\mathbf{e}^{(1)}$ becomes parallel to $\mathbf{f}^{(1)}$ and the orthogonal sub-space of these vectors remains unchanged. As a result both $\mathbf{e}^{(2)}$ and $\mathbf{u}^{(2)}$ become orthogonal to $\mathbf{f}^{(1)}$. In the second step $\mathbf{e}^{(2)}$ is made parallel to $\mathbf{u}^{(2)}$ by rigidly rotating the whole set again in a special way, while leaving the orthogonal subset of these vectors unchanged. Consequently the previously set $\mathbf{e}^{(1)}$ also remains unchanged. This procedure determines the $\{\mathbf{e}^{(i)}|i=3, 4,..., DOF\}$ unit vectors for the matrix in Eq. (13).

This construction supposes that both of the blocks of **f** are different from zero. For these special cases, additional "tricks" must be used which practically switch off the "identification" of such cases and use the previously identified matrix or the unit matrix, or add some minor vectors to the input of the identification in order to avoid these cases. Later these ideas are investigated in the mirror of stability requirements.

4. STABILITY CONSIDERATIONS

From a purely mathematical point of view the control problem can be formulated as follows: there is some *imperfect model of the system* that is based on some *excitation* that is calculated for a desired input $\mathbf{i}^{d}$ as $\mathbf{e}=\varphi(\mathbf{i}^{d})$. The system has its *inverse dynamics* described by the *unknown function* $\mathbf{i}^{r}=\psi(\varphi(\mathbf{i}^{d}))=f(\mathbf{i}^{d})$ and this results in a realized $\mathbf{i}^{r}$ instead of the desired one $\mathbf{i}^{d}$. (In Classical Mechanics these values are the *desired* and the *realized joint accelerations*, while the external free forces and the joint velocities serve as the parameters of this temporarily valid and changing function.) It is evident, that we can normally obtain information via observation only on the "net" function $f()$, and that this function varies considerably in time. Furthermore, we do not have the practical tools to "manipulate" the nature of this function directly: generally we can manipulate or *deform* its actual input $\mathbf{i}^{d*}$ by comparing it with the *desired one*. The aim here is to achieve and maintain the $\mathbf{i}^{d}=f(\mathbf{i}^{d*})$ state. We can directly manipulate only the nature of the *model function* $\varphi()$.

Farly similar thing often occurs in theoretical physics. This example could be called a kind of "renormalization". For instance, suppose, that for given function $g()$,where an iteration is defined as $x_{n+1}=g(x_{n+1})$, it is given a *hypothetical fixed point* pertaining to this iteration on the basis of some other considerations, x^{d}. Actually it may well happen that the given function $g()$ does not yield a convergent series, or the series may be convergent but it converges to some other value that differs considerably from the desired fixed-point. For a single-dimensional (Single Input, Single Output or SISO) system there is thepossibility of *so manipulating* $g()$ with a scalar scaling factor s as $\gamma(x):=s^{-1}g(sx)$, so that the deformed function yields the desired fixed point. For determining the proper deformation factor the following iteration can be invented:

$$x^{d}=s_{n+1}^{-1}g\left(s_{n}x^{d}\right)\text{which may be convergent as } s_{n}\xrightarrow[n\to\infty]{}s \tag{17}$$

Let us suppose that in the given region $g()$ is *contractive*, i.e. for the arbitrary input values a and b

$$\left\|g(a)-g(b)\right\|\le K\left\|a-b\right\|, 0<K<1 \tag{18}$$

In this case, for the above sequence, it holds that

$$\left\|\left(s_{n+1}-s_{n}\right)x^{d}\right\|=\left\|g\left(s_{n}x^{d}\right)-g\left(s_{n-1}x^{d}\right)\right\|\le K\left\|\left(s_{n}-s_{n-1}\right)x^{d}\right\|\le\ldots\le K^{n}\left\|\left(s_{1}-s_{0}\right)x^{d}\right\|\xrightarrow[n\to\infty]{}0, \tag{19}$$

and a Cauchy sequence can be obtained. For a SISO system this sequence is confined in the set of real numbers, namely it is also convergent. For non-zero x^{d} this means a convergent sequence $\{s_{n}\}$ which trivially yields the solution for Eq. (17). It is interesting that the contractive nature of $g()$ itself guarantees this result.

It also is quite plausible that the above considerations could be extended to SISO systems of finite dimensions, in which case scalar parameters should be replaced with some quadratic and invertible matrices $\{S_n\}$. In this context the first difficulty is that Eq. (17) does not have a unique solution, and that the calculation of the inverse of some general invertible matrix is very inefficient from computational point of view. The second problem is that even if *g*() is contractive, then by using only Eq. (19) no conclusion can be obtained for the convergence of the sequence $\{S_n\}$. It only contains information for the behavior of the sequence in relation to the special vector $\mathbf{x}^d$.

It is plausible that both of these difficulties can be conveniently evaded if special restrictions are imposed on the set $\{S_n\}$: a) let its element be the members of some special Lie groups outlined in Paragraph 3: this immediately solves the problem withthe calculation of the inverse matrices; b) within the mathematical framework of the appropriate Lie groups the ambiguity of the solution still remains an open problem; to resolve this the simple constructions that were discussed in Paragraph 3 e.g. the MOST algorithm, the generalized Lorentzians, etc. can be applied; since these constructions control the behavior of the matrix sequence in the directions linearly independent of the special vector $\mathbf{x}^d$, the convergence of the sequence is also guaranteed to be simply process due to the contractive nature of *g*(). If the processes of this iteration are much faster than the change of $\mathbf{x}^d$ in time, the above method would be used for real-time control purposes. (Similar considerations that are related to the concept of *complete stability* of the CNNs are applied.)

Though these considerations seem to be very attractive, from a *phenomenological point of view,* it cannot be realized in real time control, because there is normally no manipulate the system's response with S^{-1}. Only the input can be deformed, and the output can be measured. This is why it is expedient to consider the possible and simple modification of the "renormalization algorithm". For a SISO system let us consider the following sequence:

$$i_1 = s_1 i_0, \ldots, i_{n+1} = s_{n+1} i_n, \text{ in which } i_0 = s_{n+1} f(i_n), \text{ and } f(i_0) < i_0 \tag{20}$$

If, for example, for a positive x $f(x)$ is positive, with monotone increasing, and there exists a constant K for which

$$1 < \frac{x}{f(x)} < K, \text{and } g(x) := \frac{x}{f(x)} \text{ is monotone and } f(x) \text{ is not bounded}, \tag{21}$$

the following estimation can be done:

$$i_{n+1} - i_n = \frac{i_0}{f(i_n)} i_n - \frac{i_0}{f(i_{n-1})} i_{n-1} \equiv s_{n+1} i_n - s_n i_{n-1} = i_0 \left(\frac{i_n}{f(i_n)} - \frac{i_{n-1}}{f(i_{n-1})} \right) > 0 \tag{22}$$

that is the sequence $f(i_n)$ is monotone increasing. Consequently the sequence $s_n = i_0 / f(i_{n-1})$ is monotone decreasing with $s_0 > 1$. This sequence evidently converges to $s_\infty = 1$. A trivial example for exemplifying that the given set of functions is not empty is $f(x) = ax + b$, a, $b>0$, $x/f(x) < 1/a$ monotone increasing and bounded.

It is natural that a similar scenario can be imagined, for MIMO systems also, in analogy with the modified renormalization algorithm with similar special matrices replacing the scalar multiplication factors as

$$\mathbf{i}_0;\mathbf{S}_1 f(\mathbf{i}_0)=\mathbf{i}_0;\mathbf{i}_1=\mathbf{S}_1\mathbf{i}_0;\ldots;\mathbf{S}_n f(\mathbf{i}_{n-1})=\mathbf{i}_0;\mathbf{i}_{n+1}=\mathbf{S}_{n+1}\mathbf{i}_n;\mathbf{S}_n \xrightarrow[n\to\infty]{} \mathbf{I}\,. \tag{23}$$

In the MIMO case, the desired convergence can be guaranteed in several manners. The $f(\mathbf{i}_n)\to\mathbf{i}_0$ requirement can be more restricted e.g. in a monotone form in the norm:

$$\|\mathbf{d}_n\|\equiv\|f(\mathbf{S}_n\mathbf{i}_{n-1})-\mathbf{i}_0\|=\|(f(\mathbf{S}_n\mathbf{i}_{n-1})-f(\mathbf{i}_{n-1}))+(f(\mathbf{i}_{n-1})-\mathbf{i}_0)\|=\|\Delta\mathbf{d}_n+\mathbf{d}_{n-1}\|<\|\mathbf{d}_{n-1}\| \tag{24}$$

When using an Euclidean norm, the previous inequality can be imposed on the squares of the norm. By taking into account the special restrictions in Eq. (22)

$$f(\mathbf{S}_n\mathbf{i}_{n-1})-\mathbf{i}_0=f(\mathbf{S}_n\mathbf{i}_{n-1})-\mathbf{S}_n f(\mathbf{i}_{n-1})+\mathbf{S}_n f(\mathbf{i}_{n-1})-\mathbf{i}_0=f(\mathbf{S}_n\mathbf{i}_{n-1})-\mathbf{S}_n f(\mathbf{i}_{n-1}) \tag{25}$$

the last two terms cancel. This leads to the special structure

$$\begin{aligned}\mathbf{d}_{n-1}+\Delta\mathbf{d}_n&=f(\mathbf{S}_n\mathbf{i}_{n-1})-\mathbf{S}_n f(\mathbf{i}_{n-1})\\ \Delta\mathbf{d}_n&=f(\mathbf{S}_n\mathbf{i}_{n-1})-f(\mathbf{i}_{n-1})\end{aligned} \tag{26}$$

Applying this for the inequality imposed on the squares of the norms an inequality can be obtained which does not contain explicitly the S matrices:

$$f^2(\mathbf{i}_n)-f^2(\mathbf{i}_{n-1})<2\mathbf{i}_0^T(f(\mathbf{i}_n)-f(\mathbf{i}_{n-1})) \tag{27}$$

This can be imposed e.g. as a "general requirement" which can be met by a class of functions. Though these considerations do not guarantee convergence to the exact behavior, however, model improvement can be achieved by them. In the next paragraph simulation examples are presented for the most "popular" paradigm, the control of the inverted pendulum.

5. SIMULATION RESULTS

The inverted pendulum has the usual structure with one linear (q_1 describing the translation of a coach in *m* unit), and one rotational (q_2 corresponding to the angle of the pendulum with respect to the vertical direction in *rad* unit) degree of freedom. The generalized forces to be exerted by the drives of the appropriate degrees of freedom are Q_1 force in *N*, and Q_2 torque in *Nm* units. Three kinds of control are compared: a kinematically prescribed PD control with an *exact dynamic model*, the same kinematically prescribed PD control with arough dynamic model *without any adaptation*, and the same kinematic requirements with adaptive control that are realized by the generalized Lorentzian matrices. It is evident from Fig. 1 that the rough model differs considerably from the real one, and that the generalized Lorentzians-based adaptation considerably increases control quality.

In Fig. 2 the generalized momentums are displayed for the above three cases. It reveals that the dynamic model is "under-estimated", and adaptivity increases the range of the generalized forces. Though the shape of the momentum curves in the graphs differs considerably from that of the real model, no drastic changes appear in it as for example, is case with a bang-bang controller, or in the case of a VS controller's transient phase. (The dynamic under-estimation of the rough model is used for the expected "flatness" of the overall function to be observed experimentally.) It is worth noting that in the control elements described here fluctuation was observed in the

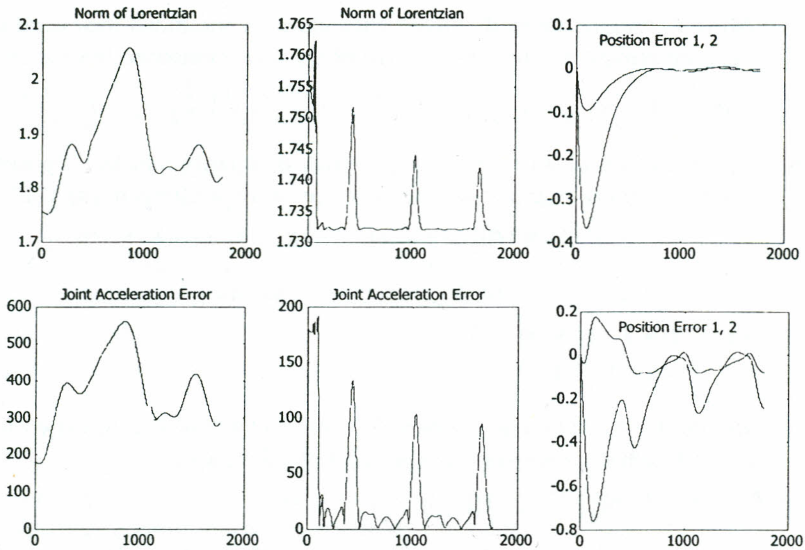

Fig. 1. The quality of control for the same, kinematically expressed PD-type, the nominal trajectory reproduction with the *exact dynamic model, the rough model without adaptation, and the rough model supported with the GLG-based adaptation* (R stands for the *realized*, N denotes the *nominal* motion, q_1 in *m*, q_2 in *rad* units)

To obtain more detailed information about the control in Fig. 3 the norms of the Lorentzians and the joint acceleration errors are described, for boththe "*simple rough*" model and the "*adaptive*" control. (In the case of the *simple rough model* the Lorentzians are not used in the control. These figures convey information on the first Lorentzian needed if the adaptive algorithm were to be switched on just at a given moment.) The *adaptive law* is switched on at the 50^{th} time unit and this can be seen on the graphs.

It is easy to see that the norm of the *unused Lorentzians* are in strict correlation with the acceleration error, and this is defined as the difference between the *desired acceleration* obtained only on the basis of kinematical considerations and the PD control, and the realized (simulated) acceleration. This considerable difference is brought about because of the very rough nature of the dynamic model used.

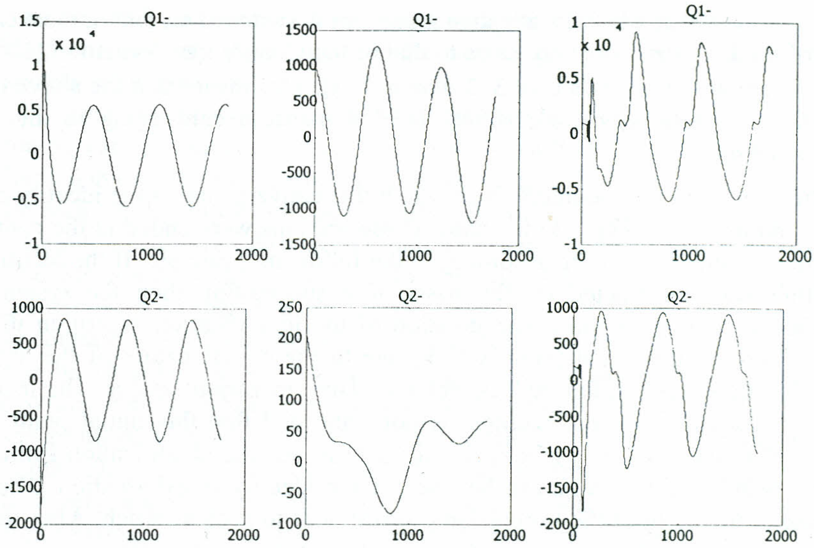

Fig. 2. The generalized forces (Q_1 N" for the linear, and Q_2 in "Nm" for the rotary axis) for the exact dynamic model, the rough model without adaptation, and the rough model supported by the GLG-based adaptation

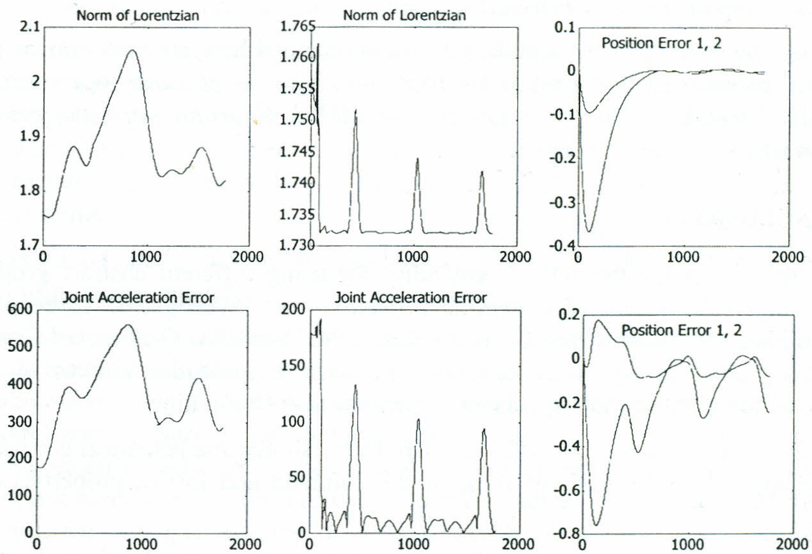

Fig. 3. The norm of the identifying GLG matrix, and the joint accelerations' error for the *rough model without adaptation* (that is without using this matrix in the control), *and the rough model with adaptation* (utilizing the GLG) in correlation with the position error. (For comparison, the position error for the exact dynamics is also given.)

It is evident that when the adaptive algorithm is switched on that in most cases the norm of the Lorentzians is very close to that of the identity transformation ($3^{1/2}$ for a 2 DOF system with Lorentzians of 3×3 dimensions). This means that the above outlined "*modified renormalization algorithm*" yielded matrices converging to the identity transformation.

There are typical time-intervals in which this convergence to the identity operator ceases and the norms begin to increase. These sessions were ended in the control law by built-in "curbs" operating according to the following principle: If the identification algorithm was constructed on the basis of a supposition then the system to be identified is "static" in time. The iteration of the identification algorithm in such a session starts at the initial point, which, due to the causal nature of the algorithm, influences those results, that will be obtained later, in various stages. The increase in the norm indicates that the problem is not static and that this initial point became "obsolete" in the session, e.g. it is expedient to restart the identification from a "new starting point" and the same can be said for the identity transformation. The abrupt decrease in the norm of the Lorentzians corresponds to such a switch. After this jump a stable session is initiated again with near unity transformations, etc.

It also is evident that the little acceleration error of the "stable sessions" results in fast improvement of the precision of trajectory reproduction. (For the sake of comparison, the trajectory reproduction error is also displayed for the case in which the *exact dynamic model* was used.)

It can be stated that the simulation results obtained here are very similar to those obtained recently for a special 3DOF robot arm used for polishing operation. In that case the *"stretched orthogonal group", the MOST algorithms and the generalized Lorentzians were compared.*

6. CONCLUSIONS

In this paper the theoretical possibility for using different abstract groups as a simple algebraic means of system-identification was investigated. Three particular possibilities, the "*Generalized Lorentz Group*", the "*Stretched Orthogonal Group*" and the "*Symplectic Group*" were described in detail as potential sources of uniform structures to be utilized in a special new brand of Soft Computing.

It was found that in each case considered very similar mathematical considerations can be applied on the basis of some of the common and formal properties of these groups.

Convergence problems were discussed at quite a general level, and they were discussed independently of the particular group that was going to be used to solve a problem.

The new approach has certain essential advantages over the "traditional means" of Soft Computing:

- The size and the number of the free parameters of the uniform structure to be used are uniquely determined by the degree of freedom of the mechanical system to be controlled;
- The generally obscure process of adaptive machine learning or parameter tuning can be executed by applying simple, definite algebraic steps that are limited in number; no "stochastic" approach is necessary;
- In each of the cases considered, these algebraic steps consist of traditional rotations, stretches or shrinks and subtractions in different abstract spaces;
- The dimensions of the appropriate abstract spaces also depend on the particular group chosen for this purpose;
- In contrast to the "backpropagation"-based learning, the solutions obtained thus can be regarded as not "local" but "global" optima, in the sense that the essential transformations are restricted to only those sub-spaces of the abstract spaces on which the actual information is available. No transformation happens in the orthogonal sub-spaces for which no any information is available;
- It is very important that in each of the cases considered, the inverse of the appropriate matrices can be calculated in a very cost-efficient way;
- In each case we have a Lie group, therefore the control being considered here can be accomplished with some extrapolation between two subsequent control steps, in the tangent space of the group elements.
- With extrapolation, more new parameters could be introduced, and they too could be "tuned" according to traditional stochastic tuning methods.

One particular group, the *generalized Lorentzian,* was investigated via simulation in the case of a mechanical system, and in the control of the *inverted pendulum*, which is a popular paradigm in control technology. In the simulation only the c=1 value was applied. The role of this parameter whitin the control needs to be investigated further. We can expect that different values of this parameter may fit into different particular physical systems.

It was found that considerable improvement in the control quality by can be achieved using these methods.

It can be expected that those considerations presented herecan be extended to a wider class than just the control of mechanical systems. For instance, the inductance of the volatage-controlled DC motors causes an inertia in the change of the motor current, and consequently in the motor torque, which is very similar to mechanical inertia. So electric and electro-mechanical systems seem to form a prospective problem-class together with the purely mechanical systems. Though no exact theorem was formulated for guaranteeing the convergence of the method, theoretical considerations indicate that in a quite wide class of practical problems the method may well work.

Generally speaking, more "exact" proof of convergence is needed, because the idea of "adaptive control on the basis of the Symplectic Group" is natural only in the case of mechanical systems. (At this present stage the new approach was investigated only in connection with mechanical systems.)

Generally it can be concluded that this approach is appealing and it deserves further theoretical investigations as well as simulation tests for different physical systems.

7. REFERENCES

[1] R. Reed. Pruning Algorithms - A Survey. *IEEE Transactions on Neural Networks*, 4., pp.- 740-747, 1993.

[2] S. Fahlmann, C. Lebiere.The Cascade-Correlation Learning Architecture. *Advances in Neural Information Processing Systems*, 2, pp. 524-532, 1990.

[3] T. Nabhan, A. Zomaya. Toward Generating Neural Network Structures for Function Approximation. *Neural Networks*, 7, pp. 89-9, 1994.

[4] G. Magoulas, N. Vrahatis, G. Androulakis. Effective Backpropagation Training with Variable Stepsize. *Neural Networks*, 10, pp. 69-82, 1997.

[5] C. Chen, W. Chang. A Feedforward Neural Network with Function Shape Autotuning. *Neural Networks*, 9, pp. 627-641, 1996.

[6] W. Kinnenbrock. Accelerating the Standard Backpropagation Method Using a Genetic Approach. *Neurocomputing*, 6, pp. 583-588, 1994.

[7] A. Kanarachos, K. Geramanis. Semi-Stochastic Complex Neural Networks. *IFAC-CAEA '98 Control Applications and Ergonomics in Agriculture*, pp. 47-52, 1998.

[8] J. K. Tar, O. M. Kaynak, J. F. Bitó, I. J. Rudas. Generalised Canonical Equations of Classical Mechanics: A New Approach towards the Adaptive Treatment of Unmodelled Dynamics. *Proceedings of the International Conference on Recent Advances in Mechatronics (ICRAM'95)*, 14-16 August, 1995, Istanbul, Turkey, pp. 1092-1097.

[9] J. K. Tar, I. J. Rudas, J. F. Bitó. Group Theoretical Approach in Using Canonical Transformations and Symplectic Geometry in the Control of Approximately Modeled Mechanical Systems Interacting with Unmodelled Environment. *Robotica*, Vol. 15, pp. 163-179, 1997.

[10] J. K. Tar, I. J. Rudas, J. F. Bitó, O. M. Kaynak. A New Utilization of the Hamiltonian Formalism in the Adaptive Control of Mechanical Systems Under External Perturbation. *Intelligent Automation and Soft Computing*, Vol. 5, Issue: 4, 1999, pp. 303-312.

[11] V. I. Arnold. *Mathematical Methods of Classical Mechanics.* (in Hungarian) Műszaki Könyvkiadó, Budapest, Hungary 1985.

[12] J. K. Tar, I. J. Rudas, Touré Balladji. Minimum Operation Algebraic Adaptive Control for Robot Based on a Novel Branch of Soft Computing. In the: *Proc. of EUREL '2000*, Vol. 2, Session 3, (European Advancded Robotics Systems - Masterclass and Conference - Robotics 2000), 12-14 April 2000, The University of Salford, Manchester, UK.

Received: 13 Mart 2000
Accepted: 27 May 2000

József K. Tar
Miklós Rontó

ADAPTIVNA KONTROLA BAZIRANA NA PRIMJENI POJEDNOSTAVLJENIH UNIFORMNIH STRUKTURA I POSTUPCI UČENJA

Sažetak

Prikazano je trenutno stanje pristupa kojemu je cilj stvoriti novu granu Soft Computinga (SC) za određene katergorije problema koji bi mogli biti širi od problema kontrole mehaničkih sustava. Kao i "tradicionalni" SC, ovaj pristup izbjegava razvoj analitičkih modela sustava i pokušava koristiti jednostavne uniformne strukture, ali, za razliku od tradicija, ove strukture dobivene su od različitih Lievih grupa koje se koriste u raznim područjima fizike. Glavne prednosti su drastično smanjenje veličine i povećanje jasnoće u usporedbi s "konvencionalnim" arhitekturama. Druga prednost je što izgleda da se općenito "opskurno", bilo da je strogo uzročno, polustohastičko ili potpuno stohastičko, "učenje" ili podešavanje parametara mogu zamijeniti jednostavnim eksplicitnim algebarskim postupcima ograničenih koraka u slučaju novih struktura. Temeljna ideja proizašla je iz područja kontrole mehaničkih sustava i glavne opće unutarnje simetrije mehaničkih sustava, a kasnije se dalje razvijala uzimajući u obzir određena opća svojstva ove unutarnje simetrijske grupe mnogo detaljnije. U ovom radu također su razmatrana proučavanja konvergencije za MIMO i SISO sustave. Prikazani su primjeri simulacije za kontrolu invertiranog njihala pri čemu su u tu svrhu korištene uopćene Lorentzove matrice. Zaključeno je da je ova metoda obećavajuća i vjerojatno će zadati prihvatljive konvergencijske uvjete za mnoge slučajeve.

Ključne riječi: adaptivna kontrola, Lie-grupe, parcijalna identifikacija sustava, Lorentzove grupe.